Deciphering landscape-scale plant cover and biodiversity from soil eDNA

Goodall, T. ORCID: https://orcid.org/0000-0002-1526-4071, Griffiths, R. I., Gweon, H. S. ORCID: https://orcid.org/0000-0002-6218-6301, Norton, L., Busi, S. B. and Read, D. S. ORCID: https://orcid.org/0000-0001-8546-5154 (2025)
Deciphering landscape-scale plant cover and biodiversity from soil eDNA. Environmental DNA, 7 (5). e70191. ISSN 2637-4943 doi: 10.1002/edn3.70191 Available at https://centaur.reading.ac.uk/125035/

Environmental DNA

WILEY

Environmental DNA
Open Access
Dedicated to the study and use of environmental DNA for basic and applied sciences

ORIGINAL ARTICLE OPEN ACCESS

Deciphering Landscape-Scale Plant Cover and Biodiversity From Soil eDNA

Tim Goodall[1] | Robert I. Griffiths[2] | Hyun S. Gweon[3] | Lisa Norton[4] | Susheel Bhanu Busi[1] | Daniel S. Read[1]

[1]UK Centre for Ecology & Hydrology (UKCEH), Wallingford, UK | [2]Bangor University, Bangor, UK | [3]University of Reading, Reading, UK | [4]UK Centre for Ecology & Hydrology (UKCEH), Lancaster, UK

Correspondence: Tim Goodall (timgoo@ceh.ac.uk)

Received: 15 January 2025 | **Revised:** 26 August 2025 | **Accepted:** 8 September 2025

Funding: This project was funded by UKCEH under the ASSIST programme (NERC Reference: NE/N018125/1) and BBSRC Institute Strategic Programme: Decoding Biodiversity (DECODE) (BBSRC Reference: BB/X020037/1).

Keywords: DNA sequencing | landscape biology | metabarcoding | molecular ecology | plant biodiversity | plant ITS | soil eDNA

ABSTRACT

Biodiversity surveys are critical for detecting environmental change; however, undertaking them at scale and capturing all available diversity through observation is challenging and costly. This study evaluated the potential of soil-extracted eDNA to describe plant communities and compared these findings to traditional, observation-based field surveys. We analyzed 789 soil samples using high-throughput amplicon sequencing and compared DNA-based diversity metrics, indicator taxa, predicted vegetation class, and plant cover in a comparison with co-located field survey data. The results indicated that taxonomically aggregated (genus) eDNA-derived data, while showing slightly reduced Shannon's diversity scores, yielded remarkably similar overall richness and composition estimates. However, the DNA indicator taxa and predictive power for vegetation community classification were also lower overall than those recorded by the field survey. In many cases, plant cover could be inferred from amplicon abundance data with some accuracy despite widely differing scales of sampling—0.25 g crumb of soil versus a 1 m^2 quadrat. Overall, results from eDNA demonstrated lower sensitivity but were broadly in accordance with traditional surveys, with our findings revealing comparable taxonomic resolution at the genus level. We demonstrate the potential and limitations of a simple molecular method to inform landscape-scale plant biodiversity surveys, a vital tool in the monitoring of land use and environmental change.

1 | Introduction

In the face of accelerating climate change, landscape-scale monitoring of biodiversity is essential for detecting shifts in species composition, ecosystem function, and habitat health (Belaire et al. 2022), which can have profound implications for ecosystem resilience and the services upon which human societies rely. Landscape-scale temporal biodiversity inventories, when combined with comprehensive environmental metadata, allow the observation of ecosystem trends and are vital tools for understanding and estimating habitat change, informing models, and enabling the forecast of future change (Franklin et al. 2016). Such models can be used to estimate the response of ecosystems and their services to anthropogenic impacts, the effects of climatic change, or even government policy (Wood et al. 2017). Arguably, because of the increasing rate of change and our dependence on ecosystem services, monitoring the natural environment through biodiversity assessment has never been as important, yet monitoring is often patchy or piecemeal in coverage and is dependent upon resource availability and legislative drivers (Donaldson et al. 2017).

Conventional approaches for national, landscape-scale, biodiversity surveys require an expansive group of field ecologists

to visit many hundreds of locations to perform consistent and accurate measurements of plant species (Carey et al. 2008; Maskell et al. 2008). Because these surveys offer a snapshot of visible plant growth, dormant, ephemeral, or cryptic plants may not be observed (Hiiesalu et al. 2012). Resource availability, both human and financial, is a limiting determinant of survey scale and sampling intensity. New methods to maximize efficiency, facilitate greater sampling depth, and increase the scale of surveys are desirable. At large scale, methods such as drone (Cruzan et al. 2016), aircraft or satellite remote sensing are useful tools (Sharma et al. 2017), however, their resolution is limited. Environmental DNA (eDNA) monitoring offers the potential to simplify the field effort required and enable the processing of vast numbers of samples with high taxonomic resolution (Ruppert et al. 2019).

eDNA analysis has provided insight into many otherwise difficult-to-monitor environments and assists in estimating biodiversity and distribution of both micro-organisms and, more recently, macro-organisms (Ruppert et al. 2019; Thomsen and Willerslev 2015). The accuracy of eDNA for plant community analysis from soil is relatively novel and untested at scale. Fahner et al. (2016) assessed a suite of plant taxonomic markers in 35 forest soils, upon which we built herein by applying the best-performing taxonomic markers to examine the ability of soil eDNA to represent local plant communities from different soils across a national landscape. In this study, we extracted eDNA from 789 soil samples collected as part of the UK Centre for Ecology and Hydrology (UKCEH) Countryside Survey of 2007 (Carey et al. 2008), where each sampling location was simultaneously subjected to vegetative studies by trained plant ecologists. The limitations of amplicon sequencing to resolve taxonomies below genus level with accuracy (Fahner et al. 2016; Alteio et al. 2021; Odom et al. 2023; Vallin et al. 2025) were taken into account, and we used taxonomies collapsed to genus level to assess how data derived from a high-throughput molecular and classical field survey methods compared to each other and explored the merits and limitations of this eDNA-amplicon approach within the context of a national survey. Specifically, we compared the key indicators of Aggregate Vegetation Classification (AVC) types and examined the potential for molecularly derived abundance data to describe plant cover.

Additionally, we used a machine learning approach to assess the predictive ability of the data to ascribe a sample's AVC—a potentially important use of an eDNA sample in predicting broad habitat within a survey context.

Our findings highlight the merits of each approach and, importantly, inform the potential for molecularly derived methods, specifically amplicon-based soil-eDNA, to describe plant cover, overall biodiversity, and habitat classification.

2 | Materials and Methods

2.1 | Vegetation Survey

Surveyors undertook vegetation surveys as part of the 2007 UK Countryside Survey (https://www.ceh.ac.uk/our-science/projects/countryside-survey) following published guidelines (Maskell et al. 2008) (Figure 1). The Countryside Survey is a long-running survey of the UK landscape and is designed to encompass multiple sites from each of the recognized land classes of the UK. For the purposes of this study, we used the 1 m^2 plant species recordings (nest 0). The surveys were conducted at a minimum of one meter and a maximum of 2.5 m distance from the location of the soil sample.

2.2 | Soil Collection

Soil sampling and vegetation surveys were conducted simultaneously. A clean, unused plastic tube of 5 cm diameter was used to collect soil core samples from the top 15 cm at each sample site; cores were sealed into pre-labeled plastic bags to prevent the transfer of soil residue between samples. The cores were transferred to the laboratory on the day of collection and subjected to multiple analyses (Emmett et al. 2008). For molecular work, cores were frozen at −20°C for later processing, where the cores were lightly defrosted, and a sub-sample of soil collected from below the organic horizon (thus excluding fine roots) was homogenized and archived at −20°C for later DNA extraction and plant ITS2 amplicon sequencing.

2.3 | Vegetation Classification

Vegetation communities are closely aligned with habitat definitions and assessments of habitat health. An Aggregate Vegetation Classification (AVC) was applied to each sample site based on the surveyor's plant species cover estimates. AVCs are determined as per Bunce et al. (1999); briefly, plant survey data is lumped and subjected to multivariate analysis using DECORANA and ordinated. Clustering of the sample within one of eight groups forms the basis for the classification, where the eight AVC classes are defined as: (1) "Crops and weeds," (2) "Tall grass and herb," (3) "Fertile grassland," (4) "Infertile grassland," (5) "Lowland wooded," (6) "Upland wooded," (7) "Moorland grass mosaics," and (8) "Heath and bog."

