

Trait-based Machine-Learning Identifies Functional predictors of Plant Responsiveness to Arbuscular Mycorrhizal Fungi

Tariq Shah

The University of North Carolina at Greensboro

Mohamed Akram Errahmouni

University Mohammed VI Polytechnic

Mohamed Hijri

`mohamed.hijri@umontreal.ca`

University Mohammed VI Polytechnic

Research Article

Keywords: arbuscular mycorrhizal fungi, mycorrhizal growth response, machine learning, Random Forest, plant functional traits, root economics spectrum, Trait2Myco

Posted Date: October 26th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11307374/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Arbuscular mycorrhizal fungi (AMF) are widespread plant symbionts that enhance nutrient acquisition, growth, and stress tolerance, yet plant responsiveness to AMF varies substantially and remains difficult to predict. We developed Trait2Myco, a Random Forest–based machine-learning framework that integrates plant functional traits and environmental variables to predict mycorrhizal growth response (MGR). The framework was trained using 80 observations spanning four grass species grown under four precipitation regimes. Predictor variables included leaf and root economic traits and precipitation, while total MGR served as the response variable. A full model incorporating ten functional traits and precipitation was trained using 80% of the observations and evaluated on an independent testing dataset (20%, $n = 16$). The full model explained 41% of variation in MGR ($R^2 = 0.41$), whereas a reduced ecological model containing only specific leaf area, specific root area, root carbon concentration, and precipitation explained 51.7% of the variation ($R^2 = 0.517$). Ten-fold cross-validation of the full model yielded an R^2 of 0.983, RMSE of 0.046, and MAE of 0.036. Variable importance analysis identified specific leaf area, precipitation, root carbon concentration, and specific root area as the strongest predictors of MGR. Partial dependence analysis revealed nonlinear relationships between these predictors and MGR, while principal component analysis showed that highly responsive plants occupied distinct regions of functional trait space characterized by more resource-acquisitive trait syndromes. Leaf and root traits contributed nearly equally to model performance, accounting for 41.9% and 41.3% of cumulative importance, respectively, while environmental factors contributed 16.8%. An exploratory analysis using an independent AMF colonization dataset produced comparable predictive relationships, suggesting that trait-based machine-learning approaches may provide a useful framework for predicting plant responsiveness to AMF across environmental contexts.

Introduction

Arbuscular mycorrhizal fungi (AMF) form one of the most ancient and widespread mutualistic associations in terrestrial ecosystems, colonizing the roots of a large proportion of vascular plant species and playing a central role in global nutrient cycling (Bonfante et Genre, 2010; Brundrett, 2009; Smith et D.J., 2008; van der Heijden et al., 2015). Through this symbiosis, AMF enhance phosphorus and nitrogen acquisition, improves water relations, increases tolerance to abiotic and biotic stress, and contribute substantially to ecosystem productivity (Begum et al., 2019; Smith et D.J., 2008; van der Heijden et al., 2015). In exchange, host plants allocate a significant fraction of recently fixed carbon to their fungal partners, establishing a reciprocal exchange that can influence both plant performance and belowground community function (Bever et al., 2009; Smith et D.J., 2008; van der Heijden et al., 2015).

Despite the widespread occurrence of AM symbiosis, the magnitude and even the direction of plant growth response to AMF colonization vary substantially among species, genotypes, developmental stages, and environmental conditions (Janos, 1980; Johnson et al., 2008; Klironomos, 2003). This variation is commonly quantified as the mycorrhizal growth response (MGR), defined as the proportional difference in plant performance between mycorrhizal and non-mycorrhizal plants. Responses span a

continuum from strongly positive to neutral to clearly negative (Johnson et al., 2008; Klironomos, 2003), and this variability has made MGR difficult to predict from colonization intensity alone, as colonization and plant growth responses are often decoupled (Hart et Reader, 2002; Smith et Smith, 2011). Identifying the drivers of this variation is a central problem in mycorrhizal ecology, with direct consequences for effectiveness of inoculation-based strategies succeed in agriculture and ecological restoration (Hoeksema et al., 2010; Lehmann et al., 2012; Treseder, 2013).

A range of mechanisms have been proposed to explain variation in MGR, including nutrient supply demand dynamics (Koide, 1991), carbon-cost efficiency of the symbiosis (Corrêa et al., 2011), host phylogeny and fungal community identity (Klironomos, 2003; Maherali et Klironomos, 2007), soil nutrient availability and abiotic context (Hoeksema et al., 2010; Treseder, 2013; Veresoglou et al., 2012), and root system architecture (Hart et Reader, 2002; Wilson et Hartnett, 1998). More recently, plant functional traits have emerged as a promising integrative framework for explaining this variability. The leaf economics spectrum describes a continuum from resource-acquisitive species, characterized by high specific leaf area and rapid nutrient turnover, to resource-conservative species with denser, longer-lived tissue (Reich et Cornelissen, 2014; Wright et al., 2004). An analogous root economics spectrum exists belowground, where root traits such as specific root length, root diameter, and tissue density reflect trade-offs between independent nutrient acquisition and reliance on fungal symbionts (Bergmann et al., 2020; Roumet et al., 2016). Importantly, root economic variation does not collapse onto a single acquisitive–conservative axis: Bergmann et al. (2020) demonstrated that a distinct collaboration gradient, orthogonal to the conservation gradient, captures the degree to which species invest in traits favoring mycorrhizal partnership over independent root foraging. Building on this framework, Chagnon et al. (2013) proposed that both plant and fungal traits can be organized along life-history axes that jointly predict the ecological outcome of the symbiosis, and subsequent work has reinforced the view that leaf and root economic traits jointly shape both the degree of AMF colonization and the magnitude of benefits that hosts derive from the association (Delavaux et al., 2017; Freschet et al., 2021).

Despite this conceptual progress, predicting MGR from trait data in practice remains difficult, because trait effects on mycorrhizal responsiveness are frequently nonlinear, depend on the abiotic environment, and may arise from interactions among multiple traits rather than from any single trait acting in isolation (Chagnon et al., 2013; Hoeksema et al., 2010). Conventional linear statistical models are poorly suited to capturing this complexity and typically require a priori specification of functional form and interaction terms that are rarely known beforehand in ecological trait datasets. Machine-learning algorithms offer a promising alternative: Random Forest models in particular can accommodate nonlinear relationships and complex interactions among predictors, while being relatively robust to multicollinearity and providing quantitative measures of variable importance (Breiman, 2001; Cutler et al., 2007; Prasad et al., 2006). Machine-learning approaches have seen growing use across plant and soil ecology (Cutler et al., 2007; Prasad et al., 2006), yet predictive frameworks specifically designed to identify the functional trait syndromes associated with mycorrhizal responsiveness remain limited.

To address this gap, we developed Trait2Myco, a Random Forest–based machine-learning framework that integrates leaf and root functional traits with an environmental covariate to predict plant responsiveness to AMF. Using 80 observations spanning four plant species and four precipitation regimes, we examined whether leaf and root economic traits could predict variation in MGR, asking specifically: (i) which functional traits contribute most strongly to responsiveness; (ii) whether a reduced set of ecologically meaningful traits can explain a substantial proportion of variation in MGR; and (iii) how highly responsive and weakly responsive plants are distributed within multivariate trait space. To evaluate whether the trait-based relationships identified in the primary dataset extend to a different measure of mycorrhizal association, we additionally applied the framework to an independent dataset relating host growth traits to AMF root colonization intensity (Werner et Kiers, 2015; Zhao et al., 2026). By combining machine learning with ecological interpretation across two datasets, Trait2Myco provides a reproducible approach for exploring trait-mediated variation in plant responses to AMF and for identifying functional characteristics associated with stronger mycorrhizal responses.

Materials and Methods

Development of the Trait2Myco framework

Trait2Myco was developed as a Random Forest-based machine-learning framework to identify plant functional trait patterns associated with responsiveness to AMF. The framework integrates ecological trait data and an environmental covariate to predict MGR and to identify the most influential predictors of responsiveness. Random Forest was selected because it can model nonlinear relationships, accommodates interactions among predictors without requiring their pre-specification, and is relatively robust to the multicollinearity commonly encountered among correlated functional traits (Breiman, 2001; Cutler et al., 2007). The overall workflow of Trait2Myco, including data integration, model development, prediction, variable-importance assessment, and ecological interpretation, is illustrated in Fig. 1.

