Mapping the relative abundance of soil microbiome biodiversity from eDNA and remote sensing

Andrew K. Skidmore [a,b,*], Andjin Siegenthaler [a], Tiejun Wang [a], Roshanak Darvishzadeh [a], Xi Zhu [a], Anthony Chariton [b], G. Arjen de Groot [c]

[a] *Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, P.O. Box 217, 7500 AE, Enschede, the Netherlands*
[b] *Macquarie University, Sydney, NSW, 2109, Australia*
[c] *Wageningen Environmental Research, Wageningen UR, P.O. Box 46, 6700 AA, Wageningen, the Netherlands*

ABSTRACT

Although an enormous number of plant and animal species have been directly observed and recorded in online databases, the spatial variation in the composition of the microbiome remains relatively largely unknown. In this study, for the first time, we demonstrate mapping of the relative abundance of the soil microbiome for three terrestrial ecosystems across North America (savanna, boreal and tundra) using airborne image spectroscopy and environmental DNA (eDNA) data. We identified field plots of publicly available eDNA data co-occurring with AVIRIS-NG hyperspectral imagery. An eDNA processing pipeline was developed to generate a consistent profile of the relative abundance for thousands of microbiome operational taxonomic units (OTU) and 225 microbiome families. Using Linear Discriminate Analysis (LDA) scores for the eDNA data, we identified 81 families with the greatest explanatory power based on the community composition between the three ecosystems. A case study example demonstrates our conceptual approach by selecting a dominant and functionally important bacterial family for each ecosystem, with each family representing a specific biomarker. A partial least squares regression (PLSR) was applied to spatially predict the relative abundance of each bacteria family (in the boreal, tundra, and savanna ecosystems) from the hyperspectral imagery. For the boreal, Pseudomonadaceae is shown to be a dominant family taxon, as it is a saprophytic family specialized in decomposing the dense organic matter of boreal forest soils. Members of an understudied family of Acidobacteria, so far indicated as AKIW659, are abundant in acidic Arctic soils and peat bogs. Finally, the Micromonosporaceae are dominant and functionally important in drier regions with grass-tree dominated woodlands, being a member of Actinobacteria with a high relative abundance in soils with high carbon content and nitrate leaching. We demonstrate for the first time how the spatial prediction of relative abundance of these bacteria taxa based on remote sensing, showing patterns of the soil microbiome biodiversity and ecosystem function within and across the three ecosystems.

1. Introduction

Traditional biodiversity monitoring involves the field observation of species by specialized biologists, aided by skilled volunteers, whose expertise is restricted to specific biotic groupings. Botanists may identify and record the presence of plant species and ornithologists the bird biota, but 'unpopular' biotic groups such as fungi, bacteria and insects are under-observed or escape identification altogether. Traditional methods for identifying soil microbial communities from *in situ* samples have been based on direct human observation of organisms cultured under laboratory conditions. Though traditional techniques have been

steadily refined they remain expensive, time consuming, inconsistent across time and space, as well as only capturing a fraction of true biodiversity (Chariton et al., 2010). In recent years, environmental DNA (eDNA) metabarcoding has enabled the rapid collection of comprehensive species occurrence of a sample's microbial community composition and functional potential (Taberlet et al., 2018).

Though eDNA remains expensive, the costs per amount of data are rapidly reducing, while protocols (for field sampling of eDNA as well as laboratory analysis) are standardizing, and from massive online reference databases it is possible to identify taxa from DNA-sequences (Chariton et al., 2010; Taberlet et al., 2018). Another key innovation

is the rapid growth in the number of meta-databases of sequenced DNA samples, though samples are thin on the ground especially when upscaled to landscape or global levels. For the microbiome, the spatial distribution of taxa remains essentially unknown, as more intensive sampling to 'fill the gaps' at landscape to global levels would be prohibitively expensive in terms of fieldwork and laboratory costs.

Concurrent with these advances in microbiology, imaging spectroscopy with a finer spatial resolution pixel (2–5 m) can remotely sense the chemical, compositional, and functional characteristics of above-ground plant communities using spectral reflectance and vegetation indices (Lausch et al., 2016; Feilhauer et al., 2016; Rahman et al., 2003). Interestingly, even ecosystem functions such as phenology and photosynthetic activity have been linked to vegetation indices using coarser resolution hyperspectral satellites (250–500 m) such as MODIS and Hyperion, (Zhang et al., 2003; Guerschman et al., 2009). The combination of very high-resolution image spectroscopy and eDNA approaches link disparate ecosystem components by capturing dominant ecosystem functions across multiple domains of life. The use of very high-resolution image spectroscopy allows contiguous mapping, by 'filling the gaps' between *in situ* eDNA field samples.

As biodiversity change has many dimensions and occurs at a range of spatial and temporal scales, relying on observational *in situ* samples is proving inadequate for the 196 parties to the UN Convention on Biodiversity (CBD) to meet their monitoring commitments (Marques et al., 2014). The Essential Biodiversity Variables developed by the Group of Earth Observation Biodiversity Observation Network (GEO BON) provide a framework for measuring biodiversity at multiple levels (Pereira et al., 2013). Although some essential biodiversity variables can already be monitored using broad-band optical satellite imagery such as from Landsat and Sentinel-2, these are primarily measures of gross biodiversity metrics, e.g. land cover and phenology (Potapov et al., 2017; Skidmore et al., 2015), and lack the high spectral resolution needed to model ecosystem functions from eDNA *in situ* samples. To address this shortfall requires fundamentally new approaches for rapidly quantifying and understanding the complexity and changes in global biodiversity, including the relative abundance of species and community composition. Exciting actual and planned innovations in next-generation hyperspectral satellites (*viz.,* the Environmental Mapper and Analysis Programme EnMAP (Storch, 2022) from DLR-Germany, PRecursore IperSpettrale della Missione Applicativa PRISMA (PRISMA, 2022) from ASI-Italy , Copernicus Hyperspectral Imaging Mission for the Environment CHIME (ESA, 2022) from ESA-Europe, and Surface Biology and Geology SBG (Nastal and Pavlick, 2022) from NASA-USA) further our abilities to upscale biodiversity products to a global level and further understanding of how biodiversity responds to stress and environmental change (Lausch et al., 2016).

Here we demonstrate for the first time the potential to spatially predict microbial community composition from remote sensing. We further highlight examples of how ecosystem functions, such as soil respiration and plant growth reduction in waterlogged anaerobic conditions, can be synchronously measured by remote sensing as well as deduced from the high relative abundance of specific soil microbiome families. This builds on the work of spatial ecology and image spectroscopy, highlighting how functional ecology has been linked to image spectroscopy (Asner and Martin, 2016) as well as Plant Functional Types (Ustin and Gamon, 2010) as indicators of soil properties and the soil microbiome (Aponte et al., 2013). Our research, using the image spectroscopy response to inform on environmental niche/habitat, reveals for the first time how biodiversity such as the soil properties as well the soil microbiome can be derived from image spectroscopy. In other words, through analysis of co-occurring eDNA and image spectroscopy data, we aim to show how ecosystem function can be interpreted from the literature, thereby filling the Linnaean gap (i.e., the gap caused by most species not being described by traditional sampling) and the Wallacean gap (i.e., the unknown geographical distribution and extent of most species) (Bini et al., 2006).

2. Methods

2.1. Remote sensing data

Using the NASA Jet Propulsion Laboratory (JPL) portal we located high resolution hyperspectral (image spectroscopy) imagery obtained with the Airborne Visible InfraRed Imaging Spectrometer - Next Generation (AVIRIS-NG) between 2014 and 2017. AVIRIS-NG measures the radiance of wavelengths between 380 nm and 2510 nm with a 5 nm sampling interval. Data were processed by the AVIRIS-NG instrument ground data system (IGDS). All publicly available images were free-of-cloud. As there was noise in the radiance values, the useable wavelength range was truncated to 431–2470 nm.

The hyperspectral data are delivered as standardized ortho-corrected as well as radiometrically-corrected reflectance (https://avirisng.jpl.nasa.gov/ang_data_pipeline.html). Depending on the height of the aircraft, the spatial resolution of the imagery ranged from 2 m to 5.2 m. To standardize the image pixel size to the eDNA plot data (with some field locations having more than 2 replicate samples) we assumed replicate points were located with a RMSE $\leq$10 m. The image pixels intersecting with each eDNA field plot (of 10 m × 10 m) were selected and the average reflectance values of all the pixels in the 10 m × 10 m plot retrieved.