2.4 | Molecular Analyses of Plant ITS2

DNA was extracted from 0.25 g of the archived soil. Briefly, soil was weighed by means of pre-sterilized (immersion in 5% bleach and 70% ethanol wash) apparatus (Figure 1, inset) into Powersoil DNA 384 Isolation Kit (Qiagen Ltd.) plates, and the DNA was extracted according to the manufacturer's instructions. Samples were randomly distributed across extraction plates, and each extraction plate incorporated negative extraction controls. Amplicons were generated using a 2-step amplification approach, with Illumina Nextera tagged ITS2 region primers, forward primer ITS2-S2 ATGCGATACTTGGTGTGAAT and reverse primer ITS4 TCCTCCGCTTATTGATATGC following the approach of Fahner et al. (2016), each primer was modified at the 5′ end with the addition of Illumina pre-adapter and Nextera sequencing primer sequences. Amplicons were generated using high-fidelity DNA polymerase (Q5 Taq; New England Biolabs). After initial denaturation at 95°C for 2 min, the PCR conditions were as follows: denaturation at 95°C for 15 s, annealing at 55°C, annealing for 30 s with extension at

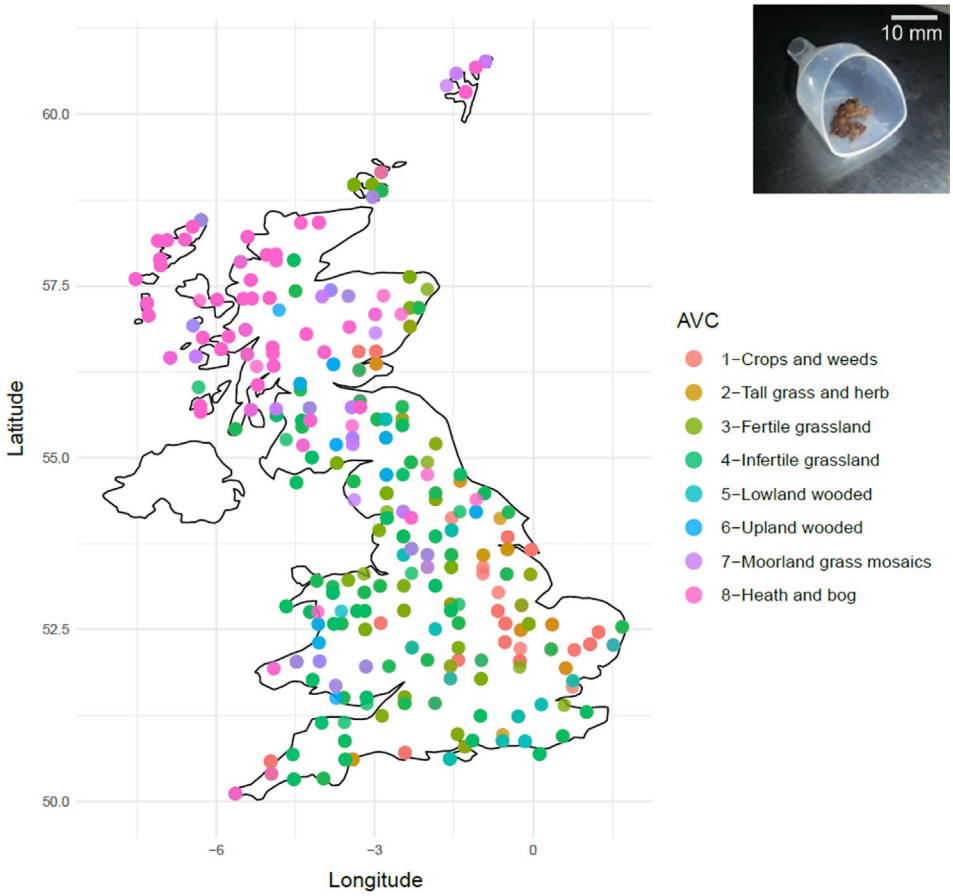

FIGURE 1 | Distribution of survey sites within England, Scotland, and Wales; colored by AVC classification. For data protection site co-ordinates are rounded to the nearest 10 km. (*inset*) 0.25 g of homogenized soil prior to DNA extraction.

72°C for 30 s, repeated for 35 cycles. A final extension step of 10 min at 72°C was performed. PCR products were cleaned using a Zymo ZR-96 DNA Clean-up Kit (Zymo Research, US) following the manufacturer's instructions. MiSeq adapters and 8 nt dual-indexing barcode sequences were added during the second PCR amplification step. After an initial denaturation at 95°C for 2 min, the PCR conditions were as follows: denaturation at 95°C for 15 s, annealing at 55°C, annealing for 30 s with extension at 72°C for 30 s, repeated for eight cycles with a final extension of 10 min at 72°C. Both extraction and PCR negative control samples were verified as negative using gel electrophoresis. Amplicon sizes were determined using an Agilent 2200 TapeStation system. Libraries were normalized using the SequalPrep Normalization Plate Kit (Thermo Fisher Scientific), quantified using the Qubit dsDNA HS kit (Thermo Fisher Scientific), and pooled. The pooled library was diluted to 400 pM after denaturation and neutralization. Denaturation was achieved with 0.2N NaOH for 5 min, followed by neutralization with 0.2N HCl. The library was then diluted to a load concentration of 12 pM with HT1 Buffer and a 10% denatured PhiX control library. The final denaturation was performed by heating to 96°C for 2 min, followed by cooling in crushed ice. Sequencing was performed on an Illumina MiSeq using V3 600 cycle reagents. The 789 samples were randomly split into three sequencing libraries, each being sequenced on its own flow cell, and with each flow cell generating more than 17 million raw reads.

2.5 | Molecular Bioinformatics

Illumina demultiplexed sequences for each of the three sequencing libraries were processed independently using HONEYPI (Oliver et al. 2021), a bioinformatics pipeline developed for the processing of ITS2 sequences for the UK National Honey Monitoring scheme (https://github.com/hsgweon/honeypi). Amplicon Sequence Variant (ASV) tables and sequence taxonomies were generated using the standard workflow. Briefly, raw sequences were quality-filtered using DADA2 within HONEYPI, with the parameters: reads truncated at a quality score drop below $Q = 30$, and reads shorter than 150 bp were removed. On average, ~17,605 reads per sample passed the initial quality ($Q \geq 30$) filtering step, with a standard deviation of 13,566 reads per sample. Denoising and ASV inference were performed using default DADA2 parameters as implemented in HONEYPI. Subsequently, taxonomy was assigned using a custom-trained naive Bayesian classifier on an updated NCBI nucleotide (nt; https://www.ncbi.nlm.nih.gov/nucleotide/) database, with assignments based on 97% sequence similarity. The output files included taxonomies and ASV tables. After passing through HONEYPI, taxonomies and ASV tables for the three libraries were merged by ASV sequence using R to generate a composite table for all samples. Samples with fewer than 1000 reads were deemed, through examination of sample read depth frequency, to lie outside the range of normal distribution and

were therefore removed. Samples with greater than 1000 reads ($n = 798$) were maintained, and all samples were rarefied to this read depth for subsequent analysis.

2.6 | Analysis

Tests were conducted to compare (i) diversity measures of genera, (ii) richness of the genera recorded, (iii) AVC indicator genera, and (iv) the AVC predictive accuracy of each survey method.

The number of sites examined totaled 789, of which 126 were of AVC-1 "Crops and weeds," 35 of AVC-2 "Tall grass and herb," 173 of AVC-3 "Fertile grassland," 190 of AVC-4 "Infertile grassland," 19 of AVC-5 "Lowland wooded," 28 of AVC-6 "Upland wooded," 69 of AVC-7 "Moorland grass mosaics," and 149 of AVC-8 "Heath and bog."

All non-flowering plant data were removed from both the ASV and survey tables; thus, bryophytes, algae, gymnosperms, bare ground, leaf litter, and rock were excluded. To minimize the risk of spurious species-level assignments, we conducted all downstream analyses at the genus level. For example, sequences derived from Oil-seed Rape (*Brassica napus*) were variably misassigned as one or both of its parental lineages (*Brassica rapa* and *Brassica oleracea*), necessitating collapsing all *Brassica* detections to the genus level. This conservative approach ensures consistency and avoids introducing artifacts from uncertain species identifications.