Primary dataset: sources and preparation

The primary dataset was obtained from the Dryad Digital Repository (Zhao et al. (2026); <https://doi.org/10.5061/dryad.3j9kd51v3>), comprising measurements from four plant species exposed to four precipitation regimes, with five biological replicates per species–precipitation combination (80 observations in total after data intergration). Component data files containing plant biomass and shoot nutrients, mycorrhizal response measurements, and leaf and root functional traits were merged using species identity and replicate number as shared identifiers, producing a single dataset containing trait, environmental, and response variables for every observation. Functional trait and environmental data were used as predictor variables in the Trait2Myco Random Forest framework, which generated predictions of mycorrhizal growth response (MGR) and associated model outputs, including variable importance rankings, partial dependence plots, and trait-space ordinations (Fig. 1).

Functional trait predictors

Trait2Myco incorporated leaf and root traits representing resource-acquisition and resource-conservation strategies, consistent with established leaf and root economics frameworks (Bergmann et al., 2020; Reich et Cornelissen, 2014; Roumet et al., 2016; Wright et al., 2004). Leaf traits included leaf dry matter content, specific leaf area, leaf carbon concentration, and leaf nitrogen concentration. Root traits included root diameter, specific root length, root tissue density, specific root area, root carbon concentration, and root nitrogen concentration. Precipitation treatment was included as an environmental predictor, given the potential influence of water availability on AMF functioning and plant responsiveness (Hoeksema et al., 2010; Veresoglou et al., 2012).

Response variable

Plant responsiveness to AMF was quantified as total mycorrhizal growth response (MGR), defined as the relative change in plant growth attributable to AMF inoculation. MGR was calculated as:

$$MGR = \frac{B_{AMF} - B_{control}}{B_{control}}$$

Where B_{AMF} is plant biomass in AMF-inoculated plants and $B_{control}$ is plant biomass in non-inoculated control plants. Positive MGR values indicate greater biomass production in AMF-inoculated plants, whereas negative values indicate reduced biomass relative to non-inoculated plants. An MGR of zero indicates no difference in biomass between AMF-inoculated and control plants.

MGR was used as the primary response variable rather than colonization intensity because the two are frequently decoupled: a plant can be heavily colonized without a correspondingly large growth benefit, or vice versa (Hart et Reader, 2002; Smith et Smith, 2011). The full Trait2Myco model incorporated ten leaf and root functional traits together with precipitation as predictor variables (Table 1).

Table 1
Functional trait and predictor variables incorporated in the Trait2Myco framework.

Variable	Description	Unit
LDMC	Leaf dry matter content	mg g ⁻¹
SLA	Specific leaf area	cm ² g ⁻¹
LCC	Leaf carbon concentration	%
LNC	Leaf nitrogen concentration	%
RD	Root diameter	mm
SRL	Specific root length	m g ⁻¹
RTD	Root tissue density	g cm ⁻³
SRA	Specific root area	cm ² g ⁻¹
RCC	Root carbon concentration	%
RNC	Root nitrogen concentration	%
Precipitation	Precipitation treatment	mm
MGR	Mycorrhizal growth response	response variable

Random Forest model development

The full Trait2Myco model incorporated all ten functional traits and precipitation as predictors of total MGR. Models were fitted using 1,000 trees, with variable importance estimation enabled, using the randomForest package in R (Liaw et Wiener, 2002). The model was fitted using the training dataset only, with the held-out testing dataset excluded from model fitting and predictor selection.

Variable importance analysis

Predictor importance was quantified as the percentage increase in mean squared error (%IncMSE) under permutation-based variable importance analysis, in which predictors producing larger increases in error following permutation were interpreted as having greater relative importance for model prediction. Variable importance was calculated from the training dataset and was used to identify the predictors retained in the reduced ecological model. The held-out testing dataset was not used for predictor selection. Specific leaf area, precipitation, root carbon concentration, and specific root area emerged as the four most important predictors in the training dataset.

Reduced ecological model

To identify a minimal, biologically interpretable predictor set, a second, reduced model was constructed using only the four variables identified as most important in the training dataset: specific leaf area, specific root area, root carbon concentration, and precipitation. The reduced model was fitted using the training observations and subsequently evaluated on the same held-out testing dataset used to evaluate the full model. Predictor selection was completed before the test dataset was used for model evaluation. This reduced model tested whether a more parsimonious trait set could achieve predictive performance comparable to that of the full ten-predictor model while retaining ecological interpretability.

Training, testing, and cross-validation

The merged dataset consisted of 80 observations and was randomly partitioned into training (80%; $n = 64$) and testing (20%; $n = 16$) subsets using the `createDataPartition()` function in the `caret` package (Kuhn, 2008). The training dataset was used for model fitting and predictor selection, whereas the testing dataset was held out from model-development procedures and used only for final evaluation of predictive performance on previously unseen observations.

Model performance was assessed using the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). The held-out test-set performance was considered the primary estimate of model generalization because these observations were not used for model fitting or predictor selection.

Model robustness was additionally assessed using ten-fold cross-validation applied to the full model, implemented through the `caret` package (Kuhn, 2008). Because cross-validation folds were assigned randomly, observations originating from the same species $\times$ precipitation combination could occur in different folds. Consequently, cross-validation results represent model consistency across resampled observations but may provide an optimistic estimate of predictive performance owing to shared biological characteristics among replicate observations. Cross-validation of the full model yielded an R^2 of 0.983, RMSE of 0.046, and MAE of 0.036. Because the full and reduced models were both evaluated on the same held-out testing dataset ($R^2 = 0.41$ and $R^2 = 0.517$, respectively), these held-out test-set values rather than the cross-validation result were considered the primary and more conservative estimates of predictive performance on unseen observations.

External application to an independent mycorrhizal response dataset

To evaluate whether the trait-based modelling approach could be applied to a different dataset and mycorrhizal response measure, Trait2Myco was additionally applied to an independent dataset relating host growth traits, including plant length, total plant biomass, aboveground dry weight, flower number, and root dry weight to AMF root colonization score ($n = 8$; Werner et Kiers (2015); <https://doi.org/10.5061/dryad.08c2k>). A Random Forest model was fitted using the same general modelling approach, with colonization score as the response variable, and evaluated using R^2 , RMSE, and

MAE. Variable importance and partial dependence analyses were conducted using the same procedures as for the primary model. Results from this external application are reported in the Supplementary Material.

Partial dependence analysis

Partial dependence plots (PDPs) were generated using the pdp package (Greenwell, 2017) to estimate the average marginal relationship between individual predictors and model predictions while averaging over the observed values of the remaining predictors. Analyses focused on specific leaf area, root carbon concentration, and precipitation, as these variables consistently ranked among the most important predictors in both the full and reduced models. Principal component analysis (PCA) was performed on the functional trait matrix to visualize multivariate trait space and to assess whether highly responsive and weakly responsive plants occupied distinct regions of functional trait space.

Software availability

Trait2Myco (v0.1.0) is an open-source R package for predicting mycorrhizal growth response (MGR) from plant functional traits and environmental variables. The package is available at <https://github.com/tariq6640/Trait2Myco>.

Results

Model performance

Model performance differed among validation approaches, with cross-validation producing higher performance estimates than the held-out testing dataset (Table 2). The Trait2Myco framework integrated plant functional traits and precipitation data to predict mycorrhizal growth response and evaluate the relative contribution of ecological predictors (Fig. 1). The full Random Forest model, incorporating all ten functional traits and precipitation, accounted for 41% of the variance in mycorrhizal growth response on the held-out test set ($R^2 = 0.41$, $n = 16$; Fig. 2), with predicted values showing agreement with observed MGR across the range of responses, although prediction errors increased at higher observed MGR values.

Table 2
Performance of the Trait2Myco Random Forest models under different validation approaches.

Model	Predictors	Validation Strategy	R ²
Full model	10 traits + precipitation	Independent test set	0.41
Reduced model	SLA + SRA + RCC + precipitation	Independent test set	0.517
Full model	10 traits + precipitation	10-fold cross-validation	0.983

The reduced ecological model, restricted to the four most influential predictors (specific leaf area, specific root area, root carbon concentration, and precipitation), was evaluated on the same held-out test set and achieved a higher R^2 (0.517), indicating that this smaller, ecologically interpretable predictor set provided better predictive performance on hold-out observations than the full ten-predictor model. Because the four predictors were selected using the training dataset before evaluation of the held-out observations, the test-set performance provides an evaluation of the reduced model on observations not used for predictor selection or model fitting.