2.2. eDNA data

From the thousands of records of field measured eDNA that are publicly available, only 23 eDNA plots were found that co-occur with 5 hyperspectral (AVIRIS-NG) images across North America (Supplementary Table S1). The hyperspectral images, coincident with the eDNA plots, were acquired during summer months (June–August) between 2014 and 2017. The eDNA plot samples were derived from two prominent international databases: the European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI) (Mitchell et al., 2018) and the Earth Microbiome Project (EMP) (Thompson et al., 2017).

The publicly available eDNA data have bacterial 16S gene profiles, allowing inter-comparison of bacteria family taxonomic data across the three ecosystems (savanna, boreal and tundra). These profiles have been constructed using the well-established 515F-806R primer set targeting the V4 region of the 16S SSU rRNA (Caporaso et al., 2011). The DNA profiles were downloaded from the public databases as operational taxonomic unit (OTU) tables annotated using the Greengenes reference database (v13.8). Data in these databases has been collected and processed according to standardized protocols (dx.doi.org/10.17504/protocols.io.nuudeww) and comply with the high data standards of these databases (https://earthmicrobiome.org/protocols-and-standards/16s/, https://www.ebi.ac.uk/ena/browser/about/data-standards), with the aim to reduce variation between individual studies (Yilmaz et al., 2011). Meta-analyses of microbial community samples collated in these databases have resulted in over 100 high-quality publications exploring microbial diversity at local to global scales by utilizing the information stored about the DNA extraction, amplification, sequencing and bioinformatic (QIIME 1) protocols (Caporaso et al., 2010; Thompson et al., 2017; Nuccio et al., 2016; Gilbert et al., 2014). See Supplementary Table S2 for source information of each dataset used in this study, including the sampling and sequencing procedures. We proceeded with only those taxa assigned to Eubacteria (i.e., true bacteria), removing any taxa assigned to Eukaryotic organelles (i.e., mitochondria and chloroplast) and Archaea (i.e., organisms with molecular characteristics separating them from bacteria). Replicate samples from the same plot were averaged and taxonomic units with a maximum of 9 or less counts per sample were removed on a sample-by-sample basis to avoid false positives and low-frequency noise (De Barba et al., 2014; Siegenthaler et al., 2019). The resulting table was rarefied to 10,000 reads per sample to normalize for variation in sequencing depth (mean $\pm$ SD read count per sample: 90,693 $\pm$ 49,086) (Sanders, 1968; Weiss et al., 2017) and taxa were grouped into families and then the relative read abundance per family was calculated by

dividing the raw reads of each family by the total number of reads in a sample. Relative abundances show the relative contribution of each taxon to the overall composition of the amplified product and are less influenced compared to raw read counts by stochastic processes introduced by the amplification and sequencing process (Philippot et al., 2009; Taberlet et al., 2012). Relative abundances is a well established approach for studying microbial community composition (Orwin et al., 2018; Lozupone and Knight, 2007; Roesch et al., 2007; Jesmok et al., 2016) as well as analysing diversity patterns in eDNA datasets (Philippot et al., 2009; Albaina et al., 2016; Evans et al., 2016). We conducted Linear Discriminant Analysis (LDA) to select bacterial families that form representative biomarkers for the 3 ecosystems. Each ecosystem (i.e., savanna, boreal and tundra) represented a *class* in a Linear Discriminant Analysis (LDA). The *variables* in this LDA were the eDNA families and values were the relative abundance of each family. Differentially abundant families between the three ecosystems (boreal, savanna and tundra) were identified using LDA Effect Size (LEfSe) analysis (Segata et al., 2011) with $\alpha' \leq 0.05$ (factorial Kruskal-Wallis test) and a minimum logarithmic LDA score ≥ 3 thresholds (Segata et al., 2011).

2.3. Modeling the bacterial families from remote sensing using PLSR

Based on the Linear Discriminant Analysis (LDA) thresholds, a total of 81 discriminant family taxonomic units were extracted for each sample. Then, using a Partial Least Squares regression (PLSR) the relative abundance of each family was predicted using the hyperspectral imagery (with the AVIRIS image band being the predictor variable in the PLSR analysis). PLSR reduces predictor variables (i.e., image bands) to a small set of uncorrelated components, and regression is then started on this small set of components. As spectral measurements have predictors (wavebands) which are highly collinear and heavily outnumber observations, PLSR has become the preferred method of analysis for image spectroscopy and is frequently used for modeling multiple response variables (Axelsson et al., 2013; Ramoelo et al., 2013; Bi et al., 2019). The optimal number of PLSR components (with 'components' being a term often used in remote sensing, versus 'latent variables' which is used more in biology) was chosen based on the lowest corrected Akaike Information Criterion (AICc) threshold values and lowest RMSE, where AICc is the AIC corrected for small sample sizes (Segata et al., 2011). PLSR models were run in R (v. 4.1.1) using 'pls' (v. 2.8-0) with NIPALS (Nonlinear Iterative Partial Least Squares) (Mevik and Wehrens, 2007) as well as the Python (3.8.2) environment. Models were validated using leave-one-out (LOO) cross-validated datasets (Philippot et al., 2009; Albaina, 2016; Evans et al., 2016) (see also Supplementary Fig. S3). Examples of discriminant bacterial families (Log-LDA score >3) which were highly correlated ($R^2 > 0.5$) to the image spectra in the PLSR models) are summarized in Table 1.

To demonstrate the potential to map and further understand the distribution of microbial families, we selected three example microbial families (from the 81 pre-selected families) as a representative biomarker for each of the three ecosystems (Table 2). These three families were selected based on a high R-squared value from the PLSR and known ecological information explaining the presence of each

family taxon across a landscape based on its ecological function. We considered when selecting the three family taxa whether the typical habitat of the family taxon could be deduced from known hyperspectral remote sensing wavelengths. For example, it is known that 940–960 nm are wavelengths important for discriminating tundra vegetation (Bratsch et al., 2017), while the 1305–1355 nm range is associated with the waterlogged conditions in the Arctic tundra which exposes a greater component of the soil/ground/water reflectance signal (Knipling, 1970). From the literature, we also considered the habitat typical of the family taxon and links to ecological function (e.g., amount of green biomass and non-photosynthetic components such as lignin and cellulose linked to photosynthesis, the role of canopy nitrogen and phosphorus in nutrient cycling, and role of biochemical compounds such as leaf proteins and polyphenols etc.) The first selected microbial family AKIW659 is abundant in waterlogged, anaerobic, and highly acidic soils, and proportionally more abundant in tundra wetlands. The second family is Pseudomonadaceae, which is saprophytic and specialized in decomposing the thick organic matter of boreal forest floors. Finally, the Micromonosporaceae family is dominant and functionally important in savannah soils, and forms in herbage with a higher carbon content and

Table 2

Known ecological habitat of the three families (AKIW659, Pseudomonadaceae, Micromonosporaceae) with respectively high relative abundance in the tundra, boreal and savanna ecosystems.

Family	Description
AKIW659	AKIW659 is in the phylum Acidobacteria being characterized as an aerobic heterotroph which thrives in acidic conditions and have a specialized role in carbon cycling by oxidizing simple and complex carbohydrates such as starch, cellulose and lignin, in wetlands as well as drier tundra areas which typically have a buildup of carbon rich organic matter associated with a scarcity of nutrients and low quality soils (de Chaves et al., 2019; Stark et al., 2014; Koyama et al., 2014). The location of the tundra sample plots is in an ice-wedge polygonal tundra landscape with flat-center polygons, typically with sedges, grasses and various mosses (Davidson et al., 2016).
Pseudomonadaceae	Pseudomonadaceae are in the phylum Proteobacteria, and a metabolically diverse family that degrade a variety of low-molecular-weight organic compounds and hydrolytic products generated by actinomycete and fungal growth. They are aerobic and rarely grow below pH 5.0–6.0 (Cousins et al., 2004). As free-living saprophytic family, Pseudomonadaceae are specialized in decomposing the organic matter (Palleroni Starr et al., 1981) typical of boreal forest ecosystem.
Micromonosporaceae	Micromonosporaceae of the phylum Actinobacteria, are aerobic and chemo-organotrophic occurring on less acidic soils typical of the savannah. They can survive high temperature exposure but are considered to be mesophilic (optimum temperature between 20 °C and 40 °C). Micromonosporaceae are stress tolerant of drought. They are known to fix nitrogen as well as decompose carbon compounds.

Table 1

Examples of bacterial families selected from the 81 family taxonomic units as extracted for each sample. Families were selected that had the following thresholds: they showed a significant (P < 0.05; Kruskal-Wallis test) and discriminant association with a ecosystem (Log-LDA score >3) and they were highly correlated ($R^2 > 0.5$) to the image spectra in the PLSR models. The number of PLSR components was selected based the lowest AICc value. Mean relative read abundance (MRA% in Table 1) are shown per family for the full dataset and for individual ecosystems.

