A Perspective on Plant Phenomics: Coupling Deep Learning and Near-Infrared Spectroscopy

François Vasseur1*, Denis Cernot2,3, Grégory Beurier2,3, Julie Messier4, Lauriane Rouan2,3, Justine Bresson1, Martin Ecarnot3, Mark Stahl5, Simon Heumos6,7, Marianne Gérard1, Hans Reijnen1, Pascal Tillard6, Benoît Lacombe4, Amélie Emanuel1,8, Justine Floret1,10, Aurélien Estarague1, Stefania Przybylska1, Kevin Sartori1, Lauren M. Gillespie1, Etienne Baron1, Elena Kazakou9, Denis Vile10 and Cyrille Violle1

1 CEFE, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France, 2 CIRAD, UMR AGAP Institut, Montpellier, France, 3 UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France, 4 Department of Biology, University of Waterloo, Waterloo, ON, Canada, 5 Center for Plant Molecular Biology (ZMBP), University of Tübingen, Tübingen, Germany, 6 Quantitative Biology Center (QBIC), University of Tübingen, Quantitative Biology Center (QBIC), University of Tübingen, Germany, 7 Biomedical Data Science, Department of Computer Science, University of Tübingen, Tübingen, Germany, 8 BPMP, Univ Montpellier, CNRS, INRAE, Montpellier, France, 9 CEFE, Univ Montpellier, CNRS, EPHE, Institut Agro, IRD, Montpellier, France, 10 LEPSE, Univ Montpellier, INRAE, Institut Agro, Montpellier, France

The trait-based approach in plant ecology aims at understanding and classifying the diversity of ecological strategies by comparing plant morphology and physiology across organisms. The major drawback of the approach is that the time and financial cost of measuring the traits on many individuals and environments can be prohibitive. We show that combining near-infrared spectroscopy (NIRS) with deep learning resolves this limitation by quickly, non-destructively, and accurately measuring a suite of traits, including plant morphology, chemistry, and metabolism. Such an approach also allows to position plants within the well-known CSR triangle that depicts the diversity of plant ecological strategies. The processing of NIRS through deep learning identifies the effect of growth conditions on trait values, an issue that plagues traditional statistical approaches. Together, the coupling of NIRS and deep learning is a promising high-throughput approach to capture a range of ecological information on plant diversity and functioning and can accelerate the creation of extensive trait databases.

Keywords: Arabidopsis thaliana, near-infrared spectroscopy (NIRS), multivariate analysis, machine learning, functional traits, metabolomics, trait-based ecology

INTRODUCTION

In trait-based ecology, the comparison of plant phenotype across multiple species aims at identifying general trends of variation to describe the biodiversity of plant forms and functions (Grime, 1988; Keddy, 1992; Diaz et al., 2016; Garnier et al., 2016). Ecological strategies are characterized qualitatively and quantitatively from the measurement of key functional traits, i.e., morphological, physiological, and phenological parameters that determine plant growth and reproduction (Violle et al., 2007). However, our understanding of plant diversity with comparative approaches is impeded by three main limitations. First, measuring the traits that describe ecological strategies on many individuals remains laborious. Second, intraspecific trait variability and plasticity to
the environment still remain largely unconnected to traditional cross-species studies (but see Albert et al., 2010; Anderegg et al., 2018). Third, we need to clarify if and how plant (“soft”) traits used classically to describe ecological strategies are connected to plant metabolism and physiology (“hard” traits).

The development of near-infrared spectroscopy (NIRS) has provided a unique, fast, and reliable tool enabling the collection of a myriad of traits non-destructively (Foley et al., 1998; Cozzolino et al., 2001; Pasquini, 2018; Silva-Perez et al., 2018). NIRS measures the light reflected from a sample after irradiating it with wavelengths ranging from visible (VIS, 400–700 nm), near-infrared (NIR, 700–1,100 nm), to shortwave infrared (SWIR, 1,100–2,500 nm). This provides a signature of the physical and chemical characteristics of the sample (Box 1). NIRS has been widely used for determining chemical traits in various fields. For instance, it is extensively used to characterize chemical products in pharmaceutical, agricultural, and food sectors (Shepherd and Walsh, 2007; Wójcicki, 2015; Biancolillo and Marini, 2018; Pasquini, 2018). In plant science, NIRS takes an increasingly important place as a high-throughput, cost-efficient method for the characterization of biodiversity (Arslan et al., 2018; Silva-Perez et al., 2018; Burnett et al., 2021; Kothari et al., 2021). For instance, it is widely used to predict differences in leaf palatability, digestibility, and decomposability—through lignin and fiber content—between species (Birth and Hecht, 1987; Andrés et al., 2005). The advantages of this method are numerous: spectral measurements are extremely rapid, taking only a few seconds, a single spectral measurement simultaneously captures multiple diverse plant traits (Petit Bon et al., 2020), minimal or no sample preparation is required, and the measurements are non-destructive which allows to track trait changes over time and avoids interfering with the organism.

