Botany
An International Journal for Plant Biology

OPEN ACCESS | Research Article

Bacterial diversity and community structure in native legume root nodules in intact and mining-disturbed Arctic tundra

Angie S. Li[a], **Katherine M. Standen**[a], **Katherine J. Stewart**[b], **and Eric G. Lamb**[a]

[a]Department of Plant Sciences, University of Saskatchewan, Saskatoon, SK S7N 5A8, Canada; [b]Department of Soil Science, University of Saskatchewan, Saskatoon, SK S7N 5A8, Canada

Corresponding author: **Eric G. Lamb** (email: eric.lamb@usask.ca)

Abstract

Mining restoration in Arctic regions is challenging due to harsh environmental conditions with slow natural ecosystem recovery taking decades or even centuries. Identifying native species that can both colonize disturbed areas and contribute to soil development is critical to hasten restoration timelines. Nodules were collected from four native legume species (*Astragalus alpinus* Linnaeus, *Hedysarum americanum* (Michaux ex Pursh) Britton, *Oxytropis arctica* R. Brown, and *Oxytropis maydelliana* Trautvetter) observed to be naturally colonizing gravel quarries within a mine footprint near Rankin Inlet, Nunavut, Canada. Samples came from both gravel quarries and adjacent intact tundra. Next-generation sequencing of the 16S and *nifH* regions was used to characterize the nodule bacterial community composition and diversity. Despite the large differences in soil conditions between gravel substrates and intact tundra, no significant effects of soil environment were found on bacterial community composition within plant nodules. Further, few differences were observed in nodule communities between the plant species. Overall, the study suggests that the microbial propagules necessary for successful nodulation are present in gravel quarries. While restoration efforts involving native legumes may succeed without commercial inoculants, further research is needed to determine whether the rhizobia in these environments can provide sufficient nitrogen to support robust host plant growth.

Key words: *Rhizobium*, *Mesorhizobium*, nitrogen fixation, Arctic tundra, restoration ecology

Introduction

Mining activities commonly result in the widespread removal of vegetation, primarily through the establishment of gravel quarries and roads, resulting in soil that lacks an adequate amount of organic material (Naeth and Wilkinson 2014; Miller et al. 2021). Restoring such landscapes has significant challenges, especially in tundra regions characterized by low temperatures, limited precipitation, and short growing seasons. These conditions can hinder essential ecosystem functions such as decomposition, nutrient cycling, and vegetative growth rates (Forbes and Jefferies 1999; Naeth and Wilkinson 2014; Mehlhoop et al. 2018). Given these conditions, the successful restoration of disturbed soils into productive ecosystems may span decades or even centuries (Forbes and Jefferies 1999; Hodkinson et al. 2003).

The slow growth of Arctic plants and the limited availability of commercially viable native seed stock pose challenges to restoration work in Arctic environments. Past restoration approaches have turned to the use of fast-growing non-native species for quick revegetation and erosion control. However, this practice is increasingly unacceptable in northern mine site closure plans due to the risks of non-native species invasion, reductions in biodiversity, alteration of ecosystem function, and persistence of non-target plant communities. This can result in less successful long-term recovery of ecosystem function to pre-disturbance conditions (Forbes and Jefferies 1999; Hagen et al. 2014; Kearns et al. 2015; Rydgren et al. 2016; Vloon et al. 2022). Adding to these challenges, limited infrastructure in remote Arctic locations necessitates more straightforward, cost-effective methods, with the transplantation and spreading of tundra-sourced material possibly proving more suitable.

Recent studies show that transplantation of plugs of intact tundra can be effective for restoration of upland heath Arctic tundra, but the availability of material for transplantation is a serious limitation (Hnatowich et al. 2022, 2023). Anecdotal observations during fieldwork for the above studies suggest that two species of native legumes, *Astragalus alpinus* Linnaeus and *Oxytropis arctica* R. Brown (Fig. A1), are among the first natural colonizers of these gravel quarries. Given the scarcity of soil organic material after resource extraction, primary colonizing nitrogen (N)-fixing plants may shape early successional ecosystems through the deposition of high-quality organic litter and root exudates (Bradshaw 1993; Myrold and Huss-Danell 2003; Stewart and Siciliano 2015). Litter and exudates enhance the availability of carbon (C) and N in the soil,

Table 1. Nodule collection sites and number of samples collected per legume species and site.

Site	Quarry	Substrate type	*A. alpinus*	*O. arctica*	*O. maydelliana*	*H. americanum*
1	Q2	Gravel	1	1	1	0
1	Q2	Tundra	1	1	1	1
2	Q2	Gravel	1	1	1	0
2	Q2	Tundra	1	1	1	1
3	Q1	Gravel	1	1	1	1
3	Q1	Tundra	1	1	1	1
4	Q1	Gravel	1	1	1	0
4	Q1	Tundra	1	1	1	1
5	Q3	Gravel	1	1	1	1
5	Q3	Tundra	1	1	1	1
		Total	10	10	10	7

Note: Composite samples consisted of pooled nodules from individuals located within a 30 m diameter area at the site. Each site was separated by at least 300 m. Gravel and tundra substrate areas were within 100 m of each other for each site.

influencing microbial and plant community dynamics by providing enduring nutrient sources for microbial communities. This, in turn, promotes microbial nutrient mineralization, facilitating the colonization of later successional species (Bardgett and Walker 2004; Knelman et al. 2012; Iversen et al. 2015; Stewart and Siciliano 2015).

Our understanding of the root nodule communities of Arctic legumes remains limited. Rhizobial symbionts often establish a close association with specific plant species, relying on signals exchanged between the host and symbiont(Oldroyd et al. 2011). Though certain associations are often preferred, these relationships do not seem to be strictly deterministic (Masson-Boivin et al. 2009). Rhizobial populations, for example, can be genetically distinct among legume populations of the same species, even when those populations are separated by only a few kilometres (Van Cauwenberghe et al. 2014). Significant changes in rhizobial composition within nodules have been observed along environmental gradients (Birnbaum et al. 2018). Additionally, the distribution of propagules in the field may be patchy or scarce, leading to site-specific variation (Simonsen et al. 2017). Upon the death of nodules, rhizobial propagules may be released into the soil where they can persist for years without host plants (Denison and Kiers 2011). Ultimately, variation in nodule microbial community assemblies among different legume species and soil types may play a role in the success rate of colonization and subsequent succession through differences in the deposition of soil resources, impacting soil nutrient cycling.

Here, our objectives were to compare the nodule microbial communities between four native Arctic legume species collected from both gravel substrates and intact tundra substrates in Nunavut, Canada. We use 16S and *nifH* amplicon sequencing to determine how variation in community structure, diversity, and abundance differ between substrate types and among various legume species.

Materials and methods

Study site and species

Our collection sites were located near the Agnico Eagle Mines Ltd. Meliadine mine, approximately 25 km northwest of Rankin Inlet, Nunavut, Canada (63°01′23″N 92°11′41″W). The predominant vegetation in the area consists of heath tundra, heath-lichen, and lichen-rock vegetation. Among the vascular plant species commonly found are dwarf shrubs such as *Cassiope tetragona* (Linnaeus) D. Don (Arctic Mountain heather), *Dryas integrifolia* Vahl (White Mountain-avens), *Salix reticulata* Linnaeus (net veined willow), and *Vaccinium uliginosum* Linnaeus (bog blueberry). Additionally, various forbs populate the region: *Cardamine digitata* Richardson (Richardson's bittercress), *Pedicularis flammea* Linnaeus (red-tipped lousewort), and *Stellaria longipes* Goldie (long-stalked starwort) (Hnatowich et al. 2022, 2023). Three abandoned gravel quarries and nearby undisturbed upland heath tundra, located within a 100 m radius of the gravel quarries, were selected for habitat comparison (Fig. A2). The substrate in these gravel quarries consisted primarily of gravel and sand particles, with significantly lower organic matter content compared to the surrounding tundra. For instance, in 2019, one quarry had total organic carbon (TOC) levels of 17 g/kg and total organic nitrogen (TON) levels of 0.95 g/kg, whereas the nearby tundra contained 250 g/kg TOC and 11 g/kg TON (Hnatowich et al. 2023). Five collection sites were selected from the three gravel quarries and associated undisturbed heath tundra locations. Each collection site was separated by at least 300 m. At each collection site, nodules from four native legume species found on both gravel quarry and intact tundra: *Astragalus alpinus*, *Hedysarum americanum* (Michaux ex Pursh) Britton, *Oxytropis arctica*, and *Oxytropis maydelliana* Trautvetter were collected (Fig. A1; Table 1). All of the native legumes are widely distributed Arctic legume species (Porsild and Cody 1980; Flora of North America Editorial 1993); *Astragalus alpinus* and *O. arctica* were commonly found establishing on the gravel substrates, while the other two species were rarely observed colonizing the gravel substrates. All species were abundant in adjacent intact tundra at each location.

Sample collection and preparation

Samples were collected from all four species on 17 July and 18 July 2021. These consisted of a composite sample of nodules pooled from plants within a 30 m diameter for each species and substrate type, with the goal of filling a 50 mL collection tube with nodules and associated roots. Nodules were collected by carefully digging around the plant to

extract intact shoots and roots for each plant. Digging intact shoot–root systems was important, particularly in intact tundra sites, because plant roots were entangled with neighbouring plants and this ensured that the nodules collected were from the correct species. Nodules were placed in sterile plastic collection tubes and stored at 4 °C before being packed into a cooler with ice packs for transport to the University of Saskatchewan and subsequent storage at −20 °C.

Frozen nodules were separated from the roots in each sample and weighed prior to DNA extraction. These samples of nodules weighed 5–91.3 g, with an average of 41.7 g. To ensure only DNA from nodule occupants was extracted, nodules were surface sterilized by shaking in a 0.6% *w*/*v* NaClO solution for 10 min, rinsed with sterile distilled water, and then shaken with 70% ethanol for 10 min. Finally, nodules were rinsed with sterile distilled water three times (Hakim et al. 2018). Nodules were kept on ice between steps. In total, 37 samples were processed (Table 1).