Family	Ecosystem	LDA score	PLSR R^2	MRA% All data	MRA% Savanna	MRA% Boreal	MRA% Tundra
Pseudomonadaceae	boreal	3.64	0.54	0.26	0.02	0.67	0.05
AKIW659	tundra	3.74	0.85	0.21	0.00	0.00	0.61
Anaerolinaceae	tundra	3.88	0.54	0.33	0.00	0.12	0.82
Syntrophaceae	tundra	4.66	0.72	1.77	0.00	0.04	5.04
Micromonosporaceae	savanna	4.40	0.85	1.04	3.36	0.03	0.03
Solirubrobacteraceae	savanna	4.35	0.65	1.03	3.22	0.10	0.04
Geodermatophilaceae	savanna	4.02	0.72	0.51	1.50	0.03	0.11

leaching of nitrates (see Table 2 for more details).

We ran the PLSR (NIPALS option) with leave-one-out (LOO) cross-validation of the predictions (Mevik and Wehrens, 2007). For the three selected microbial families, the optimal number of components (i.e., the lowest AICc and RMSEP) were 5 for AKIW659 and Micromonosporaceae, and 3 components for Pseudomonadaceae (Table 3). To validate the modeling of the bacterial families from the remote sensing using PLSR, the measured and predicted values derived by the leave-one-out cross validation demonstrate that the PLSR models are reliable with an accuracy comparable with similar studies (Ramoelo et al., 2013; Askari et al., 2015; Fernandez-Habas et al., 2021) (Table 3 and Fig. S3).

We utilized the AVIRIS-NG hyperspectral imagery to predict the relative abundance of the 3 microbial families (AKIW659, Pseudomonadaceae, Micromonosporaceae) across the 3 ecosystems (*viz.* tundra, boreal and savanna). The PLSR model directly predicted the relative abundance of each family derived from the eDNA profile data across each AVIRIS-NG image (i.e., Python 3.8.2 PLSR regression 'predict' function), with the relative abundance values of the three selected microbial families being written into three separate image matrices. We finally projected RGB colour composite maps, by combining the relative abundance image of each family into a colour composite image as the red (AKIW659), green (Micromonosporaceae) and blue (Pseudomonadaceae) respectively (Schmidtlein et al., 2012) (Fig. 2). Using the visible bands of the AVIRIS-NG imagery, aerial imagery was simulated as a true (also known as 'natural') colour composition (Siok and Ewiak, 2020) (Fig. 2).

3. Results

DNA profiles contained on average (±SD) 90,693 ± 49,086 reads per sample after removal of reads from Eukaryotic organelles, Archaea, and low-frequency noise. Rarefaction to 10,000 reads per sample resulted in a total number of 779 OTUs (267 ± 59 Mean ± SD OTUs per sample). Seven OTUs were removed as result of the rarefaction. A total of 81 bacterial families were detected in the dataset, with a mean (±SD) 52 ± 10 families per profile.

The PLSR provides insight into the microbial community, as well as indicating habitat requirements by highlighting significant wavelength bands associated with specific microbial families. From Supplementary Fig. S1, the plotted PLSR scores for the first 2 components (latent factors 1 and 2) show good separation of the family taxon across the three ecosystems. For example, AKIW659 has a high relative abundance with PLS1 in the tundra (green dots), and a low relative abundance in the boreal and savanna, highlighting the importance of the microbiome family as a biomaker for certain ecosystems.

Examining the PLSR coefficients in combination with the Variable Importance of Projection (VIP) scores, indicate wavelengths having a significant impact on predicting the relative abundance of each family. The wavelengths having an impact on the relative abundance of the family AKIW659 (Fig. 1a) are at 760 nm, 940–960 nm (double peak), 1120 nm, 1305–1355 nm (double peak), and 2015 nm). The wavelengths having an impact on the relative abundance of the Pseudomonadaceae family associate with the wavelength range 750 nm, 900 nm,

1120 nm with a negative association at 1970–2020 nm for the relative abundance of Pseudomonadaceae (Fig. 1b). A common family present in savanna/Mediterranean landscapes is the Micromonosporaceae, which occurred with high relative abundance in our eDNA profile data. There is a high relative abundance of Micromonosporaceae associated with wavelengths at 680 nm, 1480–1490 nm, 1690–1710 nm, with a low relative abundance at 1425 nm 2100 nm (Fig. 1c).

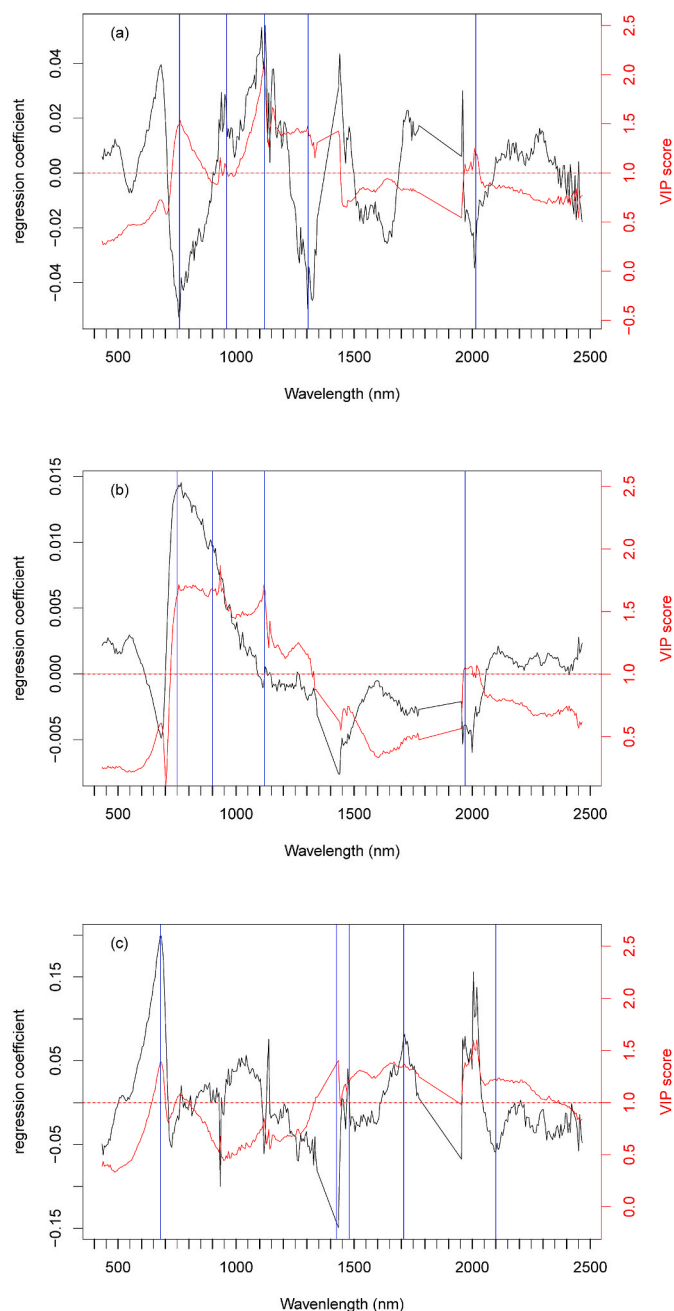The relative abundance of the 3 families in the tundra, boreal and

Fig. 1. The VIP scores and regression coefficients for families a) AKIW659 b) Pseudomonadaceae c) Micromonosporaceae generated by the PLSR analysis for the three ecosystems (boreal, savanna and tundra). VIP scores below the red line (VIP<1) indicate parts of the spectrum that are non-informative for predicting the relative abundance of the respective family. The vertical blue lines indicate the wavelengths that are informative with respect to relative abundance – for example the Micromonosporaceae family is relatively abundant at 680 nm, 1480–1490 nm,1690–1710 nm, and has a low relative abundance at 1425 nm and 2100 nm (Fig. 1c). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 3

Predictions of microbiome family relative abundances using PLSR fitted with AVIRIS-NG hyperspectral imagery. RMSE and R²values are based on bias-corrected cross-validation estimates (NIPALS option). Total model variance explained using an optimal number of components (i.e., the lowest AICc and RMSEP). See also Supplementary Fig. S1.