Ten-fold cross-validation within the training dataset produced a substantially higher R^2 (0.983). Because individual replicates from the same species $\times$ precipitation combinations could be distributed across different folds, this resampling estimate may be higher than performance expected when predicting observations from a new experimental combination. The cross-validation result was therefore interpreted as a measure of model stability under resampling rather than as an independent estimate of generalization. The held-out test-set values ($R^2 = 0.41$ and $R^2 = 0.517$) were therefore considered the primary estimates of predictive performance on unseen observations (Table 2). Overall, the difference between cross-validation and held-out test-set performance indicates that estimated model performance depended on the validation strategy.

Variable importance

Permutation-based variable importance analysis using the training dataset identified specific leaf area as the most important predictor of mycorrhizal growth response in the full model, followed by precipitation and root carbon concentration (Fig. 3). Specific root area, leaf carbon concentration, leaf nitrogen concentration, root nitrogen concentration, specific root length, root tissue density, and root diameter contributed progressively smaller increases in model error upon permutation, with leaf dry matter content contributing the smallest increase in error among the ten traits examined. Both leaf-level and root-level traits appeared among the top predictors, with specific leaf area and root carbon concentration among the four most influential predictors alongside precipitation and specific root area.

Partial dependence on key predictors

Partial dependence plots for the three most important continuous predictors revealed distinct functional relationships with predicted MGR (Fig. 4). Predicted MGR declined sharply with increasing specific leaf area before approaching a plateau at higher specific leaf area values. Predicted MGR showed a unimodal relationship with precipitation, peaking at approximately 800 mm and declining at both lower and higher precipitation levels. Root carbon concentration showed a threshold-like relationship with predicted MGR, with a marked increase in response above approximately 4.35% root carbon concentration, beyond which predicted MGR remained comparatively stable across the observed range of the predictor.

Relative contribution of trait categories

Aggregating variable importance scores by category showed that leaf traits and root traits contributed similar proportions of total variable importance, accounting for 41.9% and 41.3%, respectively, while precipitation accounted for the remaining 16.8% (Fig. 6). Thus, the summed importance of leaf and root predictors was broadly similar, whereas precipitation contributed a smaller proportion of total variable importance.

Multivariate trait space

Principal component analysis of the functional trait matrix showed that PC1 and PC2 together captured the dominant axes of trait variation (Fig. 5). Individuals with higher positive MGR values were concentrated toward the negative end of PC1, in the direction associated with root carbon concentration, whereas individuals with lower or negative MGR values were more broadly distributed toward the positive end of PC1, in the direction associated with specific root length, specific root area, specific leaf area, and leaf nitrogen concentration. The distribution of observations therefore showed partial separation along PC1, with MGR varying across a multivariate gradient in functional trait space.

Supplementary response trait correlation

Pairwise correlation analysis of all functional traits against total MGR (Fig. S1) showed that specific leaf area had the strongest negative correlation with MGR among all traits examined, followed by leaf nitrogen concentration and root diameter, while root carbon concentration and root tissue density showed the strongest positive correlations with MGR (Fig. S2). These univariate correlations were broadly consistent with the direction of the relationships identified by the Random Forest variable-importance and partial dependence analyses, with specific leaf area and root carbon concentration emerging as prominent correlates of MGR.

External application to AMF colonization

To assess whether the trait-based Random Forest workflow could be applied to a different mycorrhizal response measure, a second model was trained on an independent dataset relating plant growth traits to AMF root colonization score (Werner et Kiers, 2015). This model showed high apparent predictive performance ($R^2 = 0.94$, RMSE = 0.070, MAE = 0.054, $n = 8$; Fig. S3), with predicted colonization scores closely tracking observed values across the full range of colonization intensity. Plant length was the strongest predictor of colonization score, followed closely by total plant biomass and aboveground dry weight, with flower number and root dry weight contributing comparatively less (Fig. S3). Partial dependence plots showed that predicted colonization increased sharply with plant length up to approximately 10 cm before approaching a plateau, and showed a similar threshold-like relationships with total plant biomass and root dry weight above intermediate values (Fig. S4). Pairwise correlations among growth traits and colonization score were uniformly positive and strong ($r = 0.62$ – 0.85 ; Fig. S5). Given the small sample size ($n = 8$) and the different response variable, this analysis represents an exploratory external application of the Trait2Myco modelling workflow rather than formal validation of the primary MGR model. The complete analysis is reported in the Supplementary Material.

Discussion

Although AMF colonization is frequently used as a proxy for symbiotic success, colonization intensity and host benefit are often decoupled. Consequently, understanding the resources of variation in mycorrhizal growth response remains a central challenge in mycorrhizal ecology. While AMF associations are widespread across terrestrial ecosystems, the magnitude of plant response varies substantially among species and environments. In this study, we developed Trait2Myco, a Random Forest–based framework that integrates plant functional traits and environmental variables to predict MGR. The framework identified specific leaf area, precipitation, root carbon concentration, and specific root area as the most influential predictors of MGR in the primary dataset, suggesting that AMF responsiveness is associated with multiple dimensions of plant economic strategy and environmental context rather than with a single functional trait.

Resource-acquisitive trait syndromes influence responsiveness to AMF

Specific leaf area emerged as the strongest predictor of MGR in both the full and reduced Trait2Myco models. Specific leaf area is a central component of the plant economics spectrum and is widely associated with rapid resource acquisition, high photosynthetic capacity, and greater nutrient demand (Reich et al. 2014; Wright et al., 2004). Species characterized by high specific leaf area generally acquire resources quickly and maintain fast growth rates, whereas species with lower specific leaf area invest in denser and longer-lived tissues. The negative relationship between specific leaf area and predicted responsiveness observed in this study suggests that more acquisitive species may derive relatively lower benefits from AMF within the range of species, traits, and precipitation conditions represented by our dataset.

This pattern is broadly consistent with ecological theories proposing that plants with efficient direct nutrient acquisition pathways may rely less heavily on fungal nutrient delivery (Chagnon et al., 2013; Koide, 1991). In contrast, species exhibiting more conservative strategies may obtain greater returns from fungal-mediated nutrient acquisition because symbiotic pathways complement slower intrinsic resource-acquisition strategies. These mechanisms remain hypotheses rather than direct inferences from our model, because Trait2Myco identifies predictive associations rather than causal effects. The relationship between specific leaf area and MGR should therefore be interpreted as an association between leaf economic strategy and mycorrhizal responsiveness within the experimental system. Nevertheless, the recovery of specific leaf area as the most influential predictor by Trait2Myco suggests that leaf economic strategies are strongly associated with variation in AMF responsiveness in this dataset.

Root carbon concentration as a predictor of mycorrhizal responsiveness

Root carbon concentration was among the most influential predictors identified by Trait2Myco. While traits such as root diameter, specific root length, and specific root area are commonly discussed in studies of plant AMF interactions, root carbon concentration has rarely been evaluated explicitly as a predictor of responsiveness. Because carbon represents the principal resource supplied by plants to fungal partners, variation in belowground carbon investment may directly influence the costs and potential benefits of maintaining symbiosis (Fellbaum et al., 2014; Kiers et al., 2011; Smith et D.J., 2008). Experimental evidence indicates that carbon supplied by the host can directly regulate fungal nutrient uptake and transport processes, suggesting a mechanistic link between host carbon economy and symbiotic benefit (Fellbaum et al., 2014). The association between root carbon concentration and MGR identified here is therefore consistent with a potential role for belowground carbon economy in variation in AMF responsiveness.

The threshold-like increase in predicted MGR at higher root carbon concentration values suggests that the relationship between root carbon concentration and responsiveness may be nonlinear within the observed range. Plants allocating larger carbon pools belowground may be better able to support fungal metabolism and sustain nutrient-exchange structures, thereby increasing the net benefits derived from AMF associations. This interpretation remains mechanistic hypothesis rather than a causal conclusion from the present analysis. Experimental studies manipulating host carbon allocation while measuring fungal nutrient transfer and plant growth will be needed to test whether root carbon concentration directly influences the outcome of AMF symbiosis. The identification of root carbon concentration as a major predictor nevertheless highlights a potentially useful trait for future studies of variation in mycorrhizal responsiveness.

Root traits contribute substantially to mycorrhizal outcomes

Specific root area was consistently identified among the most important predictors of responsiveness. Root morphological traits influence nutrient foraging efficiency, soil exploration, and opportunities for fungal colonization (Comas et Eissenstat, 2009; Roumet et al., 2016). Species exhibiting higher specific root area and specific root length generally possess greater soil exploration capacity and resource acquisition potential, traits frequently associated with variation in belowground interactions (Comas et Eissenstat, 2009). Plants exhibiting greater root surface area may simultaneously increase direct nutrient uptake and alter their dependence on AMF-derived nutrient pathways. The importance of specific root area suggests that root economic strategies are associated with variation in whether plants benefit from mycorrhizal partnerships.