Proportional abundances were then calculated for each sample's molecular data using decostand (R package vegan; Oksanen et al. 2015), and each sample's rare genera (< 5% abundance) were removed from each dataset before subsequent comparison and analysis.

To assess the similarities in taxonomic observations between molecular and survey data, each site's AVC classification was used to determine indicator genera (R package labdsv; Roberts 2025), Shannon's diversity, and the richness of genera recorded by each survey method (R package vegan; Oksanen et al. 2015). Data derived from either survey method were determined by the Shapiro test to be non-normally distributed; therefore, non-parametric Spearman's correlations were used to assess the relationships between the survey methods. Statistical comparison of eDNA abundance and 1 m^2 plant cover survey methods: Spearman's Rho statistic was calculated to estimate a rank-based measure of association between the survey methods (base R), with p value adjustments made with the Benjamini–Hochberg method. Specifically, the correlation between the abundance of co-recorded genera at each site within each AVC, as well as the correlation of genus level diversity measures at each site and within each AVC.

We compared the predictive ability of the eDNA abundance and 1 m^2 plant cover surveys in ascribing the sample site's AVC classification using machine learning. To do this, we used the R package xgboost (Chen and Guestrin 2016), an approach to assess each dataset's predictive sensitivity (how many of the actual positive cases we were able to predict correctly), specificity (how many of the correctly predicted cases actually turned out to be positive), and accuracy (how often the classifier correctly predicts) with a 4:1 training to testing split, using settings: method = "xgbTree," tuneGrid = expand.grid (nrounds = c(50, 100), max_depth = c(2, 4, 6), eta = c(0.1, 0.3), gamma = c(0, 1), colsample_bytree = c(0.7), min_child_weight = c(1), subsample = c(0.8)), with a 5-fold cross validation check.

3 | Results

3.1 | Measured Genus Level Diversity and Richness and Correlation Between Methods

The community-level relative abundance data for flowering plants from each site, collapsed to the genus level, were used to calculate Shannon diversity scores per survey (Figure 2). Score averages were calculated per AVC and the results were, for molecular and 1 m^2 surveys: Crops and weeds: 0.57 (SD 0.48) and 0.16 (SD 0.35), Tall grass and herb: 0.48 (SD 0.37) and 0.69 (SD 0.58), Fertile grassland: 0.65 (SD 0.43) and 0.8 (SD 0.50), Infertile grassland: 0.85 (SD 0.49) and 1.32 (SD 0.49), Lowland wooded: 0.56 (SD 0.44) and 0.78 (SD 0.46), Upland wooded: 0.35 (SD 0.39) and 0.73 (SD 0.52), Moorland grass mosaics: 0.64 (SD 0.43) and 1.22 (SD 0.53), Heath and bog: 0.38 (SD 0.35) and 1.93 (SD 0.43). Across all samples, the average Shannon's diversity scores were 0.61 (SD 0.47) and 0.89 (SD 0.61) for molecular and 1 m^2 surveys, respectively.

Aside from AVC-1 "Crops and weeds," the molecular survey recorded, on average, lower Shannon's diversity scores. Spearman's rank correlation coefficients (Rho) between the survey method's Shannon's scores were: Crops and weeds: 0.00 ($p_{adj} = 0.96$), Tall grass and herb: 0.23 ($p_{adj} = 0.25$), Fertile grassland: 0.31 ($p_{adj} = 0.00$), Infertile grassland: 0.12 ($p_{adj} = 0.17$), Lowland wooded: 0.13 ($p_{adj} = 0.68$), Upland wooded: 0.39 ($p_{adj} = 0.08$), Moorland grass mosaic: 0.28 ($p_{adj} = 0.06$), and Heath and bog: 0.38 ($p_{adj} = 0.00$). Two of the AVCs (Fertile grassland and Heath and bog) demonstrate significant ($p_{adj} \leq 0.05$) association between the two survey methods; however, the Rho coefficients do not exceed 0.38, indicating "weak" or "moderate" associations. Diversity metrics built on community-level relative abundances did not appear to correlate well between the two survey methods.

Similarly, the richness of genera detected by molecular and 1 m^2 surveys for each site was assessed (Figure 2). Average richness scores per site were lower in the molecular survey, except for "Crops and weeds" habitats. The total genera recorded by each method within each AVC were calculated and compared, and the genera that occurred in both survey types were counted as co-recorded. Unique genera values by AVC were, for molecular and 1 m^2 surveys: Crops and weeds: 69 and 38 (27 co-recorded), Tall grass and herb: 42 and 42 (25 co-recorded), Fertile grassland: 70 and 38 (31 co-recorded), Infertile grassland: 88 and 68 (49 co-recorded), Lowland wooded: 28 and 28 (15 co-recorded), Upland wooded: 29 and 39 (17 co-recorded), Moorland grass mosaics: 41 and 42 (24 co-recorded), Heath and bog: 23 and 29 (14 co-recorded). Within the AVC "Crops and weeds," the molecular method records higher richness; in the former, this is likely due to both the residue of previous crops and the detection of ephemeral weeds. The total genera recorded by the molecular method in "Fertile" and "Infertile grassland" was also higher than traditional methods, where surveying dense grass swords may result

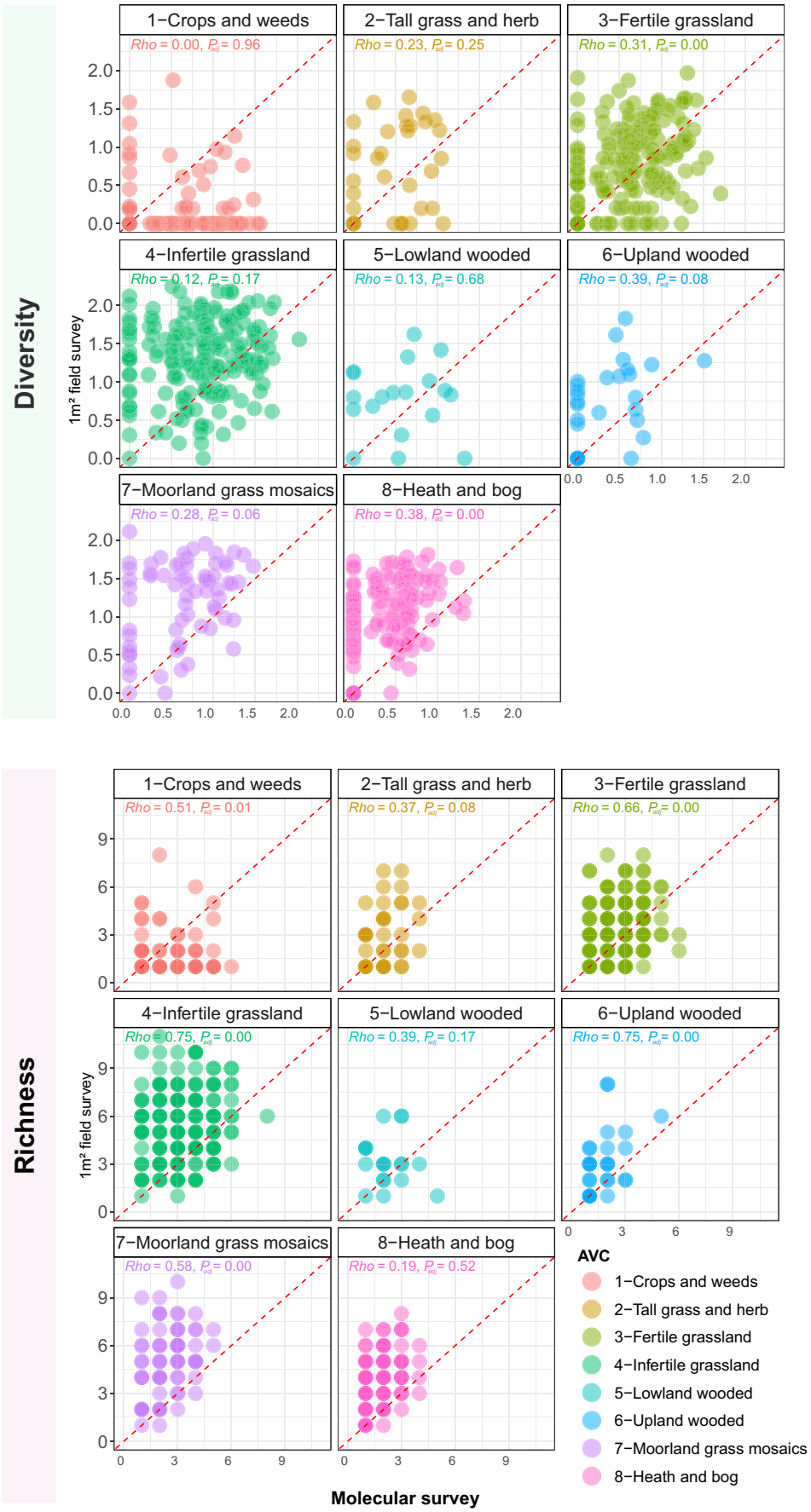

FIGURE 2 | Scatter plots, faceted by AVC, showing each site's genus level diversity score and richness of genera; 1 m^2 plant cover (*X*-axis) and molecular surveys (*Y*-axis) with dotted 1:1 line, Spearman's Rho scores and adjusted *p* values.