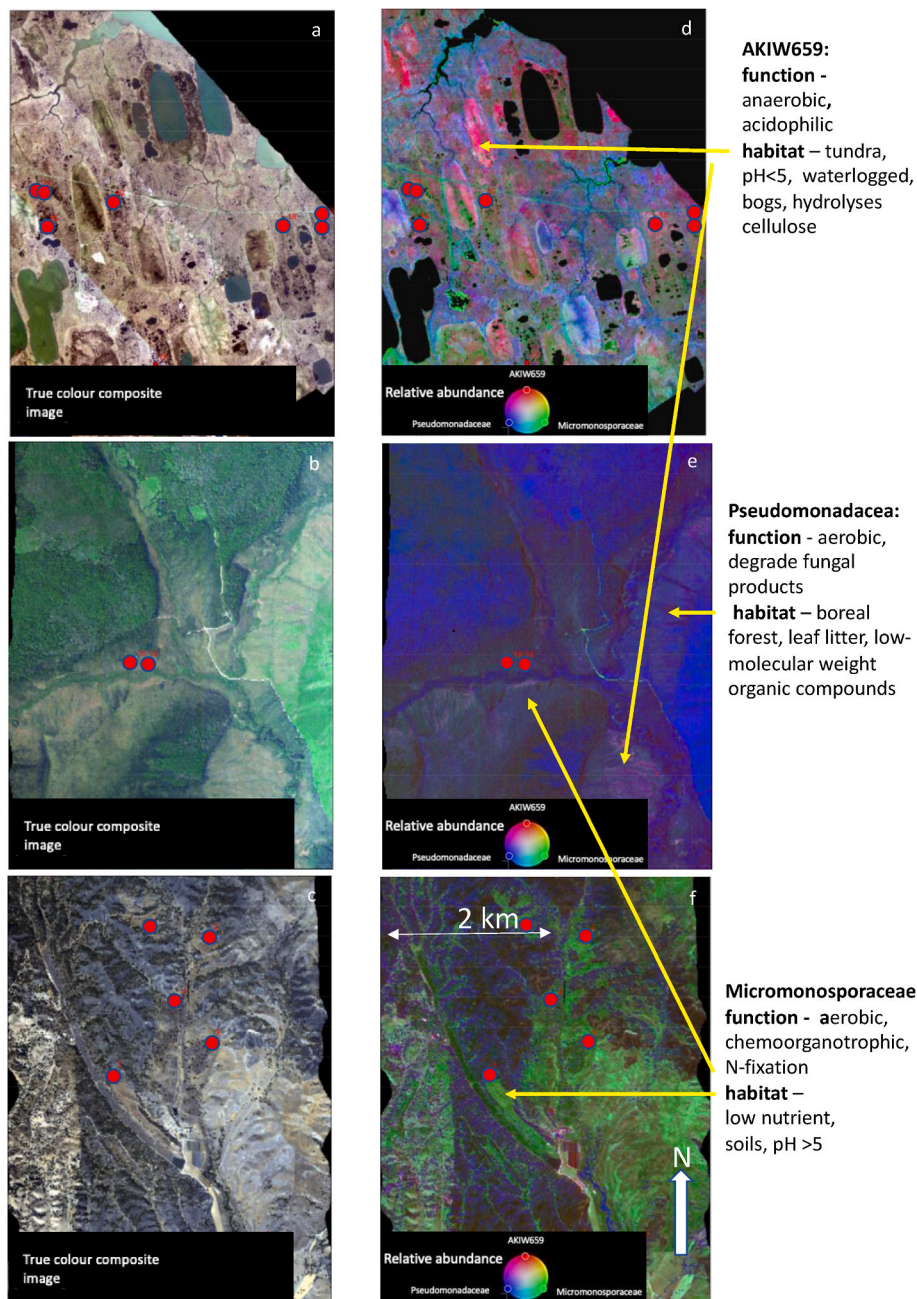| Family	R^2 (Taberlet et al., 2018)	RMSE (adjCV)	Components	Variance explained
AKIW659	0.85	0.013	5	95.34
Pseudomonadaceae	0.54	0.028	3	69.42
Micromonosporaceae	0.85	0.034	5	93.99

Fig. 2. a–f: True colour composite image segments of tundra (2a), boreal (2b), savanna (2c). Relative abundance of the three families (*viz.* AKIW659, Pseudomonadaceae, Micromonosporaceae) in the tundra (2c), boreal (2d) and savanna (2e). Red dots indicate sample locations that are overlapping with the shown image segments – note that some samples are located on adjacent imagery (which are not shown). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

AKIW659:
function - anaerobic, acidophilic
habitat – tundra, pH<5, waterlogged, bogs, hydrolyses cellulose

Pseudomonadacea:
function - aerobic, degrade fungal products
habitat – boreal forest, leaf litter, low-molecular weight organic compounds

Micromonosporaceae
function - aerobic, chemoorganotrophic, N-fixation
habitat – low nutrient, soils, pH >5

savannah hyperspectral images were modelled from reflectance values for each image and compared with a true colour image (Fig. 2). To interpret the relative abundance imagery (Schmidtlein et al., 2012) the colour of each pixel indicates the relative abundance of each family projected as a mix of the 3 colour guns, with the dominant colour indicating the respective higher relative abundance of the microbial families in the landscape. A cursory inspection of the relative abundance images for the tundra, boreal and savanna imagery (Fig. 2) highlights an overall reddish colour for the tundra (where AKIW659 (red) dominates the tundra), boreal is dominated by blue (Pseudomonadaceae), and savanna by green (Micromonosporaceae) – with the relative abundances of the families being confirmed by summary statistics (Table 1). In Fig. 2, larger areas of similar colour are hotspots of higher relative abundance for that specific mix of microbial families. Black pixels are water pixels, filtered using a negative NDVI values.

4. Discussion

The family AKIW659 occurs with high relative abundance in the tundra, the Psedomonadaceae family is abundant in the boreal, and for Mediterranean climates the Micromonosporaceae family has a high relative abundance (Table 1). There is a relative abundance of AKIW659 associated with an important (VIP >1.0) negative regression coefficient at 760 nm, highlighting the contrast between the lack of leaf structure of lichens/mosses (Kuusinen et al., 2020) (i.e., poorly developed NIR shoulder) in the tundra with the well-developed NIR shoulder structure typical of grasses, shrubs and trees in tundra, boreal and savanna vegetation (Schmidt and Skidmore, 2001; Hope et al., 1993) (Fig. 1a). The wavelength range 660–680 nm has a high regression coefficient (Fig. 1a) and has been shown to be statistically significant for differentiating vegetation communities in this region west of Barrow Alaska (Davidson et al., 2016). This part of the visible spectrum, coinciding with a peak contribution of the regression coefficient and VIP (0.7),

differentiates grasses (including hummock grass) and non-ericaceous shrubs from mosses (Tieszen and Johnson, 1968), with a higher chlorophyll *a* pigment concentration being associated with green vegetation (Liu et al., 2017). We do note that even though there is a high correlation between relative abundance of AKIW659 and this wavelength, the VIP of 0.7 indicates the wavelength does not contribute a large amount of information to the model. The double peak regression coefficient at 940–960 nm has been noted as a wavelength range important for discriminating tundra vegetation (Bratsch et al., 2017). This concurs with findings that structural carbon elements (including lignin, starch, and cellulose) are strongly associated with the SWIR at 950–1000 nm (starch) and at 1120 nm (lignin and cellulose) (Curran, 1989), and as noted in Table 2 the AKIW659 family has a specialized role in carbon cycling by oxidizing carbohydrates. The strong negative coefficient feature at 1305–1355 nm is associated with the waterlogged conditions of the Arctic tundra, which also has a lower leaf area index (LAI) contrasting with the boreal forests to the south, exposing a greater component of the soil/ground/water reflectance signal, interspersed with sparser shrub and grass vegetation (Knipling, 1970). The atmospheric absorption bands for atmospheric cloud and water vapour around 1250–1400 nm are generally not of interest to terrestrial remote sensing, being included in bands on satellites for automatic cirrus cloud detection at 1375 nm (Richter et al., 2011).

The Pseudomonadaceae family occurred with high relative abundance in all boreal samples (Table 1). In the boreal forest, tree spectra dominate an increasing part of hyperspectral image reflectance. The 1120 nm wavelength is known to be associated with lignin (Curran, 1989) which is an important structural carbohydrate component of the fibre content of leaves and litter in boreal forests (Richardson, 2004), and the PLSR analysis indicates this wavelength range is important for explaining the variability of the relative abundance of the Pseudomonadaceae family (Fig. 1b). The importance of the red edge wavelengths 680 nm (Gitelson et al., 1996) association with the relative abundance of Pseudomonadaceae (Fig. 1b) can be explained by the chlorophyll content of the tree species of the Canadian and Alaskan Shield reflecting relatively more near-infrared energy compared to the wetland areas of the tundra. The 750–925 nm range is part of the spectrum associated with light scattering due to leaf and canopy structure (Ali et al., 2016) coupled with low absorption by water, lignin and cellulose, resulting in high plant reflectance and transmittance (Rautiainen et al., 2018). The wavelength range 1970–2020 nm associates with higher protein/nitrogen as well as cellulose/starch content leading to lower reflectance (Curran, 1989). In boreal forests, there is a relative increase in the lignin and cellulose content resulting from slower litter decomposition at the cooler temperatures of more northerly latitudes (Richardson, 2004). Low soil macronutrient levels, especially nitrogen and phosphorus further reduce the rate of litter decomposition (Archibold, 1995). The boreal forests are generally podzolic, acid soils, with a well-developed leaf litter layer containing a relatively high content of cellulose and lignin (Handley, 1954) providing a suitable habitat (Table 3) for the observed high relative abundance of Pseudomonadaceae.