Aggregated importance scores demonstrated that leaf traits (41.9%) and root traits (41.3%) accounted for similar proportions to total variable importance, whereas precipitation accounted for 16.8%. These percentages describe the relative distribution of permutation-based variable importance and should not be interpreted as independent contributions to model performance. The similar importance assigned to leaf and root traits nevertheless indicates that both aboveground and belowground functional

characteristics were represented among the influential predictors in this dataset. Previous studies have emphasized that root and leaf economics are only partially coordinated and should be considered together when evaluating plant responses to environmental change and biotic interactions (Bergmann et al., 2020; Freschet et al., 2021). Our results are consistent with this whole-plant perspective and suggest that variation in mycorrhizal responsiveness is associated with coordinated variation across leaf and root functional traits.

Environmental context is associated with variation in s responsiveness

Precipitation was the second most influential predictor overall and the most important environmental factor in the model. Water availability affects AMF activity, nutrient mobility, root growth, and plant carbon allocation, all of which influence the costs and benefits of symbiosis (Smith et D.J., 2008; van der Heijden et al., 2015). Partial dependence analyses demonstrated a nonlinear relationship between precipitation and predicted MGR, with responsiveness peaking at intermediate precipitation levels and declining at both wetter and drier conditions within the range represented in the dataset.

This pattern is consistent with the view that AMF outcomes are strongly context dependent (Hoeksema et al., 2010; Johnson et al., 2008). Under high water availability, plants may acquire resources effectively without fungal assistance, reducing the relative value of symbiosis. Conversely, under severe drought, host carbon limitation and reduced fungal activity may constrain the benefits of AMF associations. An intermediate precipitation optimum could therefore reflect a balance between resource limitation and the capacity of plants and fungi to sustain the symbiosis. These mechanisms were not directly tested in the present study, and the observed nonlinear relationship should not be interpreted as evidence that AMF benefits generally peak at intermediate precipitation. The precipitation response identified by Trait2Myco should instead be interpreted as an empirical association within the environmental range represented by the experiment.

Functional trait syndromes and multivariate trait space

The PCA revealed that observations with different MGR values occupied partially distinct regions of multivariate trait space. Observations associated with greater MGR were positioned along gradients defined by specific leaf area, specific root area, specific root length, root carbon concentration, root diameter, and leaf nitrogen content, indicating that variation in responsiveness was associated with combinations of leaf and root traits rather than with a single trait alone. Previous reports have emphasized that root trait variation is multidimensional and cannot be adequately represented by a single axis of resource acquisition, highlighting the importance of trait syndromes rather than isolated traits (Bergmann et al., 2020; Weemstra et al., 2016). The PCA pattern therefore supports the interpretation that variation in mycorrhizal responsiveness is associated with multivariate functional trait variation.

The PCA findings closely complement the Random Forest analyses. Whereas Random Forest identified the most influential predictors, PCA placed these traits within the broader multivariate structure of functional trait variation. Together, these analyses suggest that MGR is associated with coordinated variation across leaf and root traits, although they do not establish that individual trait combinations causally determine mycorrhizal responsiveness.

Trait2Myco as a predictive ecological framework

An important contribution of this study is the development of Trait2Myco as a trait-based predictive framework rather than a new machine-learning algorithm. Trait2Myco successfully identified functional trait syndromes associated with AMF responsiveness while also detecting nonlinear relationships that may be difficult to recover using traditional statistical approaches. The reduced ecological model retained four predictors while achieving higher held-out test-set R^2 than the full model, indicating that a relatively small set of ecologically interpretable variables can provide useful predictive information within the experimental system examined here.

The novelty of Trait2Myco lies in its integration of functional trait theory with Random Forest to explore variation in mycorrhizal responsiveness, rather than in the underlying machine-learning methodology. Because the framework uses readily measurable plant traits and environmental variables, it provides a basis for future testing across a broader range of plant species, environments, and experimental conditions. Such validation could determine whether trait-based predictions of AMF responsiveness have practical applications in plant selection, inoculant development, restoration ecology, or agricultural management.

Limitations

First, trait measurements and mycorrhizal response measurements originated from separate but linked experimental datasets. Consequently, trait values were effectively constant across precipitation treatments within species, and the effective number of unique trait combinations remained relatively small. The model should therefore be interpreted primarily as identifying associations between species-level functional characteristics and variation in MGR, rather than as identifying individual-level causal predictors.

Second, only four plant species were included in the dataset. Although the framework identified consistent and ecologically interpretable associations, broader taxonomic representation will be necessary to evaluate the generality of these patterns.

Third, the dataset contained five replicate observations for each species $\times$ precipitation combination. Random observation-level cross-validation allowed replicate observations from the same experimental combination to occur in different folds, potentially increasing similarity between model-fitting and validation observations. The high cross-validation R^2 (0.983) should therefore not be interpreted as an unbiased estimate of predictive performance. The held-out test-set results ($R^2 = 0.41$ for the full model

and $R^2 = 0.517$ for the reduced model) provide more conservative estimates of performance on observations not used during model fitting and predictor selection.

Fourth, predictor selection for the reduced model must be performed using the training dataset only to avoid information leakage from the held-out test set. The resulting four-predictor model should therefore be interpreted as a parsimonious model selected within the present training data rather than as a definitive minimal predictor set.

Fifth, the external analysis using AMF colonization provides an exploratory application of the modelling workflow rather than formal validation of the MGR model. The external dataset contained only eight observations and used a different response variable, limiting the extent to which its high apparent predictive performance can be generalized.

Finally, Trait2Myco focused on plant traits and precipitation and did not incorporate fungal identity, soil properties, microbial community composition, or nutrient availability. Inclusion of additional biotic and abiotic variables could improve predictive performance and help determine whether trait–MGR relationships identified here remain robust across contrasting environmental and fungal contexts.

Conclusions

Trait2Myco showed that variation in plant responsiveness to AMF can be associated with a relatively small set of functional traits and an environmental variable. Specific leaf area, root carbon concentration, specific root area, and precipitation emerged as the most influential predictors of MGR in the primary dataset, indicating that both leaf and root economic strategies contribute to determining AMF outcomes. By integrating machine learning with ecological theory, Trait2Myco provides a reproducible framework for exploring trait-based variation in mycorrhizal responsiveness and identifying functional characteristics associated with stronger or weaker responses. The higher performance of the reduced four-predictor model on the held-out test set further suggests that parsimonious, ecologically interpretable predictor sets can provide useful predictive information within the experimental system examined here. Future studies incorporating a broader diversity of plant species, environmental conditions, fungal identities, and soil properties will be needed to determine whether the trait–MGR relationships identified here generalize across ecological contexts and can support more robust prediction of plant responses to AMF.

Declarations

Data availability statement

The Trait2Myco source code, trained Random Forest model, package documentation, and example workflows are publicly available through the GitHub repository Trait2Myco at <https://github.com/tariq6640/Trait2Myco>. The primary datasets used in this study were obtained from

publicly available Dryad repositories and are cited in the manuscript. Trait2Myco can be installed directly in R from GitHub using the `remotes::install_github()` function.

Acknowledgements

We thank the original authors whose work and publicly available data were used for the development, training, and testing of Trait2Myco.

Author contributions

T.S. conceived the study, developed the Trait2Myco pipeline, collected and analyzed the data, and drafted the manuscript. M.A.E. contributed to data collection and analysis. M.H. secured funding, supervised the study, and contributed to manuscript review and editing.

Funding

This research was funded by OCP Nutricrops (Project AS FN 08).

Data availability

Data supporting the findings of this study are provided in the Supplementary Information. Trait2Myco (v0.1.0) is an open-source R package for predicting mycorrhizal growth response (MGR) from plant functional traits and environmental variables. The package is freely available at <https://github.com/tariq6640/Trait2Myco>.

Competing interests

The authors declare no competing interests.