Each sample was spiked with an internal standard using 1.75 μL of 33 ng/μL *Aliivibrio fischeri* (Beijerinck 1889) Urbanczyk et al. 2007 DNA before DNA extraction following the standard recommended DNeasy® Plant Pro Kit protocol (Qiagen, Cat. No. #69204).For tissue disruption, samples were placed in a Retsch Grinder and run at 25 Hz for 2 min. Samples were rotated and run again for 2 min. DNA samples were quantified following a standard protocol for the Qubit 2.0 Fluorometer (Invitrogen, Massachusetts, USA) with the Qubit HS assay kit (Invitrogen, Massachusetts, USA) before DNA samples were standardized to 1.5 ng DNA/μL in preparation for PCR amplification.

Primer sets used for PCR amplification were 342F/806R and *nifH*-F/*nifH*-R primers with Illumina sequencing adapters (342F: 5′-ACA CTG ACG ACA TGG TTC TAC ACT ACG GGG GGC AGC AG-3′ and 806R: 5′-TAC GGT AGC AGA GAC TTG GTC TGG ACT ACC GGG GTA TCT-3′; *nifH*-F: 5′-AAA GGY GGW ATC GGY AAR TCC ACC AC-3′ and *nifH*-R: 5′-TTG TTS GCS GCR TAC ATS GCC ATC AT-3′) (Rösch et al. 2002; Mori et al. 2014).

Each 20 μL PCR reaction mix contained 0.25 μL DreamTaq polymerase, 2.5 μL DreamTaq reaction buffer, 2.5 μL 2 mmol/L dNTP (all from Thermo Fisher Scientific, Massachusetts, USA), 0.25 μL of each of the forward primer and reverse primers (10 μmol/L), 12.25 μL nuclease-free water, and 2 μL DNA template. Negative controls and PCR duplicates were included. The 16S PCR conditions were 95 °C for 5 min; 30 cycles of 95 °C for 30 s, 54 °C for 45 s, 72 °C for 1 min; and a final extension at 72 °C for 7 min. *nifH* PCR conditions were 94 °C for 5 min; 35 cycles of 94 °C for 30 s, 52 °C for 45 s, 72 °C for 30 s; and a final extension at 72 °C for 10 min. PCR products were purified using Sera-Mag Select (Cytiva, Marlborough, USA) at a 1:1 ratio and diluted to 1.5 ng/μL for library preparation.

Library preparation used the Nextera XT Index Kit v2 (Illumina, San Diego, USA) to attach dual indices and sequencing adapters. Each 25 μL PCR reaction mix contained 0.25 μL DreamTaq polymerase, 2.5 μL DreamTaq reaction buffer, 2.5 μL 2 mmol/L dNTP (all from Thermo Fisher Scientific, Massachusetts, USA), 2.5 μL of each of the forward primer and reverse primers (10 μmol/L), 12.25μL nuclease-free water, and 2 μL DNA template. Cycling conditions were 95 °C for 3 min; 8 cycles of 95 °C for 30 s, 55 °C for 30 s, 72 °C for 30 s; and a final extension at 72 °C for 7 min. Indexed samples were purified with Sera-Mag Select (Cytiva, Marlborough, USA), standardized to 4 nmol/L, and pooled.

Sequencing was performed at the Western College of Veterinary Medicine Molecular Microbiology Lab, University of Saskatchewan, Canada, on the Illumina MiSeq system using a 300-cycle MiSeq Reagent Kit v2 (Illumina, Cat. No. #MS-102-2002). A total of 41, 16S samples were sequenced, including two negative controls and two duplicates. For *nifH* primers, a total of 40 samples were sequenced, including one negative control and two duplicates.

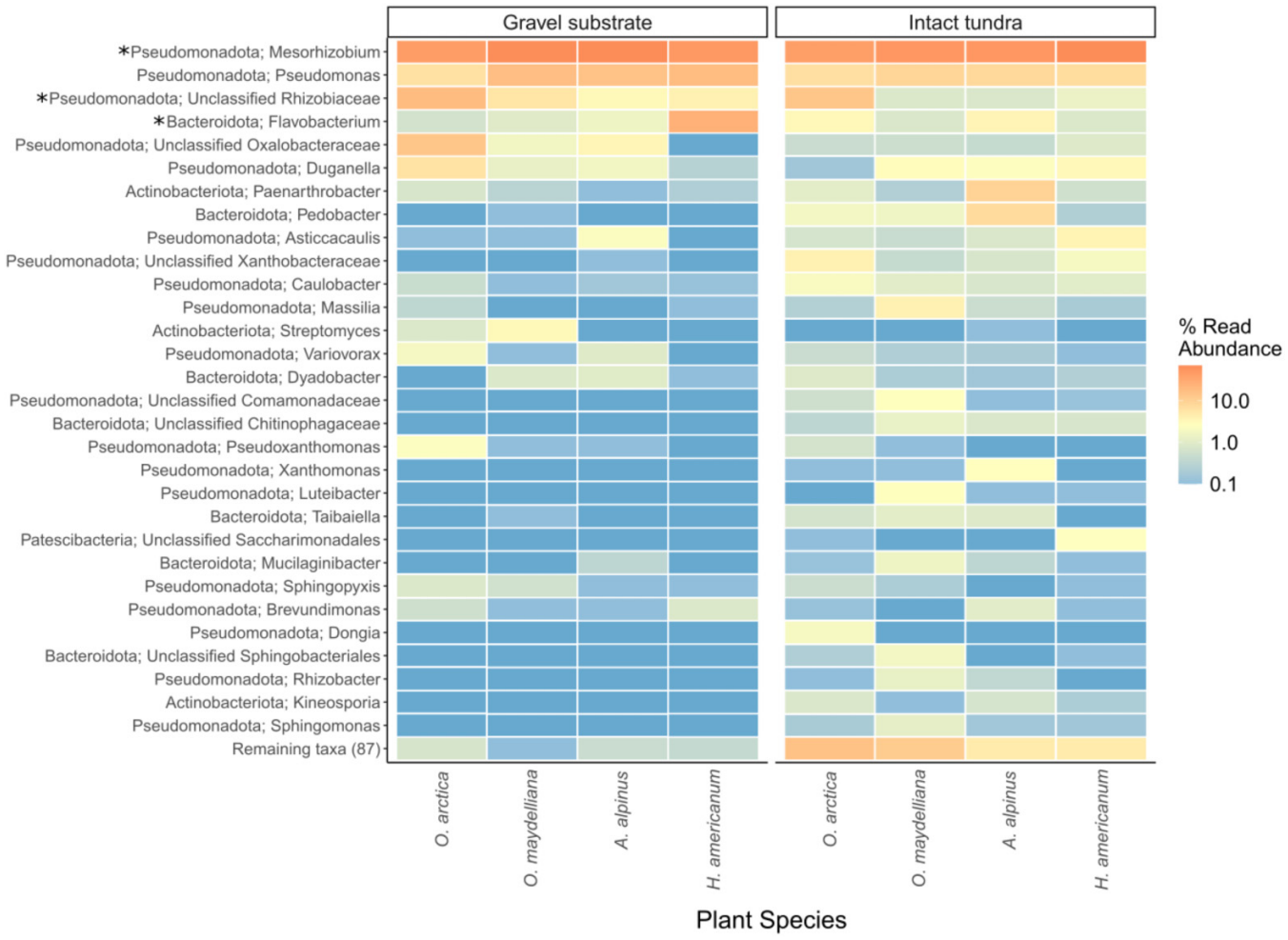
Bioinformatics

16S sequences were classified using a RESCRIPt-processed SILVA reference database trained to the 342F/806R primers (Quast et al. 2013). A *nifH* reference database was created by querying the NCBI database for all classified nitrogenase reductase entries using RESCRIPt (Robeson Ii et al. 2021), which downloads the files and generates a seven-level taxonomy. Sequences with five or more ambiguous bases or with homopolymers that are eight or more base pairs were filtered out. Sequences and taxonomies were dereplicated to remove identical sequences with identical taxonomies before extracting the *nifH*-F/*nifH*-R regions and a final dereplication to generate the *nifH* reference database.

Sequenced 16S and *nifH* reads were separately imported and processed within the QIIME2 2022.8 environment (Bolyen et al. 2019). A total of 1337 unique 16S reads and 1549 unique *nifH* reads were imported. Primers were removed with cutadapt (Martin 2011) before the DADA2 plugin was used to remove low-quality read ends and join the denoised paired-end reads (Callahan et al. 2016). Reads were truncated to a quality score of 30. A minimum quality score of 20 and an overlap of 12 bp is required for DADA2. To remove any sequencing errors, singletons and sequencing artifacts were filtered out, and low-quality samples were filtered out based on a threshold of total read counts. After filtering, 350 16S reads and 313 *nifH* reads remained. Finally, the processed reads were classified to amplicon sequence variants (ASVs) and identified taxonomically using the trained 16S and *nifH* reference databases described above (Bokulich et al. 2018).

The classified sequencing data were extracted from the QIIME2 environment and imported into *R* for statistical analysis (R Development Core Team 2022; version 4.2.2) using the Biological Observation Matrix format (McDonald et al. 2012) and the *biomformat* package (McMurdie and Paulson 2022; version 1.24.0). Abundance data were combined with taxonomy and sample metadata using package *phyloseq* (McMurdie and Holmes 2013; version 1.40.0) before further processing to remove 16S ASVs identified as chloroplasts and mitochondrial rRNA. Contaminant ASVs were identified using *decontam* (Davis et al. 2017; version 1.16.0). Four ASVs were identified from the 16S sequences and removed from analysis. Sequence counts were standardized to account for differences in sequencing depth based on the *Aliivibrio fischeri* internal standard before removing the *Aliivibrio fischeri* sequences (Smets et al. 2016). 309 16S reads and 306 *nifH* reads remained after removal of contaminant sequences.