For Mediterranean climates, the Micromonosporaceae have been isolated from soil and plant materials (Genilloud Trujillo et al., 2015), and in our data have a high relative abundance (Table 1), as this family is able to withstand and colonize drier habitats evidenced by a number of stress related genes being detected (Carro et al., 2019). Micromonosporaceae can withstand high temperatures (Khan et al., 2020). This family has been isolated in semiarid grasslands and are thought to have an important role in the degradation of cellulose (Yeager et al., 2017). The 640–660 nm wavelength associated with Micromonosporaceae (Fig. 1c) relates to increased canopy nitrogen content and chlorophyll *a*+b absorption (Curran, 1989) resulting from Micromonosporaceae taxa having a function in plant growth promotion (including nitrogen fixation) (Hirsch and Valdes, 2010) through plant/rhizosphere nitrogen fixing nodules (Trujillo et al., 2014). The ability of Micromonosporaceae to synthesize pigments, likely for UV protection (Carro et al., 2018), is indicated by association with the 640–660 nm range. Micromonosporaceae are soil-associated plant symbionts – including symbionts of grasses (Yeager et al., 2017). The 750–800 nm wavelength is

associated with internal leaf scattering of the vegetation (Liu et al., 2014). A 1480–1490 nm feature is associated with leaf cellulose and starch (Curran, 1989) of drier broadleaf plants, the feature at 1425 nm indicates a relatively lower lignin content (Curran, 1989) and the feature at 2100 nm a relatively lower non-structural carbohydrate concentration (Ramirez et al., 2015); these features have been noted for broadleaf vegetation when compared to coniferous (needle) trees of the boreal ecosystem (Prescott et al., 2004; Li et al., 2016). Lignin, starch, protein and nitrogen are associated with the 1690–1710 nm feature indicative of nitrogen fixation, and the breakdown of cell wall material (Hirsch and Valdes, 2010).

The promising results demonstrate the potential to interpret the relative abundance and ecological function of the three selected families for the savanna, boreal and tundra habitats, from the colour composite images (Fig. 2). For example, of the three selected species, in the boreal (Fig. 2e) the Pseudomonadaceae (bright blue pixels) dominate. In Fig. 2e, the abundance of Pseudomonadaceae is relatively higher (bluer colour) on the north facing slopes associated with the mesic black spruce (coniferous forest) (Fig. 2b) compared with the south facing slopes dominated by drier hardwood forests (LTER, 2021) – see the NEON field site network Caribou-Poker Creeks Research Watershed data for reference (https://www.neonscience.org/field-sites/bona). The southern facing slopes has Micromonosporaceae mixing with the Pseudomonadaceae (i.e., greener tinge in Fig. 2e), which associates with increasing mixed coniferous-broadleaf forest, broadleaf forest and open woodland/tussock vegetation (Haugen et al., 1982). The warmer microclimate on the southern facing slopes also makes this habitat more suitable for Micromonosporaceae compared to the colder northern facing slopes (Table 2). The broadleaf (birch-aspen dominated) forest is a persistent seral hardwood community maintained by repeated disturbance approximately every 75 years, with closed canopy and a shrub/herbaceous understory (Boucher and Reid, 2010), and is associated with a higher relative abundance of Micromonosporaceae. Regular disturbance results in spruce exclusion as this species is fire intolerant. Because conifer litter decays more slowly than hardwood litter (Fassnacht and Gower, 1999; McClaugherty et al., 1985) the relative abundance of Pseudomonadaceae will be higher. Further we note the mixing of the Micromonosporaceae (green-blue areas) in the boreal-taiga vegetation as well as a trace of AKWI659 (red) in the wetter bog areas (Fig. 2b).

Forest vegetation is the habitat driving the relative abundance of the microbiome. Land cover mapping, such as for the Cariboh-Poker Catchment (Haugen et al., 1982), has distinct vegetation types across which the microbial families subtly merge and these microbial families associate with specific habitats across vegetation types (e.g., AKWI659 is in the phylum Acidobacteria and characterized as an aerobic heterotroph typically occurring with sedges, grasses and various mosses). This subtle grading of microbial families occur within specific habitat types allows the relative abundance of the microbiome taxa to be visualized across a vegetation or habitat type - our novel approach provides detailed spatial information about the microbiome biodiversity across vegetation habitats, albeit requiring eDNA and hyperspectral imagery and relatively complicated processing. As discussed above for the three example microbial families that are representative biomarkers for each of the three ecosystems, we deduce that the functional ecology of each microbiome family mirrors the canopy functional traits through the soil resources, as proposed by the Spectranomics approach (Asner and Martin, 2016). Canopy plant species have evolved plant functional types and traits that echo soil structure and nutrient concentration (Massmann et al., 2022), which in turn are further adapted by the abiotic conditions of a particular location including climate, geology, and topography (Higgins et al., 2014). Image spectroscopy can associate ecosystem processes with plants, for example through biogeochemical cycles, thereby mirroring soil conditions (Asner and Martin, 2016) and specifically the soil microbiome (Aponte et al., 2013).

In our approach presented, spectral properties mirror the biochemical, structural, and functional properties of vegetation. That is, the vegetation or Plant Functional Types (PFT) are indicators of soil properties as well as the soil mirobiome. Plant Functional Types (PFTs) is an

important concept linking species with broad vegetation types, based on ecosystem structure, ecosystem function (phenology and physiology), as well as mirroring plant responses to soil and ecosystem conditions (Ustin and Gamon, 2010). Climate as well as the soil biophysical environment, control ecological niche and a biome's vegetation structure and function (Ustin and Gamon, 2010).

The relationship between above ground vegetation and soil processes are known to be both species and ecosystem specific (Ma et al., 2020; Hess and Austin, 2014), with patterns and processes even demonstrated at global scales (Ordonez et al., 2009). Understanding soil-plant relationships requires knowledge of species interactions and soil processes at the level of ecosystems. While spectral indicators of the biota may be context-dependent, many general relations have been developed (Roy et al., 2006), which allows deductive interpretation of rhizosphere functions, as demonstrated in this study for three ecosystems across North America, and further increases understanding of soil microbial processes at a global scale (Niemann et al., 2015). Soil microbiome biodiversity as measured by eDNA reveals a complex community composition. Microbiome families respond to habitat, with the habitat comprising the physical and biotic factors that support the species' survival and reproduction (Kiesewetter and Afkhami, 2021; Wang et al., 2022; Keet et al., 2019). By considering the co-occurring eDNA and image spectroscopy data, our results demonstrate that ecosystem functions and hence related processes such as the presence and relative abundance of microbiome families, can be deductively interpreted, also in ecosystems with vegetation cover.

Given the eDNA profile data contain thousands of taxonomic groups, which in this study were reduced to a list of 81 discriminative families (see Table 1 for some examples of these 81 families), it is straightforward to repeat our novel method for different families as well as at different taxonomic ranks. For example, the three families (viz. AKIW659, Pseudomonadaceae, Micromonosporaceae) chosen to demonstrate our method (Fig. 1) could be replaced in the true colour composite images with any combination of three families from the 81 discriminative families, or indeed other taxonomic ranks such as species, order, or phylum. A key insight of this paper is that it is possible, with the highly multivariate ("deep") eDNA profile data, to predict the relative distribution of any combination of the taxonomic groups present in the study area from remote sensing. We show that it is also methodologically straightforward to discover the (multivariate) relations between the microbiome taxa (in an eDNA profile database), the image spectroscopy, and environmental data, as demonstrated above for the Psedomonadaceae family.