References

1. Begum N, Qin C, Ahanger MA, Raza S, Khan MI, Ashraf M, Ahmed N et, Zhang L (2019) Role of Arbuscular Mycorrhizal Fungi in Plant Growth Regulation: Implications in Abiotic Stress Tolerance. *Front Plant Sci* 10:1068. <https://doi.org/10.3389/fpls.2019.01068>
2. Bergmann J, Weigelt A, van der Plas F, Laughlin DC, Kuyper TW, Guerrero-Ramirez N, Valverde-Barrantes OJ, Bruelheide H, Freschet GT, Iversen CM, Kattge J, McCormack ML, Meier IC, Rillig MC, Roumet C, Semchenko M, Sweeney CJ, van Ruijven J, York LM et, Mommer L (2020) The fungal collaboration gradient dominates the root economics space in plants. *Sci Adv* 6(27). <https://doi.org/10.1126/sciadv.aba3756>
3. Bever JD, Richardson SC, Lawrence BM, Holmes J, et Watson M (2009) Preferential allocation to beneficial symbiont with spatial structure maintains mycorrhizal mutualism. *Ecol Lett* 12(1):13–21. <https://doi.org/10.1111/j.1461-0248.2008.01254.x>

4. Bonfante P et, Genre A (2010) Mechanisms underlying beneficial plant-fungus interactions in mycorrhizal symbiosis. *Nat Commun* 1:48. <https://doi.org/10.1038/ncomms1046>
5. Breiman L (2001) Random Forests *Machine Learning* 45(1):5–32. <https://doi.org/10.1023/a:1010933404324>
6. Brundrett MC (2009) Mycorrhizal associations and other means of nutrition of vascular plants: understanding the global diversity of host plants by resolving conflicting information and developing reliable means of diagnosis. *Plant Soil* 320(1–2):37–77. <https://doi.org/10.1007/s11104-008-9877-9>
7. Chagnon PL, Bradley RL, Maherali H, et Klironomos JN (2013) A trait-based framework to understand life history of mycorrhizal fungi. *Trends Plant Sci* 18(9):484–491. <https://doi.org/10.1016/j.tplants.2013.05.001>
8. Comas LH, et Eissenstat DM (2009) Patterns in root trait variation among 25 co-existing North American forest species. *New Phytol* 182(4):919–928. <https://doi.org/10.1111/j.1469-8137.2009.02799.x>
9. Corrêa A, Gurevitch J, Martins-Loução MA et, Cruz C (2011) C allocation to the fungus is not a cost to the plant in ectomycorrhizae. *Oikos* 121(3):449–463. <https://doi.org/10.1111/j.1600-0706.2011.19406.x>
10. Cutler DR, Edwards TC Jr., Beard KH, Cutler A, Hess KT, Gibson J et, Lawler JJ (2007) Random forests for classification in ecology *Ecology*, 88(11), 2783–2792. <https://doi.org/10.1890/07-0539.1>
11. Delavaux CS, Smith-Ramesh LM, et Kuebbing SE (2017) Beyond nutrients: a meta-analysis of the diverse effects of arbuscular mycorrhizal fungi on plants and soils *Ecology*, 98(8), 2111–2119. <https://doi.org/10.1002/ecy.1892>
12. Fellbaum CR, Mensah JA, Cloos AJ, Strahan GE, Pfeffer PE, Kiers ET et, Bucking H (2014) Fungal nutrient allocation in common mycorrhizal networks is regulated by the carbon source strength of individual host plants. *New Phytol* 203(2):646–656. <https://doi.org/10.1111/nph.12827>
13. Freschet GT, Roumet C, Comas LH, Weemstra M, Bengough AG, Rewald B, Bardgett RD, De Deyn GB, Johnson D, Klimesova J, Lukac M, McCormack ML, Meier IC, Pages L, Poorter H, Prieto I, Wurzburger N, Zadworny M, Bagniewska-Zadworna A, Stokes A (2021) Root traits as drivers of plant and ecosystem functioning: current understanding, pitfalls and future research needs. *New Phytol* 232(3):1123–1158. <https://doi.org/10.1111/nph.17072>
14. Greenwell BM (2017) pdp: an R package for constructing partial dependence plots. *R J* 9:421–436
15. Hart MM, et Reader RJ (2002) Taxonomic basis for variation in the colonization strategy of arbuscular mycorrhizal fungi. *New Phytol* 153(2):335–344. <https://doi.org/10.1046/j.0028-646X.2001.00312.x>
16. Hoeksema JD, Chaudhary VB, Gehring CA, Johnson NC, Karst J, Koide RT, Pringle A, Zabinski C, Bever JD, Moore JC, Wilson GW, Klironomos JN et, Umbanhowar J (2010) A meta-analysis of context-dependency in plant response to inoculation with mycorrhizal fungi. *Ecol Lett* 13(3):394–407. <https://doi.org/10.1111/j.1461-0248.2009.01430.x>

17. Janos DP (1980) Mycorrhizae Influence Trop Succession Biotropica 12(2).
<https://doi.org/10.2307/2388157>
18. Johnson NC, Graham JH, et Smith FA (2008) Functioning of mycorrhizal associations along the mutualism–parasitism continuum. New Phytol 135(4):575–585. <https://doi.org/10.1046/j.1469-8137.1997.00729.x>
19. Kiers ET, Duhamel M, Beesetty Y, Mensah JA, Franken O, Verbruggen E, Fellbaum CR, Kowalchuk GA, Hart MM, Bago A, Palmer TM, West SA, Vandenkoornhuyse P, Jansa J et, Bucking H (2011) Reciprocal rewards stabilize cooperation in the mycorrhizal symbiosis. Science 333(6044):880–882. <https://doi.org/10.1126/science.1208473>
20. Klironomos JN (2003) Variation in Plant Response to Native and Exotic Arbuscular Mycorrhizal Fungi. Ecology 84(9):2292–2301
21. Koide RT (1991) Nutrient supply, nutrient demand and plant response to mycorrhizal infection. New Phytol 117(3):365–386. <https://doi.org/10.1111/j.1469-8137.1991.tb00001.x>
22. Kuhn M (2008) Building Predictive Models in R Using the caret Package. J Stat Softw 28(5). <https://doi.org/10.18637/jss.v028.i05>
23. Lehmann A, Barto EK, Powell JR, et Rillig MC (2012) Mycorrhizal responsiveness trends in annual crop plants and their wild relatives—a meta-analysis on studies from 1981 to 2010. Plant Soil 355(1–2):231–250. <https://doi.org/10.1007/s11104-011-1095-1>
24. Liaw A, et Wiener M (2002) Classification and Regression by Randomforest. R News 2:18–22
25. Maherali H, et Klironomos JN (2007) Influence of phylogeny on fungal community assembly and ecosystem functioning. Science 316(5832):1746–1748. <https://doi.org/10.1126/science.1143082>
26. Prasad AM, Iverson LR, et Liaw A (2006) Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction *Ecosystems*, 9(2), 181–199. <https://doi.org/10.1007/s10021-005-0054-1>
27. Reich PB, et Cornelissen H (2014) The world-wide ‘fast–slow’ plant economics spectrum: a traits manifesto. J Ecol 102(2):275–301. <https://doi.org/10.1111/1365-2745.12211>
28. Roumet C, Birouste M, Picon-Cochard C, Ghestem M, Osman N, Vrignon-Brenas S, Cao KF et, Stokes A (2016) Root structure-function relationships in 74 species: evidence of a root economics spectrum related to carbon economy. New Phytol 210(3):815–826. <https://doi.org/10.1111/nph.13828>
29. Smith SE, et DJ (2008) R. *Mycorrhizal symbiosis*, 3rd edn., London (3rd^e éd.). Academic Press
30. Smith SE, et Smith FA (2011) Roles of arbuscular mycorrhizas in plant nutrition and growth: new paradigms from cellular to ecosystem scales. Annu Rev Plant Biol 62:227–250. <https://doi.org/10.1146/annurev-arplant-042110-103846>
31. Treseder KK (2013) The extent of mycorrhizal colonization of roots and its influence on plant growth and phosphorus content. Plant Soil 371(1–2):1–13. <https://doi.org/10.1007/s11104-013-1681-5>