Fig. 1. Heat map showing the 30 most abundant genera in nodule communities of four Arctic legume species from gravel substrate and intact tundra areas. Nitrogen-fixing genera are marked with an asterisk (∗). Taxa outside the top 30 are grouped at the bottom of the figure. Relative abundances were log10 transformed and are displayed as colours ranging from blue to orange.



The sequencing data were then transformed to capture the relationship between the ASVs in the dataset, as sequencing data are naturally described as proportions (Gloor et al. 2017). Firstly, the large number of zeros typical of sequencing datasets were replaced using zero-imputation with the recommended Geometric Bayesian Multiplicative (GBM) method in the *zCompositions* package (Palarea-Albaladejo and Martin-Fernandez 2015; version 1.4.0-1). For the 16S dataset, the GBM method corrected 2385 of 11 433 values, while for the *nifH* dataset, 6793 of 11 322 values were corrected. Finally, the GBM zero-imputed data were transformed into centred-log ratios using the *CoDaSeq* package (Gloor and Reid 2016; version 0.99.6).

Statistical analysis

Principle component analysis was used to visualize the community composition data. After centred-log ratio transformation, Aitchison distance, which is equivalent to Euclidean distance between samples in centered log-ratio (CLR)-transformed datasets, can be used for clustering (Gloor et al. 2017). To identify statistically significant differences in the nodule communities between the four legume species and substrate types, Euclidean distances were calculated for both 16S and *nifH* datasets using wrapper functions in the package *phyloseq* before running permutational multivariate Analysis of variance (ANOVA) and homogeneity of dispersion tests using functions in the *vegan* package (Oksanen et al. 2022; version 2.6-4). Pairwise comparisons between legume species were performed using package *pairwiseAdonis* (Martinez Arbizu 2020; version 0.4).

Shannon diversity indices were calculated using modified functions adapted from the *microbiomutilities* package (Shetty and Lahti 2023; version 1.00.17). Indices were compared using a two-way ANOVA to identify statistically significant differences in nodule diversity of the legume species or substrate types and species level differences were identified using Tukey honest significant differences (Tukey HSD).

Common ASVs with strong loadings on the ordination axes were identified using species scores and selected for differential abundance analysis. Selection criteria were the 10 ASVs with (1) the most positive or negative loadings along PC 1; and (2) found in at least 10 samples. PC 1 accounted for 46.8% of the variation in the data, while PC 2 was omitted for this analysis as it only accounted for 7.2% of the variation. The *ALDEx2*

Fig. 2. (A) Principle components analysis (PCA) plot of sample scores for nodule samples from four legume species in two environment types using 16S as a marker gene. (B) Species scores for the 309 16S amplicon sequence variants (ASVs) in the PCA ordination. (C) PCA plot of sample scores for nodule samples from four legume species in two environment types using *nifH* as a marker gene. (D) Species scores for the 306 *nifH* ASVs in the PCA ordination. Points in blue represent samples taken from gravel substrate areas and points in orange represent samples taken from intact tundra areas. Samples taken from *O. arctica*, *O. maydelliana*, *A. alpinus*, and *H. alpinum* are represented by squares, circles, triangles, diamonds, respectively. Ellipses represent 95% confidence intervals.

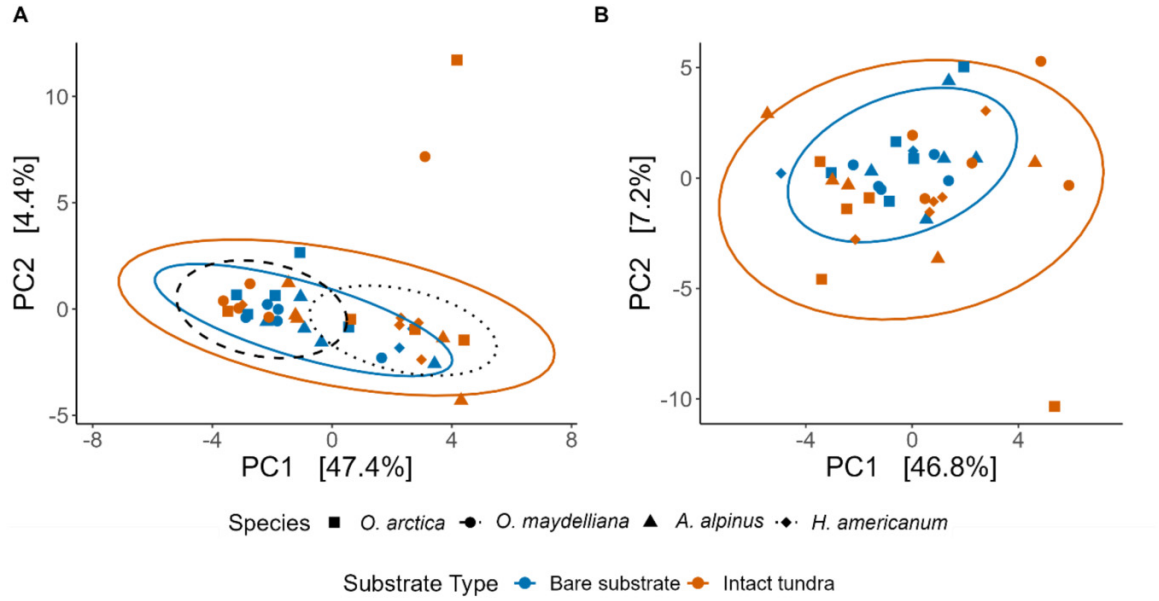


Table 2. Permutational multivariate Analysis of variation analysis testing for differences in 16S and *nifH* nodule community composition and Analysis of variance testing for 16S and *nifH* Shannon diversity differences between legume species and two substrate types.

	16S Composition	16S Diversity	*nifH* Composition	*nifH* Diversity
Species	$F_{3,29} = 0.949$; $p = 0.446$	$F_{3,29} = 0.836$; $p = 0.485$	**$F_{3,29} = 1.889$; $p = 0.045$**	**$F_{3,29} = 4.44$; $p = 0.011$**
Substrate	$F_{1,29} = 1.316$; $p = 0.193$	$F_{1,29} = 0.256$; $p = 0.624$	$F_{1,29} = 1.436$; $p = 0.161$	$F_{1,29} = 0.961$; $p = 0.335$
Species × substrate	$F_{3,29} = 1.555$; $p = 0.109$	**$F_{3,29} = 3.285$; $p = 0.035$**	$F_{3,29} = 1.043$; $p = 0.376$	$F_{3,29} = 1.091$; $p = 0.369$

tool (Fernandes et al. 2013; version 1.28.1) models a probability distribution and estimates effect size differences to identify significant differential abundances between groups. We used *ALDEx2* to identify differential abundances between substrates using a Welch's *t* test with Benjamini–Hochberg procedure adjusted *p*-values. Differential abundances of these ASVs between legume species were identified using a two-way ANOVA and Tukey HSD.

Results

Nodule taxonomic characterization

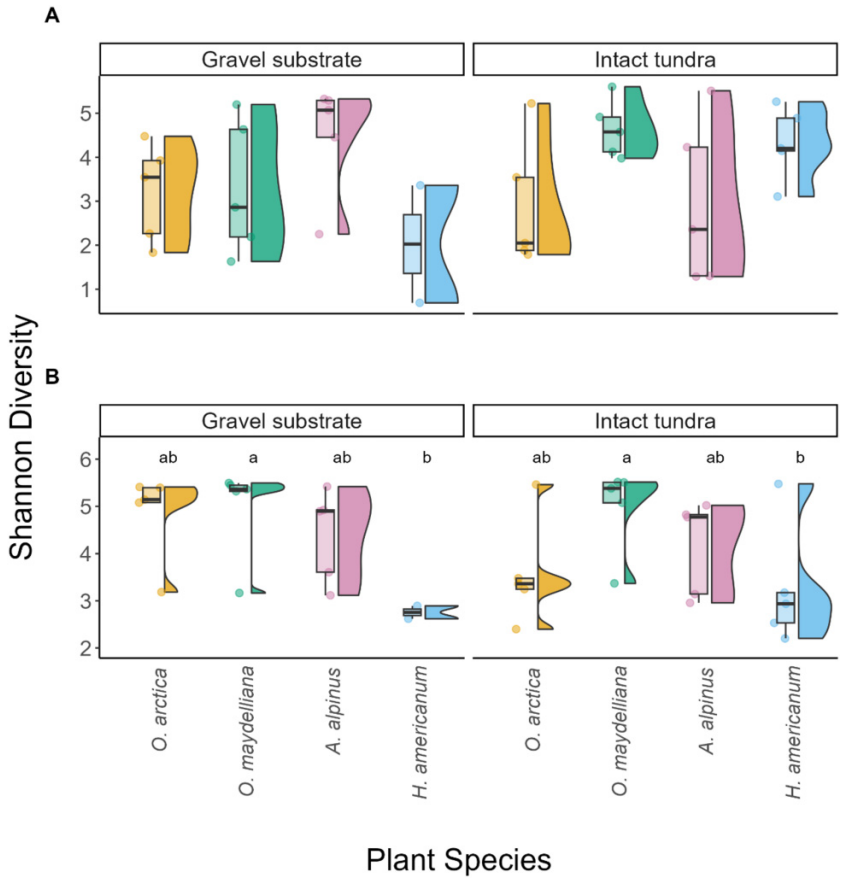
A total number of 23 417 575 reads obtained across 37 samples, ranging from 68 429 to 155 817 per sample, and classifying to 350 16S ASVs. After standardizing and filtering contaminants, 309 ASVs remained. Pseudomonadota (58.6%), Bacteroidota (22.3%), and Actinobacteriota (5.5%) were the most abundant among the 11 detected phyla. Bacteria classes *Alphaproteobacteria* and *Gammaproteobacteria* encompassed most of the 16S bacterial diversity at 29.8% and 28.8%, respectively. These ASVs were assigned to 78 distinct genera, with 112 ASVs (36.2%) remaining unresolved at the genus level. Most ASVs were classified to *Pseudomonas* (4.5%), *Pedobacter* (3.6%), *Mucilaginibacter* (2.9%), *Flavobacterium* (2.6%), *Massilia* (2.6%), and *Mesorhizobium* (2.6%). Notably, *Mesorhizobium* had the highest relative read abundances across all legume species and substrate types, followed by *Pseudomonas.* In intact tundra substrates, abundances were distributed across more taxa, whereas gravel substrates were dominated by a few genera (Fig. 1).