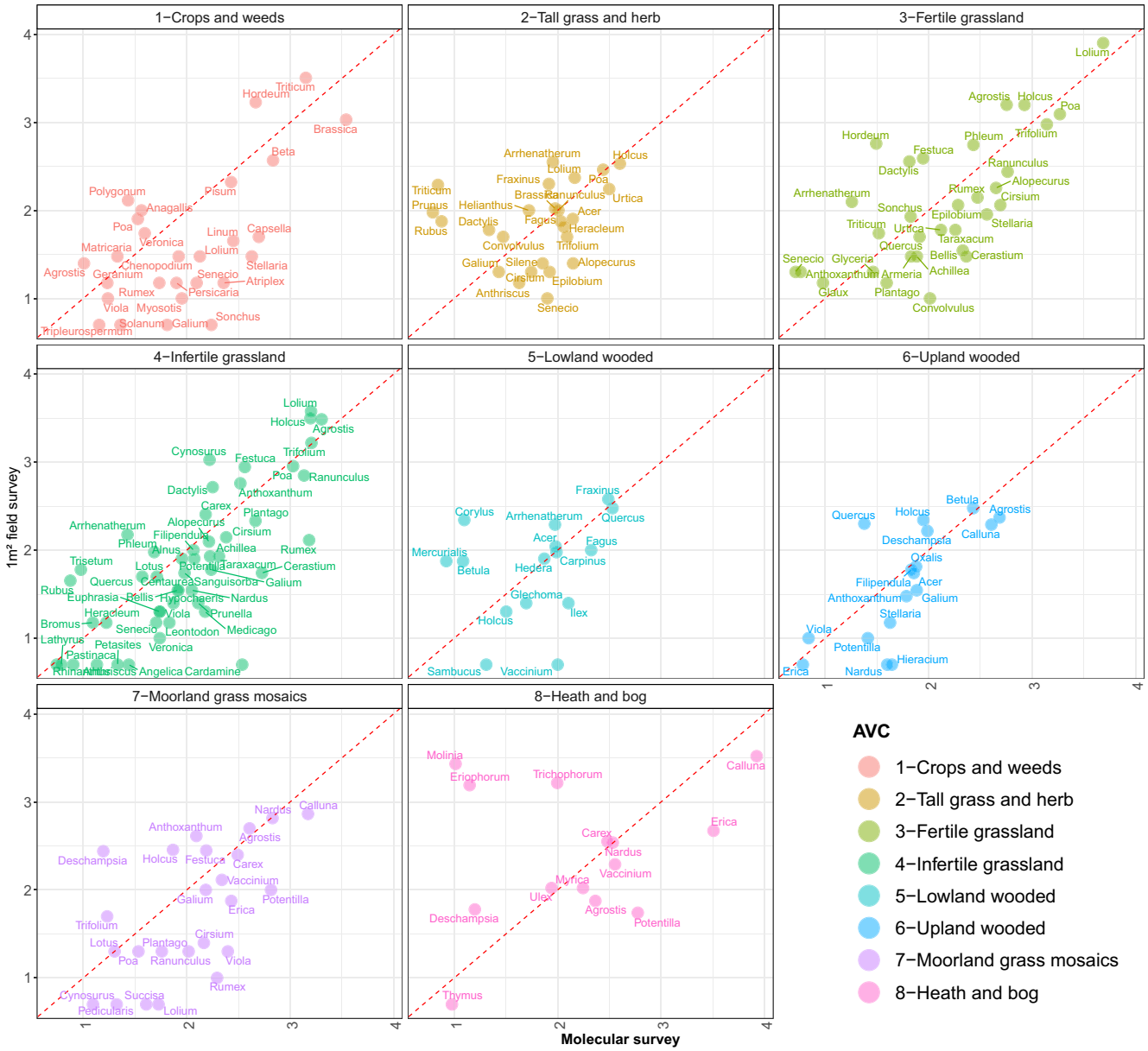

FIGURE 3 | Scatter plots of all co-recorded flowering plant genera across all sites for each AVC habitat. Axis are $\log^{10}$ cumulative percentage of genera's relative abundance (see Methods), *X*-axis molecular, *Y*-Axis 1 m^2 plant cover, red dotted 1:1 line. $\log^{10}$ percentages were used for visualization and clarity.

in lower richness being recorded. Across all samples and all AVCs, the total richness of genera was 158 and 151 for molecular and 1 m^2 surveys, respectively, with 110 genera co-recorded.

3.2 | Co-Recorded Genus Abundance Correlations Between Survey Methods

To assess the ability to predict the amount of flowering plant cover from molecular abundance data, comparisons were made against the 1 m^2 plant survey data. For this assessment, only genera that were co-recorded within each AVC by both survey types were compared using Spearman's rank correlation of the total percentage abundances of each genus. The coefficients of abundance correlations were: Crops and weeds: 0.51 ($p_{adj} = 0.01$), Tall grass and herb: 0.37 ($p_{adj} = 0.08$), Fertile grassland: 0.66 ($p_{adj} = 0.00$), Infertile grassland: 0.75 ($p_{adj} = 0.00$), Lowland wooded: 0.39 ($p_{adj} = 0.17$), Upland wooded: 0.75 ($p_{adj} = 0.00$), Moorland grass mosaic: 0.58 ($p_{adj} = 0.00$), Heath and bog: 0.19 ($p_{adj} = 0.52$). Five of the AVCs (Crops and weeds, Fertile grassland, Infertile grassland, Upland wooded, Moorland grass mosaics) demonstrate significant ($p_{adj} \leq 0.05$) association between the survey methods, with Rho coefficients > 0.50. The abundance associations were "moderate" to "good," indicating that where taxa are co-recorded, abundances are generally well correlated (Figure 3).

3.3 | Genus Level Indicators of Aggregate Vegetation Classification

We assessed the commonality of indicator genera between survey methods to determine whether molecular data could provide relevant taxonomic discriminators of AVC habitats. Indicator

genera with significance ($p < 0.05$) were determined for each AVC from each survey type. Table 1 lists the significant indicators per AVC ranked by the indicator values (indval). Genera names in bold are those recorded as significant indicators by both survey methods. As per the earlier diversity and richness metrics, the molecular data records more indicators of "Crops and weeds" than the 1 m^2 plant survey data, where the common weeds of agricultural or disturbed land are also recorded as indicators (genera Aegilops, Capsella, Vicia, and Atriplex). The molecular data also recorded the same indicator taxa for "Fertile grasslands" and more indicator taxa for "Infertile grasslands." In the remaining AVC classes, the 1 m^2 plant survey data produced more indicators of vegetation type.

3.4 | Inventory of Recorded Genera by Survey Method

A summary of genus-level abundance data is shown in Figure 4. It displays a complete inventory of the flowering plant genera recorded by each survey method, along with summary statistics for each AVC, including the number of sample sites, the total richness of genera recorded and of which the number that are co-recorded, Spearman's correlation scores of abundances, and average AVC Shannon's diversity scores. The abundance scores from each survey method (log^{10} of percentage abundance) provided the basis for the heat map.

3.5 | Aggregate Vegetation Classification Prediction Through Machine Learning

Genus-level abundance data were used to train and validate XGBoost models for the prediction of each sample's AVC classification from either eDNA or 1 m^2 plant cover survey-derived data. Across all AVCs, cross-validation confusion matrix accuracy for molecular eDNA output was 0.61 compared to the 1 m^2 plant cover survey's accuracy of 0.74. The predictive power, measured by sensitivity (how many of the actual positive cases we were able to predict correctly), specificity (how many of the correctly predicted cases actually turned out to be positive), and accuracy (how often the classifier correctly predicts), was broadly comparable between the survey types (Figure 5), with molecular survey data having lower predictive power, most notably for Lowland wooded and Tall grass and herb AVCs. The overall accuracy of the model was improved only slightly (0.64 and 0.75, molecular and 1 m^2 survey respectively, Figure S1) when using data at the highest resolution, i.e., data not collapsed to genus level and 1 m^2 survey data that included measures of bare ground, rock, and bryophytes.