The relatively limited number of samples may result in statistical models being potentially overfitted (Shen et al., 2020). We acknowledge that ideally more samples would be available in this study, though the scarcity of publicly available image spectroscopy imagery that coincide with eDNA profile data across natural areas in North America constrained the study. Though research indicates that as few as 20–30 samples may be adequate for significant and meaningful PLSR analyses (Bastien et al., 2005; Tenenhaus et al., 2005; Kock and Hadaya, 2018), other researchers propose more samples (>100) (Burnett et al., 2021). Several image spectroscopy studies using hard-to-collect biodiversity variables have successfully interpolated environmental variables using n < 25 in addition to the range 26 < n < 50 (Axelsson et al., 2013; Ramoelo et al., 2013; Bi et al., 2019; Askari et al., 2015; Cho et al., 2007). Additionally, by reducing the number of PLSR predictor variables to between 3 and 5 regression components, model overfitting is further diminished (Shen et al., 2020). We checked overfitting of the PLSR model using standard leave-one-out (LOO) cross-validation for the predicted versus measured scores, demonstrating that results are repeatable and robust (Ramoelo et al., 2013; Askari et al., 2015; Fernandez-Habas et al., 2021) (Supplementary Fig. S1) for the three family taxa, with a RMSEP of 0.01 and 0.04 as well as a R^2 of 0.54 and 0.85 between the measured and predicted values for the three families (Table 3). We anticipate other scientists will further test and develop the concept described in this paper for

robustness and repeatability when mapping the microbiome from remote sensing, as more eDNA data become available that are coincident with hyperspectral images.

As well as complementing *in situ* data collection, the method using remote sensing demonstrated here may be further upscaled with respect to the number of taxa for whom the relative abundance is estimated, potentially providing further insight into the ecosystem health of ecosystems. The so-called 'Linnean gap' (i.e., gaps caused by most species not being described with traditional sampling) may easily be filled by increasing the number of marker genes for additional species, as part of the eDNA analysis process (Bini et al., 2006). If multiple images obtained at different dates are available, then it will be possible to interpolate microbial species relative abundance using time-series imagery (Zhang et al., 2003).

5. Conclusion

The requirements for monitoring anthropogenic biodiversity change, as detailed by the UN Sustainable Development Goals as well as the UN Convention on Biodiversity Aichi Targets, necessitates the application of novel technologies to provide spatially contiguous information rapidly and accurately to update biodiversity information for companies, government, and NGOs. The post-2020 Biodiversity Targets of the UN Convention on Biodiversity are not merely aspirational; the implementation of policy is currently constrained by the lack of data and indicators (Geijzendorffer et al., 2016). Several imaging spectroscopy satellites with spatial sampling of 20–30 m have been and will launch between 2019 and 2026, ushering in real opportunities to combine this newly emerging space technology with eDNA *in situ* taxonomic data. Here we demonstrate that ecosystem function can be the bridge between these two rapidly developing technological approaches, combining the massive information content of *in situ* eDNA data with the high spatial accuracy and global overview provided by imaging spectroscopy. As reference collections of eDNA barcode taxonomic databases are coupled with next generation satellite imaging spectrometers, a more comprehensive and complete picture of biodiversity delineating species across all kingdoms of life will become possible by applying and building upon the techniques presented in this paper. Combining eDNA and imaging spectroscopy holds the promise of accurately, rapidly, and cost-effectively monitoring biodiversity over the terrestrial surface of the Earth.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.srs.2022.100065.

References

Albaina, A., Aguirre, M., Abad, D., Santos, M., Estonba, A., 2016. 18S rRNA V9 metabarcoding for diet characterization: a critical evaluation with two sympatric zooplanktivorous fish species. Ecol. Evol. 6, 1809–1824. https://doi.org/10.1002/ece3.1986.

Ali, A.M., Darvishzadeh, R., Skidmore, A.K., van Duren, I., 2016. Effects of canopy structural variables on retrieval of leaf dry matter content and specific leaf area from

remotely sensed data. IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. 9, 898–909. https://doi.org/10.1109/jstars.2015.2450762.

Aponte, C., Garcia, L.V., Maranon, T., 2013. Tree species effects on nutrient cycling and soil biota: a feedback mechanism favouring species coexistence. For. Ecol. Manage. 309, 36–46. https://doi.org/10.1016/j.foreco.2013.05.035.

Archibold, O.W., 1995. Ecology of World Vegetation, first ed. Chapman & Hall Ltd. 9,152,190,256.

Askari, M.S., O'Rourke, S.M., Holden, N.M., 2015. Evaluation of soil quality for agricultural production using visible-near-infrared spectroscopy. Geoderma 243, 80–91. https://doi.org/10.1016/j.geoderma.2014.12.012.

Asner, G.P., Martin, R.E., 2016. Spectranomics: emerging science and conservation opportunities at the interface of biodiversity and remote sensing. Glob. Ecol. Conserv. 8, 212–219. https://doi.org/10.1016/j.gecco.2016.09.010.

Axelsson, C., Skidmore, A.K., Schlerf, M., Fauzi, A., Verhoef, W., 2013. Hyperspectral analysis of mangrove foliar chemistry using PLSR and support vector regression. Int. J. Rem. Sens. 34, 1724–1743. https://doi.org/10.1080/01431161.2012.725958.

Bastien, P., Vinzi, V.E., Tenenhaus, M., 2005. PLS generalised linear regression. Comput. Stat. Data Anal. 48, 17–46. https://doi.org/10.1016/j.csda.2004.02.005.

Bi, K.Y., Gao, S., Niu, Z., Zhang, C.S., Huang, N., 2019. Estimating leaf chlorophyll and nitrogen contents using active hyperspectral LiDAR and partial least square regression method. J. Appl. Remote Sens. 13 https://doi.org/10.1117/1.Jrs.13.034513.

Bini, L.M., Diniz, J.A.F., Rangel, T., Bastos, R.P., Pinto, M.P., 2006. Challenging Wallacean and Linnean shortfalls: knowledge gradients and conservation planning in a biodiversity hotspot. Divers. Distrib. 12, 475–482. https://doi.org/10.1111/j.1366-9516.2006.00286.x.

Boucher, T., Reid, M.S., 2010. Terrestrial ecological system: western north American boreal mesic birch-aspen forest. https://explorer.natureserve.org/Taxon/ELEMENT_GLOBAL.2.817381/Western_North_American_Boreal_Mesic_Birch-Aspen_Forest.

Bratsch, S., Epstein, H., Buchhorn, M., Walker, D., Landes, H., 2017. Relationships between hyperspectral data and components of vegetation biomass in Low Arctic tundra communities at Ivotuk, Alaska. Environ. Res. Lett. 12, 14. https://doi.org/10.1088/1748-9326/aa572e.

Burnett, A.C., et al., 2021. A best-practice guide to predicting plant traits from leaf-level hyperspectral data using partial least squares regression. J. Exp. Bot. 72, 6175–6189. https://doi.org/10.1093/jxb/erab295.

Caporaso, J.G., et al., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7, 335–336. https://doi.org/10.1038/nmeth.f.303.

Caporaso, J.G., et al., 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc. Natl. Acad. Sci. U. S. A 108, 4516–4522. https://doi.org/10.1073/pnas.1000080107.

Carro, L., et al., 2018. Genome-based classification of micromonosporae with a focus on their biotechnological and ecological potential. Sci. Rep. 8 https://doi.org/10.1038/s41598-017-17392-0.

Carro, L., et al., 2019. Uncovering the potential of novel micromonosporae isolated from an extreme hyper-arid Atacama Desert soil. Sci. Rep. 9 https://doi.org/10.1038/s41598-019-38789-z.

Chariton, A.A., Court, L.N., Hartley, D.M., J, C.M., Hardy, C.M., 2010. Ecological assessment of estuarine sediments by pyrosequencing eukaryotic ribosomal DNA. Front. Ecol. Environ. 8, 233–238.

Cho, M.A., Skidmore, A., Corsi, F., van Wieren, S.E., Sobhan, I., 2007. Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least squares regression. Int. J. Appl. Earth Obs. Geoinf. 9, 414–424. https://doi.org/10.1016/j.jag.2007.02.001.

Cousins, M.A., 2004. In: Robinson, R.K., Batt, C.A., Patel, B.D. (Eds.), Encyclopedia of Food Microbiology, 997-893 (978-0-12-227070-3.

Curran, P.J., 1989. Remote-sensing of foliar chemistry. Remote Sens. Environ. 30, 271–278. https://doi.org/10.1016/0034-4257(89)90069-2.

Davidson, S., et al., 2016. Mapping arctic tundra vegetation communities using field spectroscopy and multispectral satellite data in North Alaska, USA. Rem. Sens. 8, 978. https://doi.org/10.3390/rs8120978.

De Barba, M., et al., 2014. DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. Mol. Ecol. Resour. 14, 306–323. https://doi.org/10.1111/1755-0998.12188.

de Chaves, M.G., et al., 2019. Acidobacteria subgroups and their metabolic potential for carbon degradation in sugarcane soil amended with vinasse and nitrogen fertilizers. Front. Microbiol. 10, 16. https://doi.org/10.3389/fmicb.2019.01680.