The total number of *nifH* reads obtained across 37 samples was 3505 639 reads, ranging from 14 917 to 145 585 per sample, and classifying to 313 *nifH* ASVs. After filtering contaminants, 306 *nifH* ASVs remained.The vast majority of the identified *nifH* reads classified to phylum Pseudomonadota (98%) and, within the Pseudomonadota, to genus *Rhizobium* (97.1%).

16S composition and diversity of nodule bacterial communities

Nodule 16S bacterial communities were not significantly different between either substrate type ($p = 0.194$) or among legume species ($p = 0.444$) (Fig. 2; Table 2). The 95%

Fig. 3. Boxplots and violin plots of (A) 16S and (B) *nifH* Shannon diversity in nodule communities of four Arctic legume species—*O. arctica*, *O. maydelliana*, *A. alpinus*, and *H. americanum* from gravel substrate and intact tundra areas. Boxes encompass 25%–75% percentiles of the data. The median is indicated by the horizontal black line, with each nodule sample represented by dots.



confidence intervals of the gravel substrate and intact tundra communities overlap completely, and the gravel substrate communities seem to be subset within the broader intact tundra communities. There was also no difference in Shannon diversity between substrate types ($p = 0.623$) or legume species ($p = 0.485$), although there was a significant interaction between species and substrate ($p = 0.035$) (Fig. 3; Table 2) likely due to the difference in response by *H. americanum* nodule communities between substrates.

Contribution of 16S ASVs with high loadings to nodule community structure

Of the 309 16S rRNA ASVs, the 10 with the most positive species scores on PC1 axis classified to *Alphaproteobacteria* (6), *Gammaproteobacteria* (2), and *Bacteroidia* (2). The 10 with the most negative species scores on PC1 classified to *Alphaproteobacteria* (4), *Bacteroidia* (2), and *Acidimicrobiia*, *Saccharimonadia*, *Phycisphaerae*, and *Gammaproteobacteria* at 1 each. Six of the above ASVs were found in more than 10 samples. These ASVs were classified as bacterial genera *Mesorhizium* (2), *Pseudomonas* (2), *Asticcacaulis* (1), and *Caulobacter* (1). There were five other ASVs that did not contribute greatly to the differentiation of communities along the ordination axes but were found in more than 10 samples. Differential abundance analysis using ANOVA-like differential expression (ALDEx) of these 11 ASVs showed that although the two substrates had nodule communities that overlapped in composition, the selected ASVs were differentially abundant between the two substrates (Fig. 4). In contrast to the prevailing trend, we saw no difference in abundance of the six ASVs with high loadings compared to the ASVs with smaller scores (Table 3). A two-way ANOVA also did not show differences in abundance of the six ASVs with higher species scores between the four legume species (Fig. 5; Table 3).

nifH composition and diversity of nodule bacterial communities

nifH gene sequences showed a similar pattern to the 16S rRNA sequences with no significant differences in composition between the N-fixing nodule bacterial communities of either substrate type (Fig. 2; Table 2). There was a significant difference between the *nifH* N-fixing bacterial community composition of *O. maydelliana* and *H. americanum* (Table 3; Table A1) and *nifH* Shannon diversity was different between these two species (Fig. 3; Table 2; Table A2).

Fig. 4. Analysis of variance-like differential expression plot of the most abundant (A) 16S and (B) *nifH* amplicon sequence variants (ASVs). *X* axis values are median centred-log ratio transformed abundances and *y*-axis values are per feature median centred-log ratio transformed abundance differences. Differences significantly different from zero are in red while those not significant are in black. Triangles are ASVs negatively loaded on PC1 in the ordinations in Fig. A2, and diamonds are positively loaded ASVs. Squares are ASVs that did not load strongly on PC1. Positive *y*-axis values indicate higher abundance in intact tundra and negative higher abundance in gravel substrates.

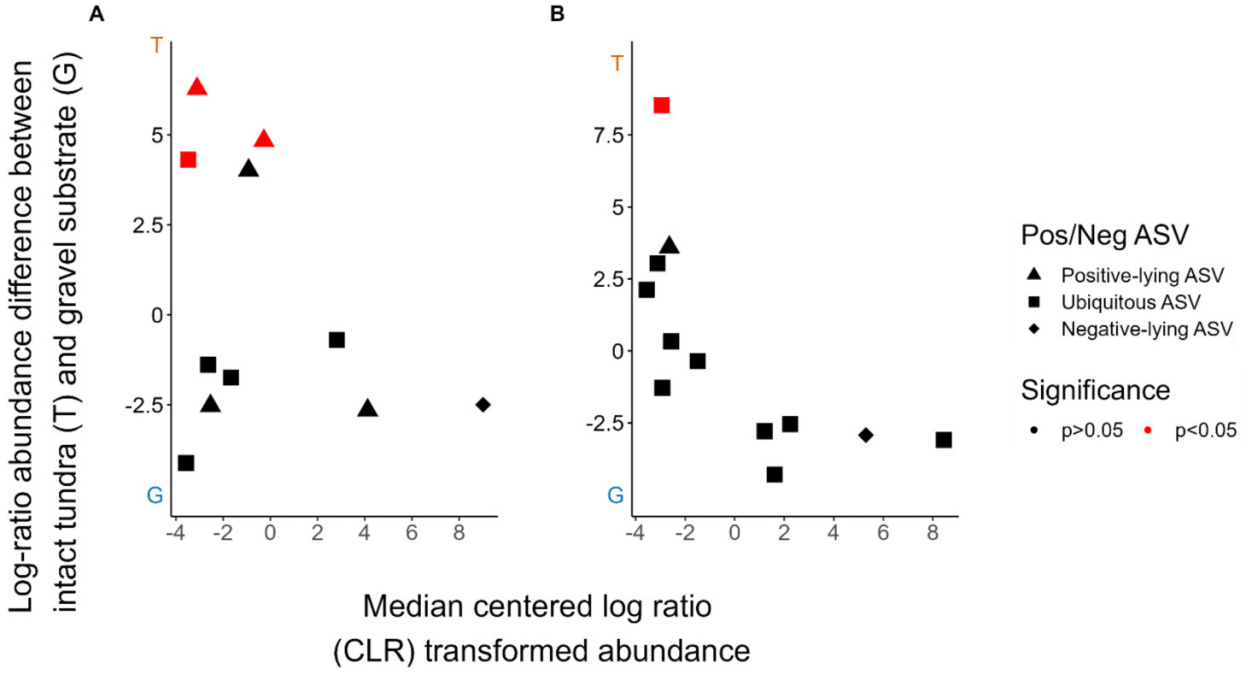


Table 3. Results from two-way ANOVA analyses on the centred-log ratio transformed abundances of the six 16S amplicon sequence variants (ASVs) and the two *nifH* ASVs that contributed greatly to their respective PC1 axis (Fig. 1; Fig. A2).

ASV	Species	Substrate	Species × substrate
16S Asticcacaulis1 (ae41a7e7ad08cd926075f8ab35eaa0eb)	$F_{3,29} = 0.725$; $p = 0.545$	$F_{1,29} = 2.137$; $p = 0.155$	$F_{3,29} = 1.833$; $p = 0.163$
16S Caulobacter1 (6d1de170c8d9c378c1bacf72d93283c2)	$F_{3,29} = 0.479$; $p = 0.700$	$F_{1,29} = 2.976$; $p = 0.095$	$F_{3,29} = 1.325$; $p = 0.285$
16S Mesorhizobium1 (e010f6ec05db89e11d623899272ecde1)	$F_{3,29} = 0.890$; $p = 0.458$	$F_{1,29} = 0.838$; $p = 0.368$	**$F_{3,29} = 4.677$; $p = 0.009$**
16S Mesorhizobium2 (7d850db295ca96fe18f859c33034a116)	$F_{3,29} = 0.786$; $p = 0.512$	$F_{1,29} = 1.723$; $p = 0.200$	$F_{3,29} = 0.894$; $p = 0.456$
16S Pseudomonas1 (3aa080fdc0ba7fe7aa9153d11c5fb3fa)	$F_{3,29} = 0.877$; $p = 0.465$	$F_{1,29} = 0.668$; $p = 0.420$	$F_{3,29} = 0.039$; $p = 0.990$
16S Pseudomonas2 (0f6640911d66a8b028b2b091f1cbf0c0)	$F_{3,29} = 0.689$; $p = 0.566$	$F_{1,29} = 4.137$; $p = 0.051$	$F_{3,29} = 0.575$; $p = 0.636$
nifH Mesorhizobium-1 (88ad6d4de9880fdeb52ea15e8f11c0e0)	**$F_{3,29} = 3.779$; $p = 0.021$**	$F_{1,29} = 1.842$; $p = 0.185$	$F_{3,29} = 0.863$; $p = 0.471$
nifH Mesorhizobium-2 (439a5c6fc811c01591732533d407d094)	$F_{3,29} = 2.118$; $p = 0.120$	$F_{1,29} = 0.033$; $p = 0.857$	$F_{3,29} = 0.558$; $p = 0.647$

Note: ASVs that were present in more than 10 samples and were also within the top 10 negative or top 10 positive loadings on PC1 were chosen for analysis. ANOVA, Analysis of variance.

Contribution of *nifH* ASVs with high loadings to nodule community structure

Of the 306 *nifH* ASVs, the 10 with the highest species scores on PC1 axis classified to *Mesorhizobium* (7) and *Rhizobium* (2) with one ASV only able to be classified as bacteria. The 10 with the most negative species scores on PC1 classified to *Mesorhizobium* (1) and *Rhizobium* (9). Two of the above ASVs were found in more than 10 samples. Both the ASVs were classified as bacterial genus *Mesorhizobium*. There were 10 other ASVs that did not contribute to the differentiation of communities along the ordination axes but were found in more than 10 samples. Differential abundance analysis using *ALDEx* of these 12 ASVs showed no significant difference in the relative abundance of ASVs between the two substrate types (Fig. 4; Table 3), similar to the *nifH* nodule community composition analysis. When comparing the two ASVs that were found in more than 10 samples, there was a significant difference in the relative abundance of these ASVs between *O. maydelliana* and *H. americanum* and between *Astragalus alpinus* and *H. americanum* (Fig. 6; Tables 3; A3; A4).