4 | Discussion

Temporal large-scale survey programs encompassing the collection of ecological variables, such as soil state, land use, and plant cover, improve our understanding of the significance, causes, and consequences of large-scale ecological change (Wood et al. 2017). The relative ease of sample collection for eDNA analysis has huge potential for large-scale surveys and citizen science schemes, and could arguably assist in expanding a survey's range, allowing a more comprehensive inventory to be taken (Biggs et al. 2015). Given the lack of similar analyses over such a large geographic area, we hope to highlight the potential caveats of utilizing eDNA-based surveys. These include but are not limited to DNA being a stable biomolecule (Yoccoz et al. 2012), with differential persistence and degradation of DNA presenting possible biases in the data. Moreover, soil properties can influence the sequestration and persistence of biological molecules. Indeed, the stability of DNA in soil has been shown to depend on moisture, temperature, management, exposure to UV, clay particle type and size, and pH (Cai et al. 2006; Strickler et al. 2015). Similarly, post-sampling soil storage can have an effect on DNA therein for similar reasons, and the consistency of the storage method should be applied to all samples within a study (Clasen et al. 2020). Options to this end would include freezing, drying, freeze-drying, or the use of proprietary storage buffers. Other biases, which are particularly pertinent for studies looking for the presence of rare or invasive species, are the relative ease of transmission of eDNA from the site of initial deposition, through a transportive phase (hydrology or disturbance), to a place of persistence (Harrison et al. 2019), ultimately leading to a study site's contamination (Pedersen et al. 2015). Additionally, Jones et al. (2025), in their assessment of sources of uncertainty in DNA metabarcoding, found that differing sequence read depth and even processing at differing laboratories could result in differences in family level richness results. It is also important to note that traditional plant surveys are not immune from inaccuracies. Plant characteristics (such as small size, rarity, ephemerality, and morphological confusion), along with environmental factors (such as topography and inclemency) and observer variability, contribute and combine to introduce significant variation and bias, which should ideally be quantified as a quality indicator (Morrison 2016; Ullerud et al. 2018; Verheyen et al. 2018). It is imperative to note that both traditional and eDNA-based survey methods are susceptible to bias and inaccuracy, highlighting the need for future research focusing on clear guidelines, standardization, and quality assurance schemes (Jones et al. 2025) to ensure accuracy and permit temporal comparability.

DNA has been shown to persist in temperate soils at very detectable levels for > 60 years (Yoccoz et al. 2012; Foucher et al. 2020), and it is this persistence that may help explain the high diversity measures for AVC-1 "crops and weeds," where the legacy DNA of previous crops and weeds enhances the diversity of plants described by the eDNA method. This persistence raises some questions regarding the time scale needed to detect vegetation change; that is, the sensitivity to land use change. Recently, Foucher et al. (2020) examined eDNA from soil samples in plots for which the crop rotation history was documented and found that the last grown crop formed the dominant taxa in the eDNA inventory, alongside variable detection of past crops up to 8 years, with relic-eDNA from historic grape-vines also present. Similarly, Ariza et al. (2023) found that eDNA, collected from the top 11 cm of forest soil, contained taxa from inventories up to 30 years previous, though inventories most closely matched the contemporary taxa composition. Detection of legacy eDNA may also be advantageous, as it negates the effect of ephemeral genera, those that are small and easily overlooked, or those that are particularly difficult to identify. Indeed, habitat reconstruction through examination of ancient eDNA can facilitate our understanding of lost landscapes, land use change, and migrations

TABLE 1 | Aggregate vegetative class indicator genera determined by molecular and 1 m^2 plant cover surveys.

	Molecular survey				1 m^2 field survey			
AVC	**Rank**	**Genera**	**Indval**	***p***	**Rank**	**Genera**	**Indval**	***p***
1. Crops and weeds	1	**Brassica***	0.381	0.001	1	**Triticum***	0.263	0.001
	2	**Triticum***	0.249	0.001	2	**Hordeum***	0.114	0.003
	3	Aegilops	0.101	0.001	3	**Brassica***	0.076	0.008
	4	Capsella	0.079	0.003	4	**Beta***	0.041	0.016
	5	**Hordeum***	0.072	0.003				
	6	Vicia	0.056	0.02				
	7	**Beta***	0.047	0.03				
	8	Atriplex	0.033	0.044				
2. Tall grass and herb	1	**Urtica***	0.115	0.002	1	**Urtica***	0.172	0.001
					2	Arrhenatherum	0.083	0.015
					3	Heracleum	0.044	0.009
					4	Anthriscus	0.043	0.012
					5	Brassica	0.039	0.043
3. Fertile grassland	1	**Lolium***	0.344	0.001	1	**Poa***	0.132	0.003
	2	**Poa***	0.139	0.001	2	**Phleum***	0.116	0.001
	3	**Phleum***	0.049	0.029	3	**Lolium***	0.503	0.001
4. Infertile grassland	1	**Agrostis***	0.110	0.006	1	**Trifolium***	0.269	0.001
	2	Rumex	0.100	0.002	2	**Agrostis***	0.257	0.001
	3	**Ranunculus***	0.094	0.006	3	**Holcus***	0.249	0.001
	4	**Trifolium***	0.094	0.021	4	**Cynosurus***	0.245	0.001
	5	**Holcus***	0.081	0.014	5	**Ranunculus***	0.128	0.001
	6	Cerastium	0.053	0.042	6	Dactylis	0.079	0.017
	7	**Plantago***	0.050	0.028	7	**Plantago***	0.057	0.017
	8	**Cynosurus***	0.047	0.023				
5. Lowland wooded	1	**Quercus***	0.217	0.001	1	Rubus	0.253	0.001
	2	**Fraxinus***	0.203	0.001	2	**Fraxinus***	0.227	0.001
	3	**Ilex***	0.143	0.001	3	**Hedera***	0.190	0.001
	4	**Hedera***	0.140	0.001	4	**Quercus***	0.166	0.001
	5	Fagus	0.139	0.001	5	**Mercurialis***	0.126	0.001
	6	Arrhenatherum	0.063	0.009	6	Crataegus	0.124	0.001
	7	**Carpinus***	0.048	0.033	7	**Corylus***	0.123	0.001
	8	**Mercurialis***	0.048	0.029	8	**Glechoma***	0.095	0.001
	9	**Sambucus***	0.048	0.026	9	**Ilex***	0.095	0.003
	10	**Glechoma***	0.047	0.026	10	Viola	0.088	0.003
	11	**Corylus***	0.045	0.009	11	**Acer***	0.058	0.008
	12	**Acer***	0.041	0.026	12	Elytrigia	0.054	0.018
	13	Nicotiana	0.039	0.03	13	**Sambucus***	0.048	0.024
	14	Tripleurospermum	0.037	0.046	14	Ballota	0.048	0.017
					15	Ulmus	0.048	0.023
					16	Rosa	0.048	0.033
					17	**Carpinus***	0.048	0.026
					18	Circaea	0.032	0.042

(Continues)

TABLE 1 | (Continued)

AVC	Molecular survey				1 m² field survey			
	Rank	Genera	Indval	*p*	Rank	Genera	Indval	*p*
6. Upland wooded	1	**Betula***	0.181	0.001	1	Deschampsia	0.101	0.006
	2	Ficaria	0.050	0.005	2	Oxalis	0.086	0.001
					3	**Betula***	0.081	0.004
					4	Ulex	0.043	0.014
7. Moorland grass mosaics	1	**Potentilla***	0.194	0.001	1	**Nardus***	0.343	0.001
	2	**Nardus***	0.176	0.001	2	Anthoxanthum	0.300	0.001
	3	**Carex***	0.090	0.007	3	Juncus	0.256	0.001
	4	Viola	0.073	0.011	4	**Carex***	0.147	0.001
					5	**Potentilla***	0.134	0.001
					6	Festuca	0.095	0.015
					7	Galium	0.056	0.022
					8	Vaccinium	0.052	0.026
					9	Sagina	0.029	0.034
8. Heath and bog	1	**Calluna***	0.450	0.001	1	Eriophorum	0.477	0.001
	2	**Erica***	0.381	0.001	2	Trichophorum	0.434	0.001
					3	**Calluna***	0.429	0.001
					4	Molinia	0.365	0.001
					5	**Erica***	0.223	0.001
					6	Narthecium	0.129	0.001

Note: The indicators are listed per AVC habitat and ranked by indicator value (indval) with significance value (*p*). Genera names given in bold with an asterisk are those recorded as significant indicators by both survey methods.