ESA, 2022. CHIME (Copernicus). https://directory.eoportal.org/web/eoportal/satellite-missions/c-missions/chime-copernicus.

Evans, N.T., et al., 2016. Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. Mol. Ecol. Resour. 16, 29–41. https://doi.org/10.1111/1755-0998.12433.

Fassnacht, K.S., Gower, S.T., 1999. Comparison of the litterfall and forest floor organic matter and nitrogen dynamics of upland forest ecosystems in north central Wisconsin. Biogeochemistry 45, 265–284. https://doi.org/10.1007/bf00993003.

Feilhauer, H., Doktor, D., Schmidtlein, S., Skidmore, A.K., 2016. Mapping pollination types with remote sensing. J. Veg. Sci. 27, 999–1011. https://doi.org/10.1111/jvs.12421.

Fernandez-Habas, J., et al., 2021. Investigating the potential of Sentinel-2 configuration to predict the quality of Mediterranean permanent grasslands in open woodlands. Sci. Total Environ. 791 https://doi.org/10.1016/j.scitotenv.2021.148101.

Geijzendorffer, I.R., et al., 2016. Bridging the gap between biodiversity data and policy reporting needs: an Essential Biodiversity Variables perspective. J. Appl. Ecol. 53, 1341–1350. https://doi.org/10.1111/1365-2664.12417.

Genilloud, O., 2015. In: Trujillo, M.E., et al. (Eds.), Bergey's Manual of Systematics of Archaea and Bacteria, 1-28. Wiley & Sons, Inc.

Gilbert, J.A., Jansson, J.K., Knight, R., 2014. The Earth Microbiome project: successes and aspirations. BMC Biol. 12 https://doi.org/10.1186/s12915-014-0069-1.

Gitelson, A.A., Merzlyak, M.N., Lichtenthaler, H.K., 1996. Detection of red edge position and chlorophyll content by reflectance measurements near 700 nm. J. Plant Physiol. 148, 501–508. https://doi.org/10.1016/s0176-1617(96)80285-9.

Guerschman, J.P., et al., 2009. Estimating fractional cover of photosynthetic vegetation, non-photosynthetic vegetation and bare soil in the Australian tropical savanna region upscaling the EO-1 Hyperion and MODIS sensors. Remote Sens. Environ. 113, 928–945. https://doi.org/10.1016/j.rse.2009.01.006.

Handley, W.R.C., 1954. Mull and More Formation in Realtion to Forest Soils. Imperial Forestry Institute, London, 1-114.

Haugen, R.K., Slaughter, C.W., Howe, K.E., Dingman, S.L., 1982. Hydrology and Climatology of the Caribou-Poker Creeks Research Watershed, p. 41. Washington, D. C.

Hess, L.J.T., Austin, A.T., 2014. Pinus ponderosa alters nitrogen dynamics anddiminishes the climate footprint in natural ecosystems of Patagonia. J. Ecol. 102, 610–621. https://doi.org/10.1111/1365-2745.12228.

Higgins, M.A., et al., 2014. Linking imaging spectroscopy and LiDAR with floristic composition and forest structure in Panama. Remote Sens. Environ. 154, 358–367. https://doi.org/10.1016/j.rse.2013.09.032.

Hirsch, A.M., Valdes, M., 2010. Micromonospora: an important microbe for biomedicine and potentially for biocontrol and biofuels. Soil Biol. Biochem. 42, 536–542. https://doi.org/10.1016/j.soilbio.2009.11.023.

Hope, A.S., Kimball, J.S., Stow, D.A., 1993. The relationship between tussock tundra spectral reflectance properties and biomass and vegetation composition. Int. J. Rem. Sens. 14, 1861–1874.

Jesmok, E.M., Hopkins, J.M., Foran, D.R., 2016. Next-generation sequencing of the bacterial 16S rRNA gene for forensic soil comparison: a feasibility study. J. Forensic Sci. 61, 607–617. https://doi.org/10.1111/1556-4029.13049.

Keet, J.H., Ellis, A.G., Hui, C., Le Roux, J.J., 2019. Strong spatial and temporal turnover of soil bacterial communities in South Africa's hyperdiverse fynbos biome. Soil Biol. Biochem. 136 https://doi.org/10.1016/j.soilbio.2019.107541.

Khan, M.J., Jurburg, S.D., He, J.Z., Brodie, G., Gupta, D., 2020. Impact of microwave disinfestation treatments on the bacterial communities of no-till agricultural soils. Eur. J. Soil Sci. 71, 1006–1017. https://doi.org/10.1111/ejss.12867.

Kiesewetter, K.N., Afkhami, M.E., 2021. Microbiome-mediated effects of habitat fragmentation on native plant performance. New Phytol. 232, 1823–1838. https://doi.org/10.1111/nph.17595.

Knipling, E.B., 1970. Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. Remote Sens. Environ. 1, 155–159.

Kock, N., Hadaya, P., 2018. Minimum sample size estimation in PLS-SEM: the inverse square root and gamma-exponential methods. Inf. Syst. J. 28, 227–261. https://doi.org/10.1111/isj.12131.

Koyama, A., Wallenstein, M.D., Simpson, R.T., Moore, J.C., 2014. Soil bacterial community composition altered by increased nutrient availability in Arctic tundra soils. Front. Microbiol. 5, 16. https://doi.org/10.3389/fmicb.2014.00516.

Kuusinen, N., Juola, J., Karki, B., Stenroos, S., Rautiainen, M., 2020. A spectral analysis of common boreal ground lichen species. Remote Sens. Environ. 247, 13. https://doi.org/10.1016/j.rse.2020.111955.

Lausch, A., et al., 2016. Linking Earth Observation and taxonomic, structural and functional biodiversity: local to ecosystem perspectives. Ecol. Indicat. 70, 317–339. https://doi.org/10.1016/j.ecolind.2016.06.022.

Li, N., He, N.P., Yu, G.R., Wang, Q.F., Sun, J., 2016. Leaf non-structural carbohydrates regulated by plant functional groups and climate: evidences from a tropical to cold-temperate forest transect. Ecol. Indicat. 62, 22–31. https://doi.org/10.1016/j.ecolind.2015.11.017.

Liu, L.Y., Huang, W.J., Pu, R.L., Wang, J.H., 2014. Detection of internal leaf structure deterioration using a new spectral ratio index in the near-infrared shoulder region. J. Integr. Agric. 13, 760–769. https://doi.org/10.1016/s2095-3119(13)60385-8.

Liu, N.F., Budkewitsch, P., Treitz, P., 2017. Examining spectral reflectance features related to Arctic percent vegetation cover: implications for hyperspectral remote sensing of Arctic tundra. Remote Sens. Environ. 192, 58–72. https://doi.org/10.1016/j.rse.2017.02.002.

Lozupone, C.A., Knight, R., 2007. Global patterns in bacterial diversity. Proc. Natl. Acad. Sci. U. S. A 104, 11436–11440. https://doi.org/10.1073/pnas.0611525104.

LTER, B.C., 2021. Study Sites & Design: Regional Site Network. University of Alaska, Fairbanks. http://www.lter.uaf.edu/research/study-sites-regional.

Ma, J., et al., 2020. Herbaceous layer determines the relationship between soil respiration and photosynthesis in a shrub-dominated desert plant community. Plant Soil 449, 193–207. https://doi.org/10.1007/s11104-020-04484-6.

Marques, A., et al., 2014. A framework to identify enabling and urgent actions for the 2020 Aichi Targets. Basic Appl. Ecol. 15, 633–638. https://doi.org/10.1016/j.baae.2014.09.004.

Massmann, A., Cavaleri, M.A., Oberbauer, S.F., Olivas, P.C., Porder, S., 2022. Foliar stoichiometry is marginally sensitive to soil phosphorus across a lowland tropical rainforest. Ecosystems 25, 61–74. https://doi.org/10.1007/s10021-021-00640-w.

McClaugherty, C.A., Pastor, J., Aber, J.D., Melillo, J.M., 1985. Forest litter decomposition in relation to soil-nitrogen dynamics and litter quality. Ecology 66, 266–275. https://doi.org/10.2307/1941327.

Mevik, B.-H., Wehrens, R., 2007. The pls package: principal component and partial least squares regression in R. J. Stat. Software 18, 1–23. https://doi.org/10.18637/jss.v018.i02.

Mitchell, A.L., et al., 2018. EBI Metagenomics in 2017: enriching the analysis of microbial communities, from sequence reads to assemblies. Nucleic Acids Res. 46, D726–D735. https://doi.org/10.1093/nar/gkx967.