Discussion

Despite the near absence of vascular plants and substantially lower organic C in the gravel substrates compared to nearby tundra (Hnatowich et al. 2023), there were no significant differences in the composition or diversity of either the

Fig. 5. Boxplots of the six 16S amplicon sequence variants (ASVs) with highest species scores on PC1 in Fig. 1 with the centred-log ratio transformed abundance on the *y*-axis and the four legume species on the *x*-axis. ASVs that were present in more than 10 samples and were also within the top 10 negative and top 10 positive loadings in PC1 were chosen for analysis. Boxplots are coloured by substrate type and separated by ASVs codes. Boxes represent 25%–75% percentiles of the data and points represent samples that lie beyond 1.5x inter-quartile range.

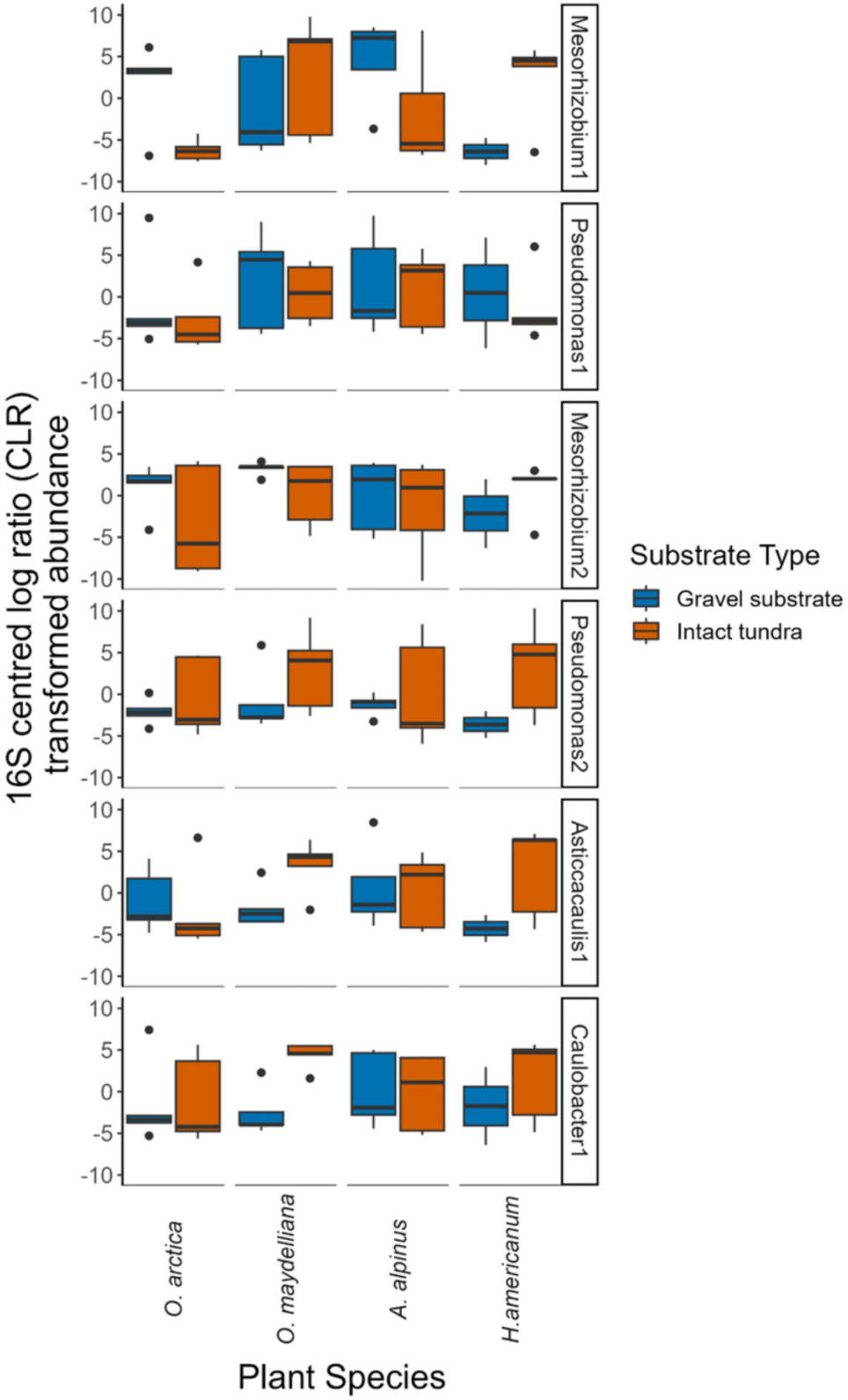
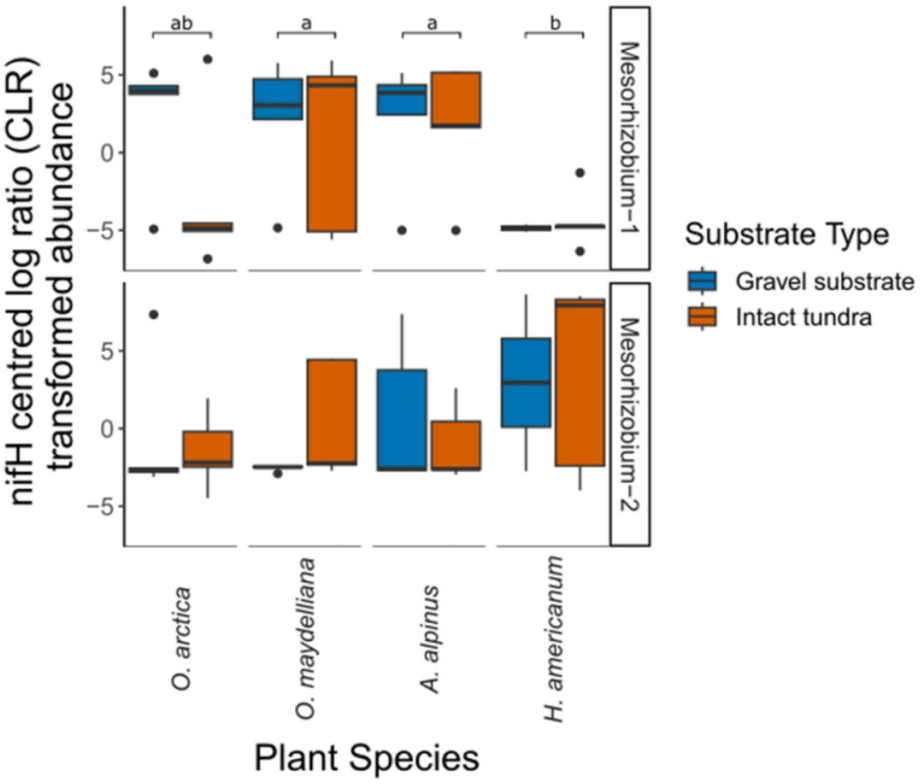


Fig. 6. Boxplots of the two *nifH* amplicon sequence variants (ASVs) with the highest species scores on PC1 in Fig. 4 with the centred-log ratio transformed abundances on the *y*-axis and the four legume species on the *x*-axis. ASVs that were present in more than 10 samples and were also within the top 10 negative and top 10 positive loadings in PC1 were chosen for analysis. Boxplots are coloured by substrate types and separated by ASVs codes. Boxes represent 25%–75% percentiles of the data and points represent samples that lie beyond 1.5∗ inter-quartile range



overall bacterial community or the potentially N-fixing bacterial community among plant nodules collected from these distinct substrates (Fig. 1; Fig. A3). Given that nodule communities are a subset of bulk and rhizosphere communities, this result is unexpected. Highly disturbed gravel substrates with limited soil development or vegetation typically exhibit lower abundance of soil microflora and distinct microbial communities (Filcheva et al. 2000; Hamidović et al. 2020). Additionally, studies in similar northern areas suggest that N cycling is influenced by topography (Stewart et al. 2014). While the high degree of disturbance might be expected to reduce populations of potential rhizobial innocula, there is evidence of successful nodulation in highly disturbed environments.

For example, both *Lotus corniculatus* and *Trifolium arvense* nodulated in early sandy overburden soils from a reclaimed mine, suggesting that the innocula necessary for nodulation can be found even in highly disturbed substrates (Boldt-Burisch et al. 2015). Rhizobia have been demonstrated to persist in substrates without host plants for up to 5 years (Hirsch 1996; Denison and Kiers 2011) and likely for much longer. Therefore, it is plausible that even small amounts of inoculum in soil may produce successful nodulation. There is also evidence that vertical transmission of N-fixing bacteria by seed is possible (Rout et al. 2013; Jooste et al. 2019; Ai et al. 2023). Whether this is a plausible mechanism of inoculation in our study is not clear because such a mechanism has not been documented in legumes (Frank et al. 2017), and a previous greenhouse study reported that *Hedysarum alpinus* and *Oxytropis campestris* grown from seed did not form nodules in the absence of inoculum treatments (Stewart and Siciliano 2015).

We found relatively few differences in the nodule communities between the four legume species. There were no differences in nodule communities when analyzed using 16S sequencing data. Some diversity differences were identified between *H. americanum* and *O. maydelliana* nodules when analyzed using *nifH* sequencing data. This finding is interesting considering that we explored a diverse group of four species spread across three genera. *Astragalus* and *Oxytropis* are closely-related sister taxa within the Fabaceae, diverging an estimated 14.8 million years ago, and the clades with the descendants of the most recent common ancestors of *Hedysarum*, *Astragalus,* and *Oxytropis* diverged 33 million years ago. There is a reasonable expectation that species-driven differences in nodule communities should be present between two genera that diverged 14.8 million years ago (Lavin et al. 2005).