32. van der Heijden MGA, Martin FM, Selosse MA, et Sanders IR (2015) Mycorrhizal ecology and evolution: the past, the present, and the future *New Phytol*, 205(4), 1406–1423.
<https://doi.org/10.1111/nph.13288>
33. Veresoglou SD, Chen B, et Rillig MC (2012) Arbuscular mycorrhiza and soil nitrogen cycling. *Soil Biol Biochem* 46:53–62. <https://doi.org/10.1016/j.soilbio.2011.11.018>
34. Weemstra M, Mommer L, Visser EJ, van Ruijven J, Kuyper TW, Mohren GM et, Sterck FJ (2016) Towards a multidimensional root trait framework: a tree root review. *New Phytol* 211(4):1159–1169.
<https://doi.org/10.1111/nph.14003>
35. Werner GDA, et Kiers ET (2015) Partner selection in the mycorrhizal mutualism. *New Phytol* 205(4):1437–1442. <https://doi.org/10.1111/nph.13113>
36. Wilson GWT et, Hartnett DC (1998) Interspecific variation in plant responses to mycorrhizal colonization in tallgrass prairie. *Am J Bot* 85(12):1732–1738. <https://doi.org/10.2307/2446507>
37. Wright IJ, Reich PB, Westoby M, Ackerly DD, Baruch Z, Bongers F, Cavender-Bares J, Chapin T, Cornelissen JH, Diemer M, Flexas J, Garnier E, Groom PK, Gulias J, Hikosaka K, Lamont BB, Lee T, Lee W, Lusk C, Villar R (2004) The worldwide leaf economics spectrum. *Nature* 428(6985):821–827.
<https://doi.org/10.1038/nature02403>
38. Zhao H, Liao H, Biere A et, Peng S (2026) Plant economics traits predict plant carbon allocation and responsiveness to arbuscular mycorrhizal fungi under varying precipitation. *Funct Ecol* 40(6):1921–1932. <https://doi.org/10.1111/1365-2435.70340>

Figures

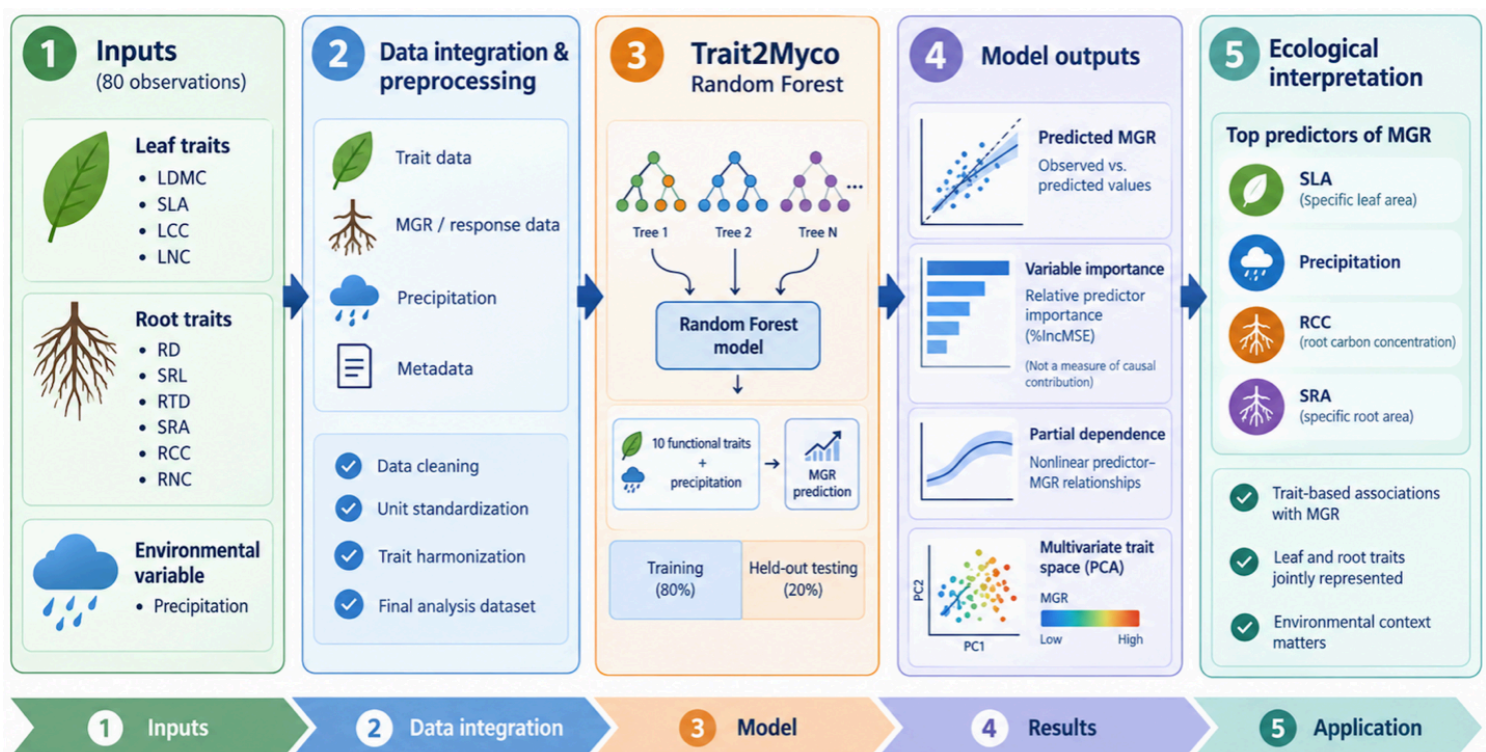


Figure 1

Conceptual overview of the Trait2Myco framework for predicting mycorrhizal growth response (MGR) from plant functional traits and precipitation. The workflow includes data integration, Random Forest model development, predictor-importance assessment, partial dependence analysis, and multivariate trait-space analysis.

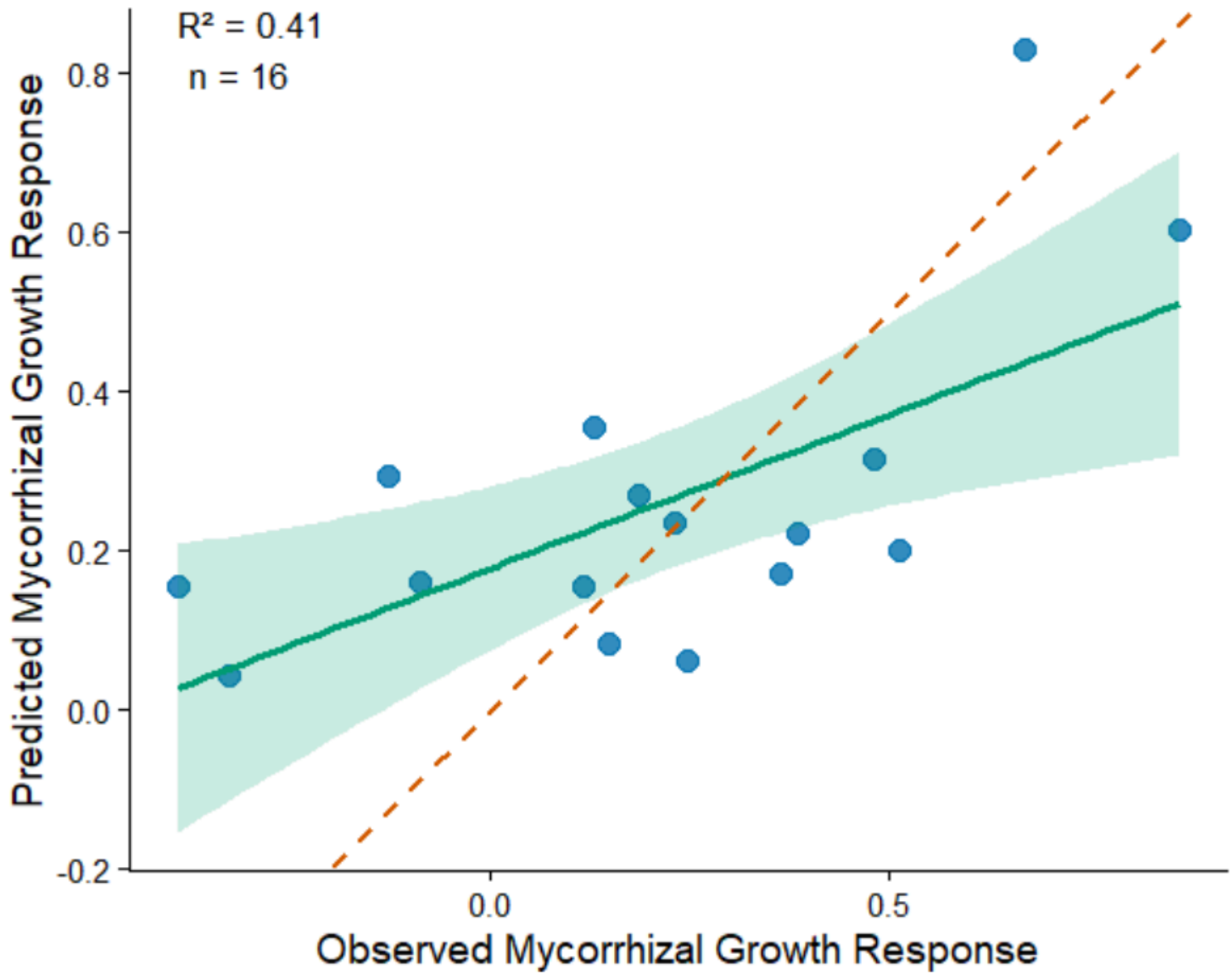


Figure 2

Performance of the Trait2Myco Random Forest model for predicting mycorrhizal growth response (MGR). Predicted MGR values are plotted against observed MGR values for the held-out testing dataset ($n = 16$). The solid green line represents the fitted relationship between observed and predicted values, with the shaded region indicating the 95% confidence interval. The dashed line represents the 1:1 relationship.