Nastal, J., Pavlick, R.P., 2022. Surface Biology and Geology (SBG). https://science.nasa.gov/earth-science/decadal-sbg.

Niemann, K.O., Quinn, G., Stephen, R., Visintini, F., Parton, D., 2015. Hyperspectral remote sensing of mountain pine beetle with an emphasis on previsual assessment. Can. J. Rem. Sens. 41, 191–202. https://doi.org/10.1080/07038992.2015.1065707.

Nuccio, E.E., et al., 2016. Climate and edaphic controllers influence rhizosphere community assembly for a wild annual grass. Ecology 97, 1307–1318. https://doi.org/10.1890/15-0882.1.

Ordonez, J.C., et al., 2009. A global study of relationships between leaf traits, climate and soil measures of nutrient fertility. Global Ecol. Biogeogr. 18, 137–149. https://doi.org/10.1111/j.1466-8238.2008.00441.x.

Orwin, K.H., Dickie, I.A., Holdaway, R., Wood, J.R., 2018. A comparison of the ability of PLFA and 16S rRNA gene metabarcoding to resolve soil community change and predict ecosystem functions. Soil Biol. Biochem. 117, 27–35. https://doi.org/10.1016/j.soilbio.2017.10.036.

Palleroni, N.J., 1981. In: Starr, Mortimer P., et al. (Eds.), The Prokaryotes: A Handbook on Habitats, Isolation, and Identification of Bacteria. Springer Berlin Heidelberg, pp. 655–665.

Pereira, H.M., et al., 2013. Essential biodiversity variables. Science 339, 277–278. https://doi.org/10.1126/science.1229931.

Philippot, L., et al., 2009. Spatial patterns of bacterial taxa in nature reflect ecological traits of deep branches of the 16S rRNA bacterial tree. Environ. Microbiol. 11, 3096–3104. https://doi.org/10.1111/j.1462-2920.2009.02014.x.

Potapov, P., et al., 2017. The last frontiers of wilderness: tracking loss of intact forest landscapes from 2000 to 2013. Sci. Adv. 3, 13. https://doi.org/10.1126/sciadv.1600821.

Prescott, C.E., Vesterdal, L., Preston, C.M., Simard, S.W., 2004. Influence of initial chemistry on decomposition of foliar litter in contrasting forest types in British Columbia. Can. J. For. Res. 34, 1714–1729. https://doi.org/10.1139/x04-040.

PRISMA, A.S.I., 2022. https://www.asi.it/en/earth-science/prisma/.

Rahman, A.F., Gamon, J.A., Sims, D.A., Schmidts, M., 2003. Optimum pixel size for hyperspectral studies of ecosystem function in southern California chaparral and grassland. Remote Sens. Environ. 84, 192–207. https://doi.org/10.1016/s0034-4257(02)00107-4.

Ramirez, J.A., et al., 2015. Near-infrared spectroscopy (NIRS) predicts non-structural carbohydrate concentrations in different tissue types of a broad range of tree species. Methods Ecol. Evol. 6, 1018–1025. https://doi.org/10.1111/2041-210x.12391.

Ramoelo, A., et al., 2013. Non-linear partial least square regression increases the estimation accuracy of grass nitrogen and phosphorus using in situ hyperspectral and environmental data. ISPRS J. Photogrammetry Remote Sens. 82, 27–40. https://doi.org/10.1016/j.isprsjprs.2013.04.012.

Rautiainen, M., et al., 2018. Spectral properties of coniferous forests: a review of in situ and laboratory measurements. Rem. Sens. 10, 207.

Richardson, A.D., 2004. Foliar chemistry of balsam fir and red spruce in relation to elevation and the canopy light gradient in the mountains of the northeastern United States. Plant Soil 260, 291–299. https://doi.org/10.1023/b:Plso.0000030179.02819.85.

Richter, R., Wang, X.J., Bachmann, M., Schlapfer, D., 2011. Correction of cirrus effects in Sentinel-2 type of imagery. Int. J. Rem. Sens. 32, 2931–2941. https://doi.org/10.1080/01431161.2010.520346.

Roesch, L.F., et al., 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. ISME J. 1, 283–290. https://doi.org/10.1038/ismej.2007.53.

Roy, D.R., Boschetti, L., Trigg, S.N., 2006. Remote sensing of fire severity: assessing the performance of the normalized Burn ratio. Geosci. Rem. Sens. Lett. IEEE 3, 112–116. https://doi.org/10.1109/lgrs.2005.858485.

Sanders, H.L., 1968. Marine benthic diversity - a comparative study. Am. Nat. 102, 243–&. https://doi.org/10.1086/282541.

Schmidt, K.S., Skidmore, A.K., 2001. Exploring spectral discrimination of grass species in African rangelands. Int. J. Rem. Sens. 22, 3421–3434.

Schmidtlein, S., Feilhauer, H., Bruelheide, H., 2012. Mapping plant strategy types using remote sensing. J. Veg. Sci. 23, 395–405. https://doi.org/10.1111/j.1654-1103.2011.01370.x.

Segata, N., et al., 2011. Metagenomic biomarker discovery and explanation. Genome Biol. 12 https://doi.org/10.1186/gb-2011-12-6-r60.

Shen, L.Z., et al., 2020. Hyperspectral estimation of soil organic matter content using different spectral preprocessing techniques and PLSR method. Rem. Sens. 12 https://doi.org/10.3390/rs12071206.

Siegenthaler, A., Wangensteen, O.S., Benvenuto, C., Campos, J., Mariani, S., 2019. DNA metabarcoding unveils multiscale trophic variation in a widespread coastal opportunist. Mol. Ecol. 28, 232–249. https://doi.org/10.1111/mec.14886.

Siok, K., Ewiak, I., 2020. The simulation approach to the interpretation of archival aerial photographs. Open Geosci. 12, 1–10. https://doi.org/10.1515/geo-2020-0001.

Skidmore, A.K., et al., 2015. Agree on biodiversity metrics to track from space. Nature 523, 403–405. https://doi.org/10.1038/523403a.

Stark, S., Mannisto, M.K., Eskelinen, A., 2014. Nutrient availability and pH jointly constrain microbial extracellular enzyme activities in nutrient-poor tundra soils. Plant Soil 383, 373–385. https://doi.org/10.1007/s11104-014-2181-y.

Storch, T., 2022. EnMAP. https://www.dlr.de/eoc/en/desktopdefault.aspx/tabid-5514/20470_read-47899/.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., Willerslev, E., 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. Mol. Ecol. 21, 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x.

Taberlet, P., Bonin, A., Zinger, L., Coissac, E., 2018. Environmental DNA for Biodiversity REsearch and Monitoring. Oxford University Press.

Tenenhaus, M., Vinzi, V.E., Chatelin, Y.M., Lauro, C., 2005. PLS path modeling. Comput. Stat. Data Anal. 48, 159–205. https://doi.org/10.1016/j.csda.2004.03.005.

Thompson, L.R., et al., 2017. A communal catalogue reveals Earth's multiscale microbial diversity. Nature 551, 457. https://doi.org/10.1038/nature24621.

Tieszen, L.L., Johnson, P.L., 1968. Pigment structure of some arctic tundra communities. Ecology 49, 370. https://doi.org/10.2307/1934476.

Trujillo, M.E., et al., 2014. Genome features of the endophytic actinobacterium micromonospora lupini strain lupac 08: on the process of adaptation to an endophytic life style? PLoS One 9. https://doi.org/10.1371/journal.pone.0108522.

Ustin, S.L., Gamon, J.A., 2010. Remote sensing of plant functional types. New Phytol. 186, 795–816. https://doi.org/10.1111/j.1469-8137.2010.03284.x.

Wang, J.J., et al., 2022. Embracing mountain microbiome and ecosystem functions under global change. New Phytol. 234, 1987–2002. https://doi.org/10.1111/nph.18051.

Weiss, S., et al., 2017. Normalization and microbial differential abundance strategies depend upon data characteristics. Microbiome 5. https://doi.org/10.1186/s40168-017-0237-y.

Yeager, C.M., et al., 2017. Polysaccharide degradation capability of actinomycetales soil isolates from a semiarid grassland of the Colorado plateau. Appl. Environ. Microbiol. 83 https://doi.org/10.1128/aem.03020-16.

Yilmaz, P., et al., 2011. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. Nat. Biotechnol. 29, 415–420. https://doi.org/10.1038/nbt.1823.

Zhang, X.Y., et al., 2003. Monitoring vegetation phenology using MODIS. Remote Sens. Environ. 84, 471–475. https://doi.org/10.1016/s0034-4257(02)00135-9.