(Pedersen et al. 2015; Haile et al. 2007). To improve the recovery and accuracy of our findings, it is plausible that samples collected above-ground, that is, surface-collected soil samples, may better align with the visual quadratic botanical surveys. Particularly, by capturing signals from plant presence such as pollen, leaf litter, or seeds, this approach may also reduce any temporal lag associated with deeper soil layers. Furthermore, above-ground sampling may enhance the spatial resolution of eDNA surveys, offering promising avenues for large-scale, harmonized vegetation monitoring in the future.

In this study, we chose ITS2 over a multimarker approach, based on previous findings by Fahner et al. (2016) who reported ITS2 in particular, alongside rbcL, was effective in plant metabarcoding for ecological monitoring. A single marker approach has some taxonomic limitations yet currently offers a more cost-effective approach for large-scale surveys. Importantly, we developed a curated, in-house workflow, and therefore chose to use our bioinformatics pipeline HONEYPI (Oliver et al. 2021) to ensure reproducibility and comparability of sequence processing. A caveat of our workflow is that the taxonomic classification utilizes an open, continuously updated NCBI database. While providing simplicity, worldwide applicability, and the ability to utilize the most recent genomic accessions, this approach also has the potential to introduce uncertainty, particularly at the species level or for rare taxa (Blackman et al. 2023). More recently, Jones et al. (2021) produced a curated database of UK flowering plants, which could increase the reliability of taxonomic assignments, while limiting geographic scope. However, the issue of the limited resolution of short-read amplicon sequencing remains (Fahner et al. 2016; Alteio et al. 2021; Odom et al. 2023), as observed by Vallin et al. (2025), taxonomic discrepancies were evident in our dataset at the species level, but unlike their study, where manual resolution was feasible, the frequency of conflicting assignments in our dataset was much higher. To account for this, we collapsed our data to genus level to ensure comparisons were reliable, reproducible, robust, and applicable beyond the UK.

On average, the eDNA molecular approach used in this study recorded lower genus level diversity scores and richness scores. Similarly, the eDNA approach produced fewer vegetation type indicators. However, within the "Crops and weeds" vegetation type, eDNA recorded greater richness and more indicator taxa. It is worth noting that within "Crops and weeds," no correlation was observed for diversity, indicating that eDNA methods are likely recording taxa through the presence of relic-eDNA. This detection of old DNA could be crucial for assessing cropping histories or land use change, but it also highlights the potential to record biases in soil eDNA methods due to the variables associated with eDNA persistence within the soil matrix. The molecular data performed well in producing indicator taxa for "Fertile" and "Infertile" grassland habitats, highlighting more indicators for "Infertile grasslands," possibly due to the increased difficulty of surveying thoroughly through dense grass swards.

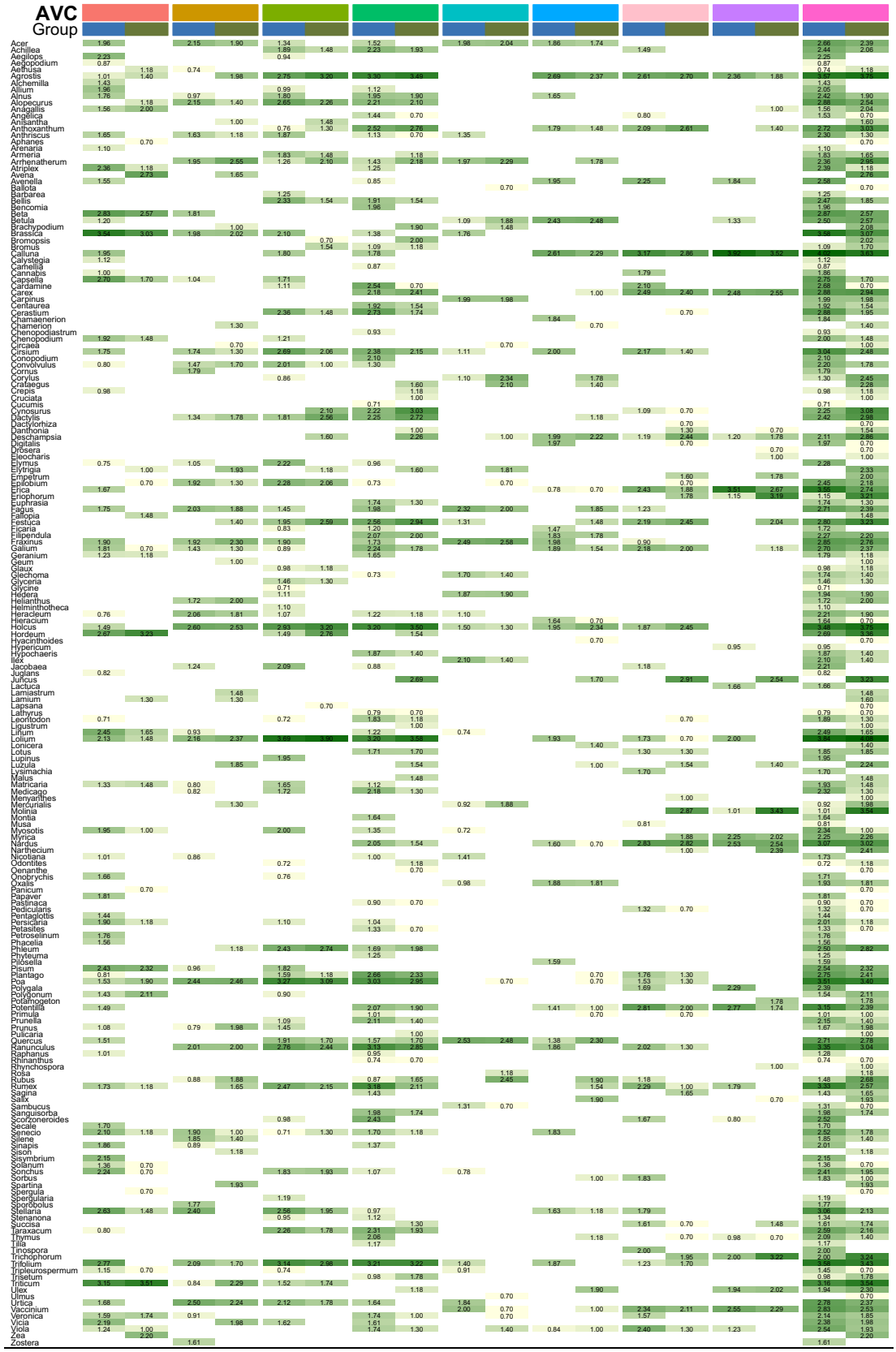

FIGURE 4 | Legend on next page.

FIGURE 4 | Flowering plant genera inventory and heat map showing occurrence and abundance of all genera recorded by Molecular and 1 m^2 plant cover surveys, with summary statistics for each AVC, including the number of sample sites, the total richness of genera recorded and of which the number that are co-recorded, Spearman's correlation scores of abundances and average AVC Shannon's diversity scores. The abundance scores from each survey method (log^{10} of percentage abundance) provide the basis for the heat-map. AVC classes: (1) "Crops and weeds," (2) "Tall grass and herb," (3) "Fertile grassland," (4) "Infertile grassland," (5) "Lowland wooded," (6) "Upland wooded," (7) "Moorland grass mosaics," and (8) "Heath and bog." log^{10} percentages were used for visualization and clarity.