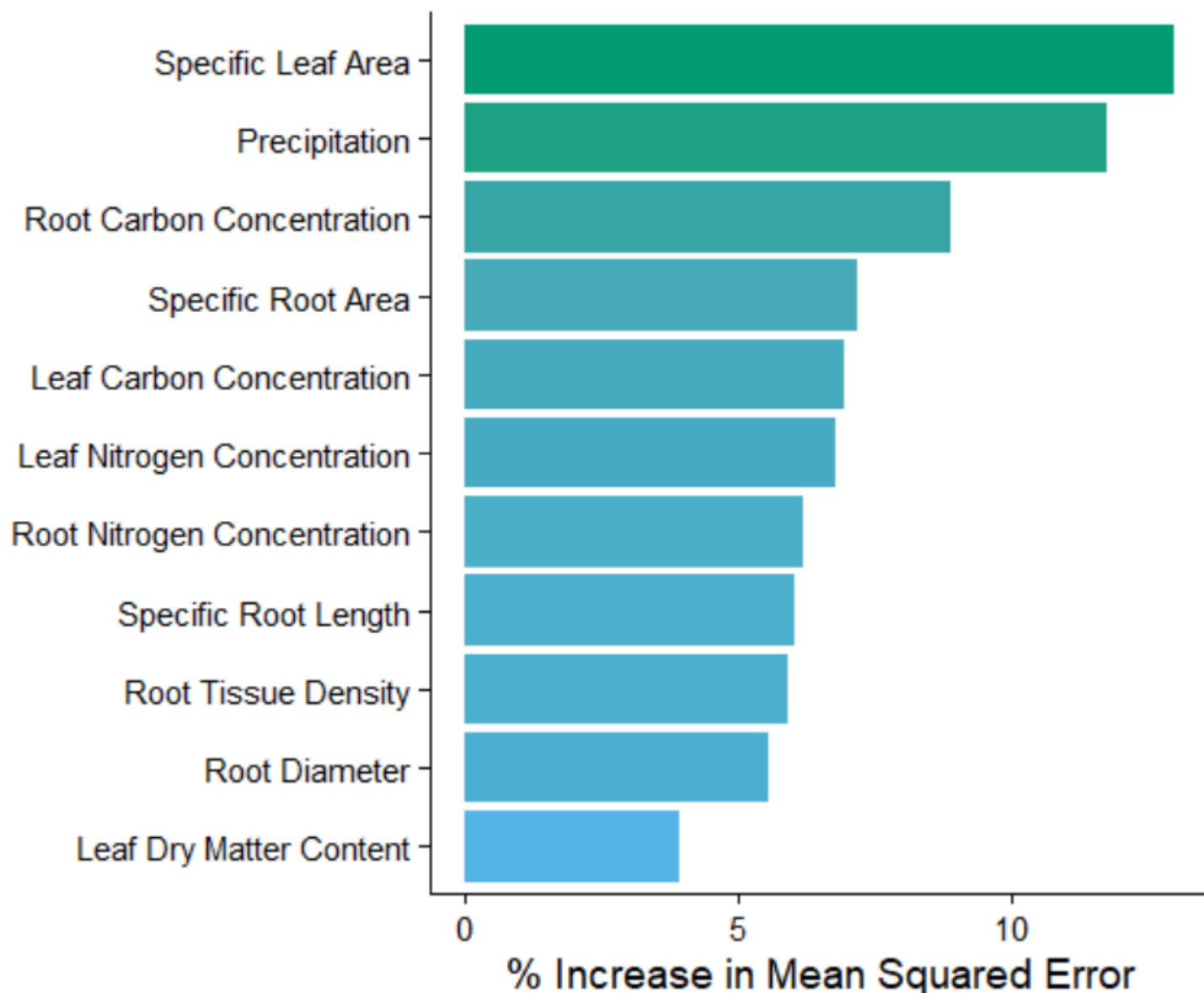


Figure 3

Variable importance ranking from the full Trait2Myco Random Forest model. Predictor importance was quantified as the percentage increase in mean squared error (%IncMSE) following permutation of individual variables. Specific Leaf Area was identified as the strongest predictor of mycorrhizal growth response, followed by precipitation, root carbon concentration, and specific root area. The remaining predictors contributed progressively smaller increases in prediction error upon permutation. Higher %IncMSE values indicate greater relative importance of model prediction.

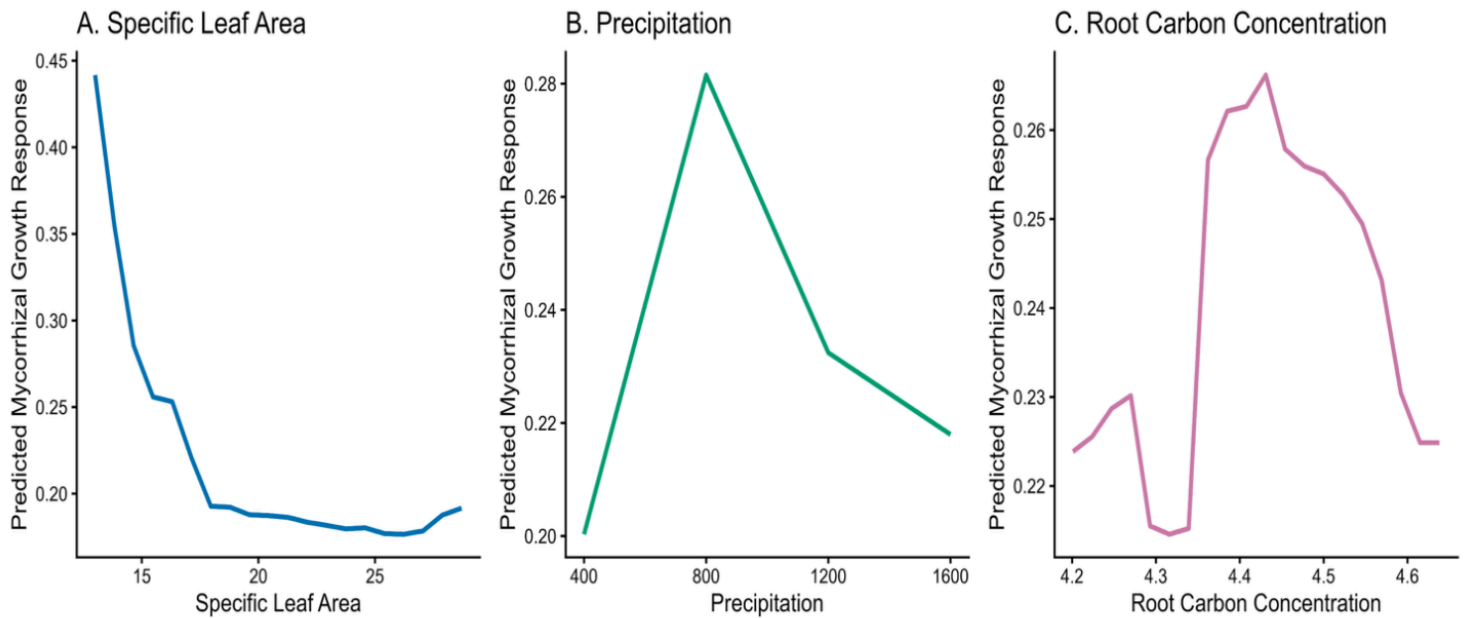


Figure 4

Partial dependence plots illustrating the marginal effects of the three most influential predictors on predicted mycorrhizal growth response (MGR). (A) Specific leaf area showed a negative relationship with MGR, with responsiveness declining across the observed specific leaf area gradient before reaching a plateau. (B) Precipitation showed a unimodal relationship with MGR, with the highest predicted values occurring at intermediate precipitation levels. (C) Root carbon concentration showed a threshold-like response, with predicted MGR increasing sharply above approximately 4.35% root carbon concentration and remaining relatively stable thereafter within the observed range. Partial dependence plots depict the effect of each predictor and model predictions while averaging over the distributions of the other predictors.