However, some studies have shown that small genetic differences can be overridden by environment to alter the nodule microbiome. For example, there was no genotype effect of two varieties of *Vigna unguiculata* (cowpea) but a strong soil type effect on the composition of the nodule microbiome

(Leite et al. 2017). The compositions of inocula in the soils at our sampling sites may be similar and be able to override the slight genetic differences in our legume species, resulting in the nodule community composition that we see here. Studies in comparable environments such as alpine regions suggest that the relationships between bacterial species and host species may be less influenced by biotic interactions (King et al. 2012; Pang et al. 2021). Many recent studies have demonstrated that soil pH plays a strong role in influencing bacterial communities in polar regions (Chu et al. 2011; Malard et al. 2019, 2021; Ji et al. 2022; Amores-Arrocha et al. 2023). Further studies may be required to compare the resident soil microorganisms at each collection site and correlate soil abiotic factors.

Our expectation was that *nifH* sequences would be conserved in N-fixers (Rösch et al. 2002). We largely observed this; however, our ability to detect differences may have been hindered by limitations in our database. Of the 306 total *nifH* sequences, 41 remained classified as "unknown bacteria" or simply proteobacteria, and of the 255 sequences that were classified to genera, all but two were identified as *Rhizobium* or *Mesorhizobium*. Given the limited research on microbial genomes in the Arctic (Edwards et al. 2020), future studies may benefit from whole-genome sequencing approaches to investigate taxa that may not be effectively captured by culturing or amplicon-based methods.

Overall, our study demonstrated that early colonizing native Arctic legumes on gravel quarries were able to acquire N-fixing nodule communities through natural colonization, with nodule community compositions resembling those found in the same plant species in adjacent natural environments. This highlights that, despite being highly disturbed, these quarries show signs of ecosystem recovery and warrant further research to compare ecosystem function similar early successional environments resulting from natural disturbance such as exposed gravel slopes near rivers. Additionally, this finding is significant for restoration practitioners, as it suggests that seeding or transplanting native legumes in disturbed Arctic sites can lead to associations with locally adapted nodule communities without the need for artificial inoculum treatments. As is the case for most Arctic forbs (Forbes and Jefferies 1999; Hagen 2002), there are currently no commercial seed sources for these species. The combination of early successional habitat preferences and natural nodulation demonstrate that *Astragalus alpinus* and *0. arctica* are good candidates for seed production as their use would not rely on commercial inoculum treatments. Integrating these native legumes with turf transplant methods (Hnatowich et al. 2022, 2023), may enhance the turf restoration efforts by increasing nutrient supplies to turf plants. Finally, since these legumes are initial colonizers, direct seeding or seedling transplantation may be an effective strategy given the challenges in scaling up turf transplantation for larger areas. By capitalizing on the natural ability of early colonizing native legumes to acquire N-fixing endophytes and stimulate ecosystem processes through N-rich plant litter deposition, restoration practitioners can enhance the success and sustainability of their restoration projects in Arctic ecosystems.

Acknowledgements

We are grateful to Ian Hnatowich for assistance in the field. Agnico Eagle Mines Ltd. Meliadine operation provided logistical support.

Article information

History dates
Received: 13 September 2024
Accepted: 6 December 2024
Accepted manuscript online: 16 December 2024
Version of record online: 20 February 2025

Data availability
The raw sequences were deposited to the sequence read archive (SRA) repository of the National Center for Biotechnology Information (BioProject PRJNA1191159, Accessions: SRR31521740–SRR31521778; SRR31522397–SRR31522435).

Author information

Author ORCIDs
Katherine M. Standen https://orcid.org/0000-0003-0370-2027
Eric G. Lamb https://orcid.org/0000-0001-5201-4541

Author notes
Present address for Katherine M. Standen is Ministry of Parks, Culture and Sport, Government of Saskatchewan. Saskatoon, SK. S7P 0B6.

Author contributions
Conceptualization: KJS, EGL
Formal analysis: ASL, KMS
Funding acquisition: KJS
Investigation: ASL, KJS, EGL
Supervision: EGL
Writing – original draft: ASL
Writing – review & editing: KMS, KJS, EGL

Competing interests
The authors declare there are no competing interests.

Funding information
National Science and Engineering Research Council (NSERC) Collaborative Research and Development grant (CRDPJ 518281-17) in collaboration with Agnico Eagle Mines Ltd.

References

Ai, J., Yu, T., Liu, X., Jiang, Y., Wang, E., and Deng, Z.-S. 2023. Seed associated microbiota and vertical transmission of bacterial communi-

ties from seed to nodule in *Sophora davidii*. Plant Soil, **491**: 285–302. doi:10.1007/s11104-023-06115-2.

Amores-Arrocha, H., Asamoah-Asare, A.K.B., Opio, J., Martin, A., Cuthbertson, L., Bradford, H.R., et al. 2023. Analysis of bacterial communities around the Adventdalen landfill site in Svalbard. Microorganisms, **11**: 1093. doi:10.3390/microorganisms11041093. PMID: 37110516.

Bardgett, R.D., and Walker, L.R. 2004. Impact of coloniser plant species on the development of decomposer microbial communities following deglaciation. Soil Biol. Biochem. **36**: 555–559. doi:10.1016/j.soilbio.2003.11.002.

Birnbaum, C., Bissett, A., Teste, F.P.T., and Laliberté, E. 2018. Symbiotic N2-fixer community composition, but not diversity, shifts in nodules of a single host legume across a 2-million-year dune chronosequence. Microb. Ecol. **76**: 1009–1020. doi:10.1007/s00248-018-1185-1. PMID: 29663039.

Bokulich, N.A., Kaehler, B.D., Rideout, J.R., Dillon, M., Bolyen, E., Knight, R., et al. 2018. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome, **6**: 1–17. doi:10.1186/s40168-018-0470-z. PMID: 29291746.

Boldt-Burisch, K., Naeth, M.A., Schneider, B.U., and Hüttl, R.F. 2015. Linkage between root systems of three pioneer plant species and soil nitrogen during early reclamation of a mine site in Lusatia, Germany. Restor. Ecol. **23**: 357–365. doi:10.1111/rec.12190.

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., et al. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat. Biotechnol. **37**: 852–857. doi:10.1038/s41587-019-0209-9. PMID: 31341288.

Bradshaw, A.D. 1993. Restoration ecology as a science. Restor. Ecol. **1**: 71–73. doi:10.1111/j.1526-100X.1993.tb00011.x.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. Nat. Methods, **13**: 581. doi:10.1038/nmeth.3869. PMID: 27214047.

Chu, H., Fierer, N., Lauber, C.L., Caporaso, J.G., Knight, R., and Grogan, P. 2011. Soil bacterial diversity in the Arctic is not fundamentally different from that found in other biomes. Environ. Microbiol. **12**: 2998–3006. doi:10.1111/j.1462-2920.2010.02277.x.

Davis, N.M., Proctor, D., Holmes, S.P., Relman, D.A., and Callahan, B.J. 2018. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome **6**: 226. doi:10.1186/s40168-018-0605-2.

Denison, R.F., and Kiers, E.T. 2011. Life histories of symbiotic rhizobia and mycorrhizal fungi. Curr. Biol. **21**: R775–R785. doi:10.1016/j.cub.2011.06.018. PMID: 21959168.

Edwards, A., Cameron, K.A., Cook, J.M., Debbonaire, A.R., Furness, E., Hay, M.C., and Rassner, S.M.E. 2020. Microbial genomics amidst the arctic crisis. Microb. Genomics, **6**: 1–20. doi:10.1099/mgen.0.000375.

Fernandes, A., Macklaim, J., Linn, T., Reid, G., and Gloor, G. 2013. ANOVA-like differential expression (ALDEx) analysis for mixed population RNA-seq. PLoS One, **8**: e67019. doi:10.1371/journal.pone.0067019. PMID: 23843979.

Filcheva, E., Noustorova, M., Gentcheva-Kostadinova, S., and Haigh, M.J. 2000. Organic accumulation and microbial action in surface coal-mine spoils, Pernik, Bulgaria. Ecol. Eng. **15**: 1–15. doi:10.1016/S0925-8574(99)00008-7.

Flora of North America Editorial, C. 1993. Flora of North America north of Mexico.

Forbes, B.C., and Jefferies, R.L. 1999. Revegetation of disturbed arctic sites: constraints and applications. Biol. Conserv. **88**: 15–24. doi:10.1016/S0006-3207(98)00095-0.

Frank, A., Saldierna Guzmán, J., and Shay, J. 2017. Transmission of bacterial endophytes. Microorganisms, **5**: 70. doi:10.3390/microorganisms5040070. PMID: 29125552.

Gloor, G.B., and Reid, G. 2016. Compositional analysis: a valid approach to analyze microbiome high-throughput sequencing data. Can. J. Microbiol. **62**: 692–703. doi:10.1139/cjm-2015-0821. PMID: 27314511.

Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., and Egozcue, J.J. 2017. Microbiome datasets are compositional: and this is not optional. Front. Microbiol. **8**: 1–6. doi:10.3389/fmicb.2017.02224. PMID: 28197127.

Hagen, D. 2002. Propagation of native Arctic and alpine species with a restoration potential. Polar Res. **21**: 37–47. doi:10.3402/polar.v21i1.6472.

Hagen, D., Hansen, T.I., Graae, B.J., and Rydgren, K. 2014. To seed or not to seed in alpine restoration: introduced grass species outcompete rather than facilitate native species. Ecol. Eng. **64**: 255–261. doi:10.1016/j.ecoleng.2013.12.030.

Hakim, S., Mirza, B.S., Zaheer, A., McLean, J.E., Imran, A., Yasmin, S., and Sajjad Mirza, M. 2018. Retrieved 16S rRNA and nifH sequences reveal co-dominance of bradyrhizobium and Ensifer (Sinorhizobium) strains in field-collected root nodules of the promiscuous host *Vigna radiata* (L.) R. Wilczek. Appl. Microbiol. Biotechnol. **102**: 485–497. doi:10.1007/s00253-017-8609-6. PMID: 29110071.