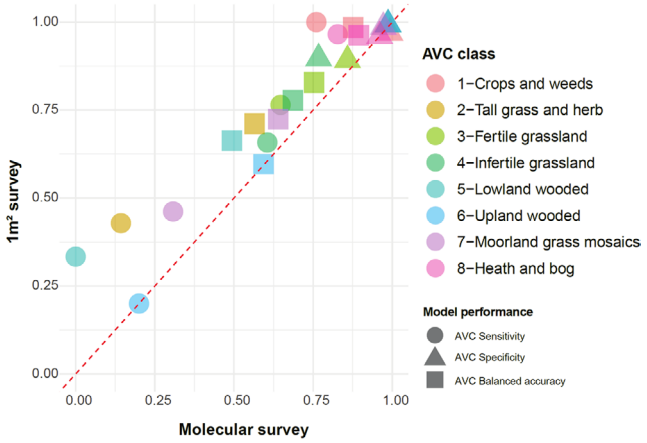

FIGURE 5 | XGBoost model for AVC assignment from genera data—scatter plot showing degree of predictability of molecular (*X*-axis) and 1 m^2 plant cover (*Y*-axis) surveys in AVC assignment. Point color = AVC classification; point shape = model's predictive sensitivity (circle), specificity (triangle), and accuracy (square).

Using the co-recorded genera of each AVC, we examined the potential of using amplicon-derived relative abundance data to estimate plant cover, a tall ask given the possibility of ploidy, copy number variation, and PCR amplification biases (length and GC content) (Álvarez and Wendel 2003). Given these concerns, we were surprised that for the majority of vegetation types, the relationship between molecular-derived relative abundance and plant survey cover was moderate to good, demonstrating that it is possible to infer the plant cover of many genera from eDNA abundance data on the proviso that the habitat and genera are appropriate.

To assess the ability to predict AVC from genus-level data, we applied a machine learning approach to each survey method and found through cross validation that, overall, the predictive performance was broadly lower for the molecular survey. In particular, lower sensitivity was observed for Lowland wooded, Tall grass and herb, and Moorland grass mosaic AVCs, where low Spearman's correlation scores were also recorded. Further, we observed that using data at the highest possible resolution, i.e., species level, led to marginal improvements in the overall accuracy of either survey type.

In concordance with the findings reported by Vasar et al., our molecular results are *broadly* in agreement with those of traditional surveys; however, the level of agreement varied by habitat. In this study, across all AVC classes the molecular survey was able to detect 73% of angiosperm genera recorded by the traditional survey method, a result in line with observations by Ariza et al. (2023) who recorded a molecular coverage of 60%, albeit using samples from a limited range of habitat types. The reduced sensitivity of the molecular data is perhaps unsurprising given the extremely small sample size. Importantly, a key aspect of our study that is perhaps surprising is the ability of the eDNA from 0.25 g of soil to describe the immediate aboveground vegetation with some accuracy. Increasing the sample size, pooling from a larger area, or increasing the sample number (Alsos et al. 2018) would increase the resolving power of the eDNA method. However, there are undesirable cost (increase) and throughput (decrease) implications that need to be weighed carefully. Once collected and stored, soil eDNA can be utilized to examine a range of taxonomic and ecological profiles (Deiner et al. 2017). The relative ease of sample collection and processing, and the large amount of genetic information held within soil eDNA makes surveys based on high-throughput eDNA methods very appealing.

The results presented here demonstrate that flowering plant communities and habitats can be described with a degree of accuracy using molecular methods based on small soil-eDNA samples, and complementarily to field survey, a molecular approach can provide useful and valuable ecological insights. While our molecular results are promising, they primarily highlight the supplementary and synergistic value of molecular techniques alongside traditional methods of biodiversity assessment. Collectively, these findings are critical for understanding the efficiencies and disparities between traditional and molecular biomonitoring methodologies. Our national-scale analysis provides insights that highlight the utility of these methods, both traditional and molecular, for large-scale ecological biodiversity measurements.

4.1 | Maximizing the Value of eDNA for Long-Term Biodiversity Monitoring

Building on these findings, we recommend that future biodiversity monitoring adopt an integrated molecular–traditional framework. eDNA metabarcoding enables rapid and cost-effective detection of taxa at scale, while traditional field surveys remain critical for validation and providing ecological context. Future versions of the UK Countryside Survey could implement a hybrid design, combining extensive eDNA-based assessments across all sites with targeted resampling via traditional surveys at a subset of locations to ground-truth molecular detections and calibrate species-level identifications. Furthermore, coordinated efforts to curate high-quality, habitat-specific reference libraries for UK flora to reduce taxonomic misassignments and increase molecular resolution could be developed. Improvements in sampling design, including replication and strategic pooling, alongside standardization of protocols across surveys, would further improve accuracy and comparability. Finally, incorporating molecular approaches into long-term monitoring frameworks

would allow us to better capture temporal and spatial biodiversity patterns, enhance comparability with historical datasets, and maximize the value of biodiversity data for long-term ecological monitoring and policy applications.

Author Contributions

T.G.: Conception and design of the study; acquisition, analysis, and interpretation of data; writing of the manuscript. R.I.G.: Conception and design of the study; analysis and interpretation of data; writing of the manuscript. H.S.G.: conception of the study; writing of the manuscript. L.N.: conception of the study; writing of the manuscript. S.B.B.: acquisition, analysis, and interpretation of the data; writing of the manuscript. D.S.R.: conception of the study; interpretation of the data; writing of the manuscript.

Acknowledgments

This project was funded by UKCEH under the ASSIST programme (NERC Reference: NE/N018125/1), and the BBSRC Institute Strategic Programme: Decoding Biodiversity (DECODE) (BBSRC Reference: BB/X020037/1).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Raw sequence files are available via the NCBI Sequence Read Archive under BioProject ID PRJNA1201089. R scripts and associated files were uploaded to Zenodo. The files are publicly available and can be accessed at 10.5281/zenodo.14644160.

References

Alsos, I. G., Y. Lammers, N. G. Yoccoz, et al. 2018. "Plant DNA Metabarcoding of Lake Sediments: How Does It Represent the Contemporary Vegetation." *PLoS One* 13: e0195403. https://doi.org/10.1371/journal.pone.0195403.

Alteio, L. V., J. Séneca, A. Canarini, et al. 2021. "A Critical Perspective on Interpreting Amplicon Sequencing Data in Soil Ecological Research." *Soil Biology and Biochemistry* 160: 108357. https://doi.org/10.1016/j.soilbio.2021.108357.

Álvarez, I., and J. F. Wendel. 2003. "Ribosomal ITS Sequences and Plant Phylogenetic Inference." *Molecular Phylogenetics and Evolution* 29: 417–434. https://doi.org/10.1016/S1055-7903(03)00208-2.

Ariza, M., B. Fouks, Q. Mauvisseau, R. Halvorsen, I. G. Alsos, and H. J. de Boer. 2023. "Plant Biodiversity Assessment Through Soil eDNA Reflects Temporal and Local Diversity." *Methods in Ecology and Evolution* 14: 415–430. https://doi.org/10.1111/2041-210X.13865.

Belaire, J. A., C. Higgins, D. Zoll, et al. 2022. "Fine-Scale Monitoring and Mapping of Biodiversity and Ecosystem Services Reveals Multiple Synergies and Few Tradeoffs in Urban Green Space Management." *Science of the Total Environment* 849: 157801. https://doi.org/10.1016/j.scitotenv.2022.157801.

Biggs, J., N. Ewald, A. Valentini, et al. 2015. "Using eDNA to Develop a National Citizen Science-Based Monitoring Programme for the Great Crested Newt (*Triturus cristatus*)." *Biological Conservation* 183: 19–28. https://doi.org/10.1016/j.biocon.2014.11.029.

Blackman, R. C., J.-. C. Walser, L. Rüber, et al. 2023. "General Principles for Assignments of Communities From eDNA: Open Versus Closed Taxonomic Databases." *Environmental DNA* 5: 326–342. https://doi.org/10.1002/edn3.382.

Bunce, R. G. H., C. J. Barr, M. K. Gillespie, et al. 1999. *Vegetation of the British Countryside—The Countryside Vegetation System. ECOFACT Volume 1*. 224pp. UKCEH.

Cai, P., Q. Huang, X. Zhang, and H. Chen. 2006. "Adsorption of DNA on Clay Minerals and Various Colloidal Particles From an Alfisol." *Soil Biology and Biochemistry* 38: 471–476. https://doi.org/10.1016/j.soilbio.2005.05.019.

Carey, P. D., S. Wallis, P. M. Chamberlain, et al. 2008. *Countryside Survey: UK Results From 2007*. NERC/Centre for Ecology & Hydrology. 105pp.

Chen, T., and C. Guestrin. 2016. "XGBoost: A Scalable Tree Boosting System." In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785.

Clasen, L. A., A. P. Detheridge, J. Scullion, and G. W. Griffith. 2020. "Soil Stabilisation for DNA Metabarcoding of Plants and Fungi. Implications for Sampling at Remote Locations or via Third-Parties." *Metabarcoding and Metagenomics* 4: e58365. https://doi.org/10.3897/mbmg.4.58365.