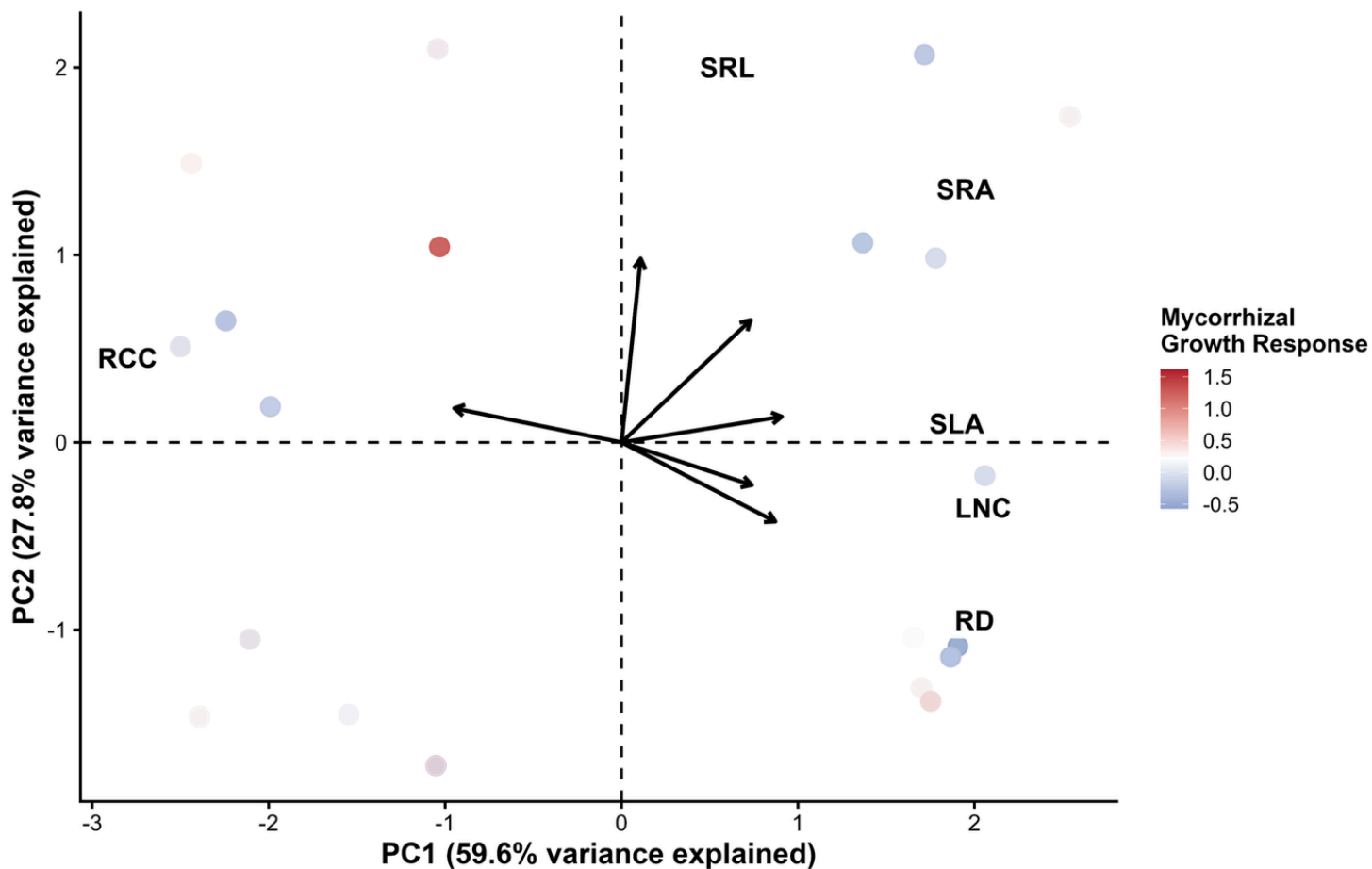


Figure 5

Principal component analysis (PCA) of plant functional traits associated with variation in mycorrhizal growth response (MGR). Points represent individual observations and are colored according to MGR, with warmer colors indicating higher response values. Trait vectors represent the direction and magnitude of trait contributions to the principal component axes.

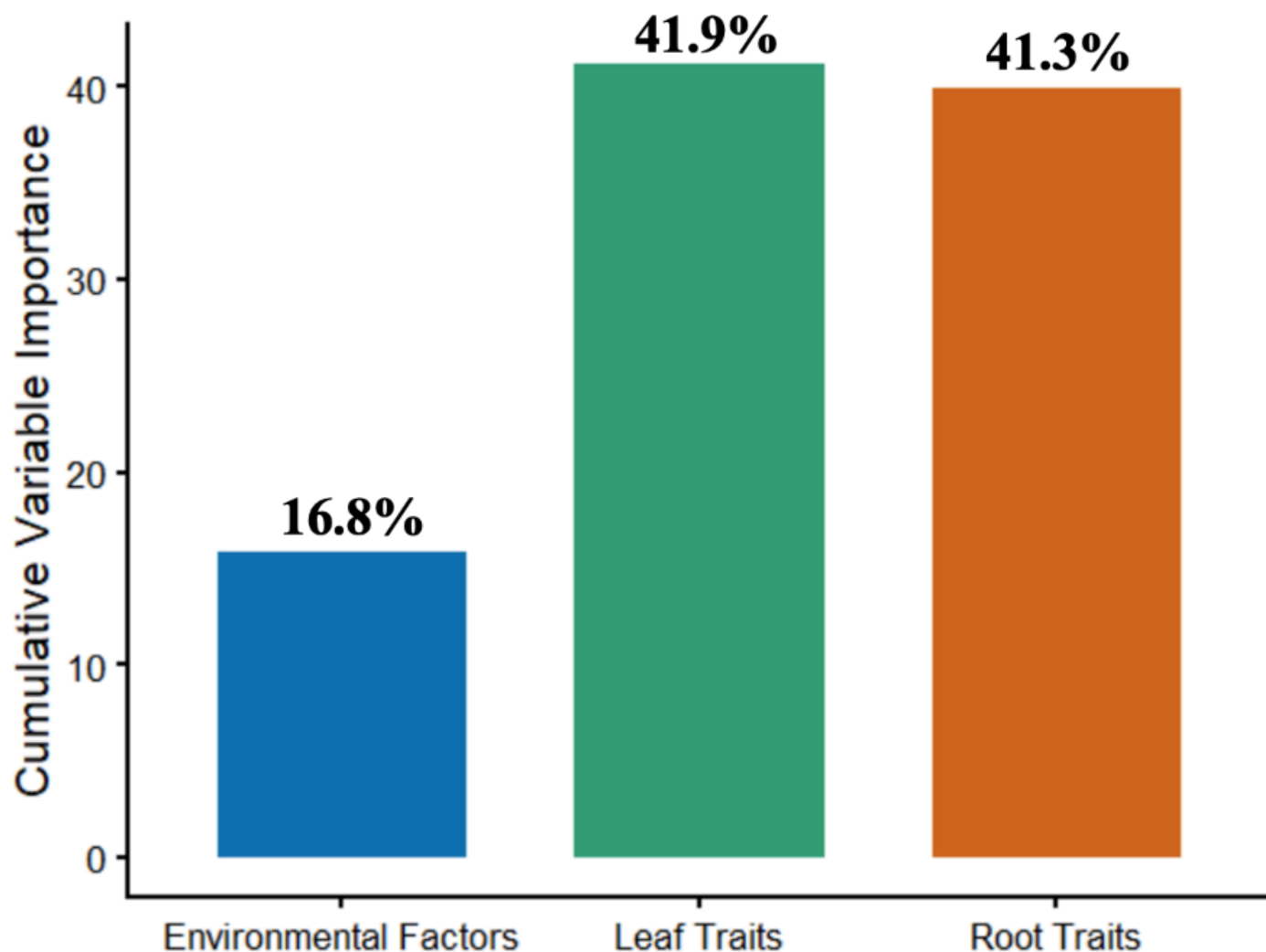


Figure 6

Relative contribution of predictor categories to total permutation-based variable importance in the Trait2Myco framework. Variable importance values were aggregated into three ecological categories: environmental factors (precipitation), leaf traits, and root traits. Leaf traits contributed 41.9% of total variable importance, root traits accounted for 41.3%, and environmental factors accounted for 16.8%. These percentages represent the relative distribution of variable-importance scores and not independent percentages of model performance explained.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryInformation.docx](#)