Hamidović, S., Cvijović, G.G., Waisi, H., Životić, L., Šoja, S.J., Raičević, V., and Lalević, B. 2020. Response of microbial community composition in soils affected by coal mine exploitation. Environ. Monit. Assess. **192**. doi:10.1007/s10661-020-08305-2.

Hirsch, P.R. 1996. Population dynamics of indigenous and genetically modified rhizobia in the field. New Phytol. **133**: 159–171. doi:10.1111/j.1469-8137.1996.tb04351.x.

Hnatowich, I.G., Lamb, E.G., and Stewart, K.J. 2022. Vegetative growth and belowground expansion from transplanted low-arctic tundra turfs. Restor. Ecol. **31**: e13716. doi:10.1111/rec.13716.

Hnatowich, I.G., Lamb, E.G., and Stewart, K.J. 2023. Reintroducing vascular and non-vascular plants to disturbed Arctic sites: investigating turfs and turf fragments. Ecol. Restor. **41**: 3–15. doi:10.3368/er.41.1.3.

Hodkinson, I.D., Coulson, S.J., and Webb, N.R. 2003. Community assembly along proglacial chronosequences in the high Arctic: vegetation and soil development in north-west Svalbard. J. Ecol. **91**: 651–663. doi:10.1046/j.1365-2745.2003.00786.x.

Iversen, C.M., Sloan, V.L., Sullivan, P.F., Euskirchen, E.S., McGuire, A.D., Norby, R.J., et al. 2015. The unseen iceberg: plant roots in arctic tundra. New Phytol. **205**: 34–58. doi:10.1111/nph.13003.

Ji, M., Kong, W., Jia, H., Delgado-Baquerizo, M., Zhou, T., Liu, X., et al. 2022. Polar soils exhibit distinct patterns in microbial diversity and dominant phylotypes. Soil Biol. Biochem. **166**: 108550. doi:10.1016/j.soilbio.2022.108550.

Jooste, M., Roets, F., Midgley, G.F., Oberlander, K.C., and Dreyer, L.L. 2019. Nitrogen-fixing bacteria and Oxalis—evidence for a vertically inherited bacterial symbiosis. BMC Plant Biol. **19**: 441. doi:10.1186/s12870-019-2049-7.

Kearns, N.B., Jean, M., Tissier, E.J., and Johnstone, J.F. 2015. Recovery of tundra vegetation three decades after hydrocarbon drilling with and without seeding of non-native grasses. Arctic, **68**: 16–31. doi:10.14430/arctic4445.

King, A.J., Farrer, E.C., Suding, K.N., and Schmidt, S.K. 2012. Co-occurrence patterns of plants and soil bacteria in the high-alpine subnival zone track environmental harshness. Front. Microbiol. **3**. doi:10.3389/fmicb.2012.00347.

Knelman, J.E., Legg, T.M., O'Neill, S.P., Washenberger, C.L., González, A., Cleveland, C.C., and Nemergut, D.R. 2012. Bacterial community structure and function change in association with colonizer plants during early primary succession in a glacier forefield. Soil Biol. Biochem. **46**: 172–180. doi:10.1016/j.soilbio.2011.12.001.

Lavin, M., Herendeen, P.S., and Wojciechowski, M.F. 2005. Evolutionary rates analysis of leguminosae implicates a rapid diversification of lineages during the tertiary. Syst. Biol. **54**: 575–594. doi:10.1080/10635150590947131.

Leite, J., Fischer, D., Rouws, L.F.M., Fernandes-Júnior, P.I., Hofmann, A., Kublik, S., et al. 2017. Cowpea nodules harbor non-rhizobial bacterial communities that are shaped by soil type rather than plant genotype. Front. Plant Sci. **7**: 1–11. doi:10.3389/fpls.2016.02064.

Malard, L.A., Anwar, M.Z., Jacobsen, C.S., and Pearce, D.A. 2019. Biogeographical patterns in soil bacterial communities across the Arctic region. FEMS Microbiol. Ecol. **95**: fiz128. doi:10.1093/femsec/fiz128.

Malard, L.A., Anwar, M.Z., Jacobsen, C.S., and Pearce, D.A. 2021. Influence of spatial scale on structure of soil bacterial communities across an Arctic landscape. Appl. Environ. Microbiol. **87**: e02220–e02220. doi:10.1128/AEM.02220-20.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. **17**: 10. doi:10.14806/ej.17.1.200.

Martinez Arbizu, P. 2020. pairwiseAdonis: Pairwise Multilevel Comparison using Adonis. R Package.
Masson-Boivin, C., Giraud, E., Perret, X., and Batut, J. 2009. Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium recipes? Trends Microbiol. **17**: 458–466. doi:10.1016/j.tim.2009.07.004.
McDonald, D., Clemente, J.C., Kuczynski, J., Rideout, J.R., Stombaugh, J., Wendel, D., et al. 2012. The Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome. GigaScience, **1**: 7. doi:10.1186/2047-217X-1-7.
McMurdie, P.J., and Holmes, S. 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One, **8**: e61217. doi:10.1371/journal.pone.0061217.
McMurdie, P.J., and Paulson, J.N. 2022. biomformat: An interface package for the BIOM file format. R/Bioconductor Package.
Mehlhoop, A.C., Evju, M., and Hagen, D. 2018. Transplanting turfs to facilitate recovery in a low-alpine environment—what matters? Appl. Veg. Sci. **21**: 615–625. doi:10.1111/avsc.12398.
Miller, V.S., Naeth, M.A., and Wilkinson, S.R. 2021. Micro topography, organic amendments and an erosion control product for reclamation of waste materials at an arctic diamond mine. Ecol. Eng. **172**: 106399–106399. doi:10.1016/j.ecoleng.2021.106399.
Mori, H., Maruyama, F., Kato, H., Toyoda, A., Dozono, A., Ohtsubo, Y., et al. 2014. Design and experimental application of a novel non-degenerate universal primer set that amplifies prokaryotic 16S rRNA genes with a low possibility to amplify eukaryotic rRNA genes. DNA Res. **21**: 217–227. doi:10.1093/dnares/dst052.
Myrold, D.D., and Huss-Danell, K. 2003. Alder and lupine enhance nitrogen cycling in a degraded forest soil in Northern Sweden. Plant Soil, **254**: 47–56. doi:10.1023/A:1024951115548.
Naeth, M.A., and Wilkinson, S.R. 2014. Establishment of restoration trajectories for upland tundra communities on diamond mine wastes in the Canadian arctic. Restor. Ecol. **22**: 534–543. doi:10.1111/rec.12106.
Oksanen, J., Simpson, G., Blanchet, F., Kindt, R., Legendre, P., Minchin, P., et al. 2022. vegan: community ecology package v.2.6-4.
Oldroyd, G.E.D., Murray, J.D., Poole, P.S., and Downie, J.A. 2011. The rules of engagement in the legume-rhizobial symbiosis. Annu. Rev. Genet. **45**: 119–144. doi:10.1146/annurev-genet-110410-132549.
Palarea-Albaladejo, J., and Martin-Fernandez, J.A. 2015. zCompositions—R package for multivariate imputation of left-censored data under a compositional approach. Chemom. Intell. Lab. Syst. **143**: 85–96. doi:10.1016/j.chemolab.2015.02.019.
Pang, J., Palmer, M., Sun, H.J., Seymour, C.O., Zhang, L., Hedlund, B.P., and Zeng, F. 2021. Diversity of root nodule-associated bacteria of diverse legumes along an elevation gradient in the Kunlun mountains, China. Front. Microbiol. **12**: 633141. doi:10.3389/fmicb.2021.633141.
Porsild, A.E., and Cody, W.J. 1980. Vascular plants of continental Northwest Territories, Canada, Ottawa.
Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic. Acids. Res. **41**: 590–596. doi:10.1093/nar/gks1219.
R Development Core Team. 2022. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Robeson II, M.S., O'Rourke, D.R., Kaehler, B.D., Ziemski, M., Dillon, M.R., Foster, J.T., and Bokulich, N.A. 2021. RESCRIPt: reproducible sequence taxonomy reference database management. PLoS Comp. Biol. **17**: e1009581. doi:10.1371/journal.pcbi.1009581.
Rösch, C., Mergel, A., and Bothe, H. 2002. Biodiversity of denitrifying and dinitrogen-fixing bacteria in an acid forest soil. Appl. Environ. Microbiol. **68**: 3818–3829. doi:10.1128/AEM.68.8.3818–3829.2002.
Rout, M.E., Chrzanowski, T.H., Westlie, T.K., DeLuca, T.H., Callaway, R.M., and Holben, W.E. 2013. Bacterial endophytes enhance competition by invasive plants. Am. J. Bot. **100**: 1726–1737. doi:10.3732/ajb.1200577.
Rydgren, K., Auestad, I., Hamre, L.N., Hagen, D., Rosef, L., and Skjerdal, G. 2016. Long-term persistence of seeded grass species: an unwanted side effect of ecological restoration. Environ. Sci. Pollut. Res. **23**: 13591–13597. doi:10.1007/s11356-015-4161-z.
Shetty, S., and Lahti, L. 2023. microbiomeutilities: Utilities for Microbiome Analytics. R package.
Simonsen, A.K., Dinnage, R., Barrett, L.G., Prober, S.M., and Thrall, P.H. 2017. Symbiosis limits establishment of legumes outside their native range at a global scale. Nat. Commun. **8**: 1–9. doi:10.1038/ncomms14790.
Smets, W., Leff, J.W., Bradford, M.A., McCulley, R.L., Lebeer, S., and Fierer, N. 2016. A method for simultaneous measurement of soil bacterial abundances and community composition via 16S rRNA gene sequencing. Soil Biol. Biochem. **96**: 145–151. doi:10.1016/j.soilbio.2016.02.003.
Stewart, K.J., and Siciliano, S.D. 2015. Potential contribution of native herbs and biological soil crusts to restoration of the biogeochemical nitrogen cycle in mining impacted sites in Northern Canada. Ecol. Restor. **33**: 30–42. doi:10.3368/er.33.1.30.
Stewart, K.J., Grogan, P., Coxson, D.S., and Siciliano, S.D. 2014. Topography as a key factor driving atmospheric nitrogen exchanges in arctic terrestrial ecosystems. Soil Biol. Biochem. **70**: 96–112. doi:10.1016/j.soilbio.2013.12.005.
Van Cauwenberghe, J., Verstraete, B., Lemaire, B., Lievens, B., Michiels, J., and Honnay, O. 2014. Population structure of root nodulating *Rhizobium leguminosarum* in *Vicia cracca* populations at local to regional geographic scales. Syst. Appl. Microbiol. **37**: 613–621. doi:10.1016/j.syapm.2014.08.002.
Vloon, C.C., Evju, M., Klanderud, K., and Hagen, D. 2022. Alpine restoration: planting and seeding of native species facilitate vegetation recovery. Restor. Ecol. **30**: 1–10. doi:10.1111/rec.13479.