Cruzan, M. B., B. G. Weinstein, M. R. Grasty, et al. 2016. "Small Unmanned Aerial Vehicles (Micro-UAVs, Drones) in Plant Ecology." *Applications in Plant Sciences* 4: 1600041. https://doi.org/10.3732/apps.1600041.

Deiner, K., H. M. Bik, E. Mächler, et al. 2017. "Environmental DNA Metabarcoding: Transforming How We Survey Animal and Plant Communities." *Molecular Ecology* 26: 5872–5895. https://doi.org/10.1111/mec.14350.

Donaldson, L., R. J. Wilson, and I. M. D. Maclean. 2017. "Old Concepts, New Challenges: Adapting Landscape-Scale Conservation to the Twenty-First Century." *Biodiversity and Conservation* 26: 527–552. https://doi.org/10.1007/s10531-016-1257-9.

Emmett, B. A., Z. L. Frogbrook, P. M. Chamberlain, et al. 2008. *CM Wood CS Technical Report No.3/07 Soils Manual*. UK Centre for Ecology and Hydrology.

Fahner, N. A., S. Shokralla, D. J. Baird, and M. Hajibabaei. 2016. "Large-Scale Monitoring of Plants Through Environmental DNA Metabarcoding of Soil: Recovery, Resolution, and Annotation of Four DNA Markers." *PLoS One* 11: e0157505. https://doi.org/10.1371/journal.pone.0157505.

Foucher, A., O. Evrard, G. F. Ficetola, et al. 2020. "Persistence of Environmental DNA in Cultivated Soils: Implication of This Memory Effect for Reconstructing the Dynamics of Land Use and Cover Changes." *Scientific Reports* 10: 10502. https://doi.org/10.1038/s41598-020-67452-1.

Franklin, J., J. M. Serra-Diaz, A. D. Syphard, and H. M. Regan. 2016. "Global Change and Terrestrial Plant Community Dynamics." *Proceedings of the National Academy of Sciences* 113: 3725–3734. https://doi.org/10.1073/pnas.1519911113.

Haile, J., R. Holdaway, K. Oliver, et al. 2007. "Ancient DNA Chronology Within Sediment Deposits: Are Paleobiological Reconstructions Possible and Is DNA Leaching a Factor?" *Molecular Biology and Evolution* 24: 982–989. https://doi.org/10.1093/molbev/msm016.

Harrison, J. B., J. M. Sunday, and S. M. Rogers. 2019. "Predicting the Fate of eDNA in the Environment and Implications for Studying Biodiversity." *Proceedings of the Royal Society B: Biological Sciences* 286: 20191409. https://doi.org/10.1098/rspb.2019.1409.

Hiiesalu, I., M. Opik, M. Metsis, et al. 2012. "Plant Species Richness Belowground: Higher Richness and New Patterns Revealed by Next-Generation Sequencing." *Molecular Ecology* 21: 2004–2016. https://doi.org/10.1111/j.1365-294X.2011.05390.x.

Jones, J. I., A. Arnold, D. Buchner, et al. 2025. "Sources of Uncertainty in DNA Metabarcoding of Whole Communities: Implications for Its Use

in Biomonitoring." *Methods in Ecology and Evolution* 16: 1658–1673. https://doi.org/10.1111/2041-210X.70093.

Jones, L., A. D. Twyford, C. R. Ford, et al. 2021. "Barcode UK: A Complete DNA Barcoding Resource for the Flowering Plants and Conifers of the United Kingdom." *Molecular Ecology Resources* 21: 2050–2062. https://doi.org/10.1111/1755-0998.13388.

Maskell, L. C., L. R. Norton, S. M. Smart, et al. 2008. *CS Technical Report No.2/07 Vegetation Plots Handbook*. UK Centre for Ecology and Hydrology.

Morrison, L. W. 2016. "Observer Error in Vegetation Surveys: A Review." *Journal of Plant Ecology* 9: 367–379. https://doi.org/10.1093/jpe/rtv077.

Odom, A. R., T. Faits, E. Castro-Nallar, K. A. Crandall, and W. E. Johnson. 2023. "Metagenomic Profiling Pipelines Improve Taxonomic Classification for 16S Amplicon Sequencing Data." *Scientific Reports* 13: 13957. https://doi.org/10.1038/s41598-023-40799-x.

Oksanen, J., F. Guillaume Blanchet, R. Kindt, et al. 2015. "Vegan: Community Ecology Package." *R Package Version 2.2-1* 2: 1–2.

Oliver, A. E., L. K. Newbold, H. S. Gweon, D. S. Read, B. A. Woodcock, and R. F. Pywell. 2021. "Integration of DNA Extraction, Metabarcoding and an Informatics Pipeline to Underpin a National Citizen Science Honey Monitoring Scheme." *MethodsX* 8: 101303. https://doi.org/10.1016/j.mex.2021.101303.

Pedersen, M. W., S. Overballe-Petersen, L. Ermini, et al. 2015. "Ancient and Modern Environmental DNA." *Philosophical Transactions of the Royal Society, B: Biological Sciences* 370: 20130383. https://doi.org/10.1098/rstb.2013.0383.

Roberts, D. 2025. "labdsv: Ordination and Multivariate Analysis for Ecology." *R Package Version 2.1-1*. https://doi.org/10.32614/CRAN.package.labdsv.

Ruppert, K. M., R. J. Kline, and M. S. Rahman. 2019. "Past, Present, and Future Perspectives of Environmental DNA (eDNA) Metabarcoding: A Systematic Review in Methods, Monitoring, and Applications of Global eDNA." *Global Ecology and Conservation* 17: e00547. https://doi.org/10.1016/j.gecco.2019.e00547.

Sharma, R. C., K. Hara, and H. Hirayama. 2017. "A Machine Learning and Cross-Validation Approach for the Discrimination of Vegetation Physiognomic Types Using Satellite Based Multispectral and Multitemporal Data." *Scientifica* 2017: 9806479. https://doi.org/10.1155/2017/9806479.

Strickler, K. M., A. K. Fremier, and C. S. Goldberg. 2015. "Quantifying Effects of UV-B, Temperature, and pH on eDNA Degradation in Aquatic Microcosms." *Biological Conservation* 183: 85–92. https://doi.org/10.1016/j.biocon.2014.11.038.

Thomsen, P. F., and E. Willerslev. 2015. "Environmental DNA—An Emerging Tool in Conservation for Monitoring Past and Present Biodiversity." *Biological Conservation* 183: 4–18. https://doi.org/10.1016/j.biocon.2014.11.019.

Ullerud, H. A., A. Bryn, R. Halvorsen, and L. Ø. Hemsing. 2018. "Consistency in Land-Cover Mapping: Influence of Field Workers, Spatial Scale and Classification System." *Applied Vegetation Science* 21: 278–288. https://doi.org/10.1111/avsc.12368.

Vallin, H., H. Hipperson, J. Titěra, L. Jones, and M. Fraser. 2025. "Comparative Analysis of Pasture Composition: DNA Metabarcoding Versus Quadrat-Based Botanical Surveys in Experimental Grassland Plots." *Ecology and Evolution* 15: e71195. https://doi.org/10.1002/ece3.71195.

Verheyen, K., M. Bažány, E. Chećko, et al. 2018. "Observer and Relocation Errors Matter in Resurveys of Historical Vegetation Plots." *Journal of Vegetation Science* 29: 812–823. https://doi.org/10.1111/jvs.12673.

Wood, C. M., S. M. Smart, R. G. H. Bunce, et al. 2017. "Long-Term Vegetation Monitoring in Great Britain—The Countryside Survey 1978–2007 and Beyond." *Earth System Science Data* 9: 445–459. https://doi.org/10.5194/essd-9-445-2017.

Yoccoz, N. G., K. A. Bråthen, L. Gielly, et al. 2012. "DNA From Soil Mirrors Plant Taxonomic and Growth Form Diversity." *Molecular Ecology* 21: 3647–3655. https://doi.org/10.1111/j.1365-294X.2012.05545.x.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Figure S1:** XGBoost model for AVC assignment from "species" data—scatter plot showing degree of predictability of molecular (*X*-axis) and 1 m^2 plant cover (*Y*-axis) surveys in AVC assignment. Point color = AVC classification; point shape = model's predictive sensitivity (circle), specificity (triangle), and accuracy (square).