Appendix A. Species descriptions

Our work focused on four common and widely distributed Arctic legume species (Fig. A1). *Astragalus alpinus* Linnaeus (Flora of North America Editorial Committee, 1993+) is a low creeping plant with branching thin stems. The leaves are strigose beneath, glabrous above, and pinnate with 15–23 leaflets. Its flowers are pale violet and arranged in a short raceme. This widely distributed alpine and Arctic species is found in sandy or gravelly areas (Porsild and Cody 1980). *Oxytropis arctica* R. Brown (Flora of North America Editorial Committee, 1993+) often forms dense branching tufts 20–30 cm in diameter. It has 9–13 leaflets on its leaves and the underside of them are densely covered with fine thin hairs. Located at the end of a peduncle slightly longer than its leaves are 3–5 dark purple flowers. It is endemic to Arctic North America and ubiquitous on dry, open tundra, though less common westward (Porsild and Cody 1980). *Oxytropis maydelliana* Trautvetter (Flora of North America Editorial Committee, 1993+) is distinguishable by the brown stipules that densely cover its branching caudex. Its leaves have 13 hairless or sparingly villous leaflets and its florescence has 5–7 pale yellow flowers arranged in a short spike. This species is found across Arctic and alpine parts of East Asia to northern Labrador (Porsild and Cody 1980). *Hedysarum americanum* (Michaux ex Pursh) Britton (Elven, R (Ed), 2007+) has large fleshy rhizomes that can be half an inch in diameter and its stems grow up to 60 cm tall. Leaves are glabrous with 9–13 leaflets and brown stipules. Its inflorescence is large compared to the other three species, with 10–20 pink or pale purple flowers arranged in a raceme of 3–12 cm long at the top of a 5–10 cm long peduncle. Above the treeline, however, the plant may be lower and fewer-flowered. Additionally, this specie has indehiscent pods composed of separable one-seeded sections. It has a general Arctic and alpine distribution across North America (Porsild and Cody 1980).

Fig. A1. Photographs of (A) *Oxytropis arctica*, (B) *Oxytropis maydelliana*, (C) *Astragalus alpinus*, and (D) *Hedysarum americanum*, taken near Rankin Inlet, Nunavut. Photos taken by Eric Lamb.

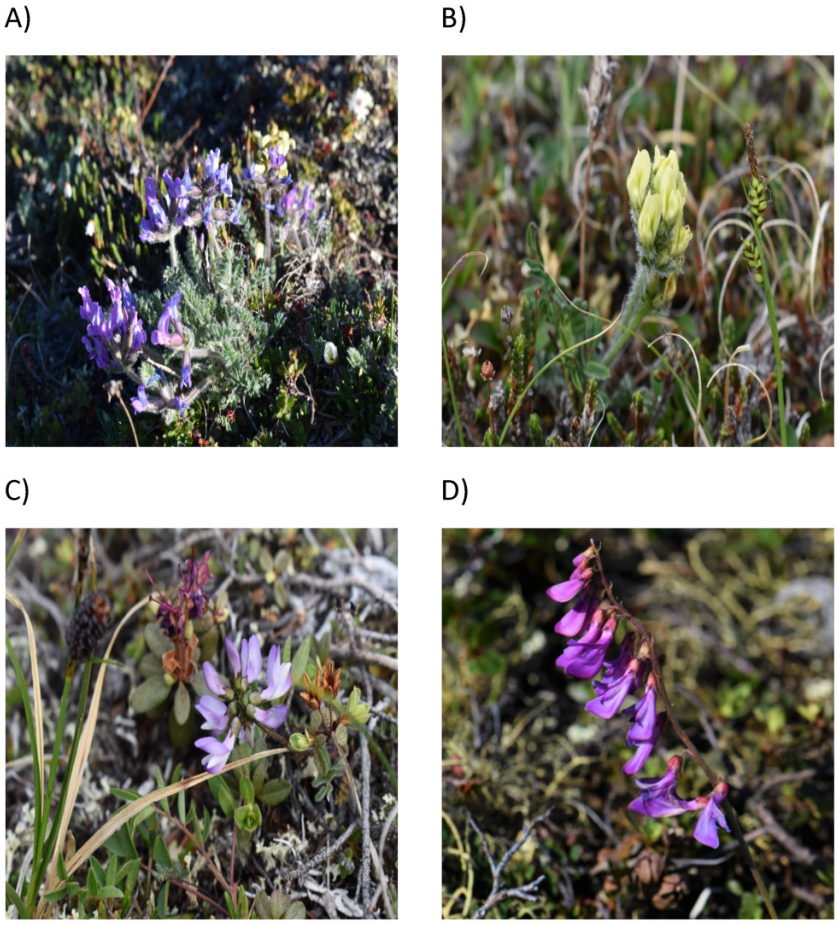


Fig. A2. Location of the three gravel quarries near the Agnico Eagle Mines Ltd. Meliadine mine. Sampling was conducted both within each quarry and in the intact tundra immediately around each quarry. Sampling area is approximately 25 km NW of Rankin Inlet, Nunavut, Canada. Map generated using Google Earth Pro. Map data Google, © 2024 Maxar Technologies.

Table A1. Results of pairwise comparisons among nodule communities of four legume species, both on bare substrate and intact tundra substrates, assessed based on the Euclidean distance matrix of centred-log ratio transformed abundance.

Pairs	Df	Sums of sqs.	*F* model	R2	*p* value
O-arc vs. O-may	1	2422.694	1.429	0.074	0.18
O-arc vs. A-alp	1	1385.736	0.748	0.04	0.53
O-arc vs. H-ame	1	2984.663	1.579	0.095	0.149
O-may vs. A-alp	1	3169.086	2.198	0.109	0.079
O-may vs. H-ame	1	6430.021	4.604	0.235	**0.011**
A-alp vs. H-ame	1	2333.143	1.472	0.089	0.153

Note: O-arc = *Oxytropis arctica*, O-may = *Oxytropis maydelliana*, A-alp = *Astragalus alpinus*, H-ame = *Hedysarum americanum*. *O. maydelliana* and *H. americanum* centred-log ratio transformed abundances are significantly different with $p = 0.011$

Table A2. Summary of Tukey's tests on *nifH* gene Shannon diversities of nodule samples from four legume species—O-arc = *Oxytropis arctica*, O-may = *Oxytropis maydelliana*, A-alp = *Astragalus alpinus*, H-ame = *Hedysarum americanum*.

	Diff	*p* adj
O-arc vs. O-may	0.748	0.379
O-arc vs. A-alp	0.052	0.999
O-arc vs. H-ame	− 1.097	0.156
O-may vs. A-alp	− 0.696	0.441
O-may vs. H-ame	− 1.846	**0.005**
A-alp vs. H-ame	− 1.149	0.129

Note: Shannon diversity was significantly different between *O. maydelliana* and *H. americanum* at $p = 0.005$.

Table A3. Summary of Tukey's tests on *Rhizobium-1 nifH* amplicon sequence variant (88ad6d4de9880fdeb52ea15e8f11c0e0) centred-log ratio transformed abundances from four legume species—O-arc = *Oxytropis arctica*, O-may = *Oxytropis maydelliana*, A-alp = *Astragalus alpinus*, H-ame = *Hedysarum americanum*.

	Diff	*p* adj
O-arc vs. O-may	1.842	0.766
O-arc vs. A-alp	2.256	0.636
O-arc vs. H-ame	− 4.217	0.201
O-may vs. A-alp	0.415	0.996
O-may vs. H-ame	− 6.059	**0.032**
A-alp vs. H-ame	− 6.473	**0.020**

Note: Centred-log ratio transformed abundances of *Rhizobium-1* were significantly different between *O. maydelliana* and *H. americanum* at $p = 0.032$, and between *A. alpinus* and *H. americanum* at $p = 0.02$.

Table A4. Summary of Tukey's honest significant differences on *Mesorhizobium-1 nifH* amplicon sequence variant (*88ad6d4de9880fdeb52ea15e8f11c0e0*) centered-log ratio transformed abundances from four legume species—O-arc = *Oxytropis arctica*, O-may = *Oxytropis maydelliana*, A-alp = *Astragalus alpinus*, H-ame = *Hedysarum americanum*.

	Diff	*p* adj
O-arc vs. O-may	1.842	0.766
O-arc vs. A-alp	2.256	0.636
O-arc vs. H-ame	− 4.217	0.201
O-may vs. A-alp	0.415	0.996
O-may vs. H-ame	− 6.059	**0.032**
A-alp vs. H-ame	− 6.473	**0.02**

Note: Centered-log ratio transformed abundances of *Mesorhizobium-1* were significantly different between *O. maydelliana* and *H. americanum* at $p = 0.032$, and between *A. alpinus* and *H. americanum* at $p = 0.02$.

Fig. A3. (A) Species scores for the 309 16S amplicon sequence variants (ASVs) in the principle component analysis (PCA) ordination. (B) Species scores for the 306 *nifH* ASVs in the PCA ordination.