While NIRS data are simple to acquire and rapidly generate a very large amount of information, they also require extensive post-processing, via chemometric and multivariate statistical

### BOX 1 | Principle of near-infrared spectroscopy (NIRS) for plant characterization

The leaf spectral reflectance is based on the low reflectivity in the visible part of the spectrum (400–700 nm), due to a strong absorption by photosynthetic pigments, and the high reflectivity in the near infrared (700–1,100 nm) produced by a high scattering of light by the leaf mesophyll tissues (Koppling, 1979). For instance, in the SWIR part of the spectrum (1100–2,500 nm), the reflectance intensity is affected by the water, cellulose, protein, and lignin content of plant tissues (Fascher et al., 2010). Healthy leaves emit radiation in the thermal infrared band (<10 μm) according to their temperature, because of their high water content (emissivity between 0.97 and 0.99). The leaves appear green because the green light band (550 nm) is reflected relatively efficiently when compared with the blue, yellow, and red bands, which are absorbed by photoactive pigments. This absorption at different wavelengths produces a spectrum of light reflectance (Figure I), which can be treated as a “signal” of the leaf physical and chemical properties. The physical association between leaf properties and light reflectance is particularly useful to investigate leaf composition, functioning, and diversity. Different leaves will have different spectral signatures depending on their structure and chemical composition. For example, leaf nitrogen concentrations are associated with wavelengths absorbed by chlorophyll a and b in the visible part of the spectrum (400–700 nm), the spectral red edge (700–760 nm), and proteins in the SWIR (1,300–2,500 nm); Gatselis and Merzlyak, 1997; Kokaly, 2001). In the SWIR (SWIR, 700–1,300 nm), structures such as palisade cell density are important determinants of the spectral reflectance because of the very low effective photon penetration distance at these wavelengths.
analyses. Usually, spectral information can be exploited through the development of calibration models relating spectra and reference trait data. Calibration models are built with a representative subsample of a complete data set, in terms of the range of spectral variation treated (Foley et al., 1998). After building and validating models linking plant spectra to independently measured traits in the calibration dataset, the trait values of new samples are predicted from their spectra using these models. For that, different statistical methods are commonly used to predict trait data from spectra, including partial least squares regression (PLSR; Wold et al., 1983), principal components analysis (Dreccer et al., 2014), and 2D correlation plots (Darvishzadeh et al., 2008). However, the performance of these methods, and especially PLSR, in estimating plant traits has been shown to vary significantly across species and growth conditions (Fu et al., 2020). In recent years, machine learning approaches have become widespread in multiple fields due to their better predictive performance. Machine learning and more particularly deep learning techniques—are promising methods to improve the statistical analysis of high-throughput data (Mishra and Passos, 2021).

Spectral predictions of functional traits have been used to screen interspecific diversity across individual leaves, canopies, and biomes (Doughty et al., 2011; Roelofsen et al., 2014; Serbin et al., 2016; Wu et al., 2016; Chavana-Bryant et al., 2017). Yet, investigating intraspecific variability is crucial to connect global trait diversity to the underlying mechanisms of selection, genetic differentiation, and evolutionary adaptation (Violle et al., 2014).

In this context, the model species Arabidopsis thaliana is an interesting model to test the predictive power of plant diversity with NIRS. Indeed, this species exhibits a large range of functional trait variation across its geographic range (Lasky et al., 2012; May et al., 2017; Price et al., 2018; Takou et al., 2018; Sartori et al., 2019), and hundreds of natural ecotypes have been fully sequenced to examine the genetic determinism of this variation (1001 Genomes Consortium, 2016). Ecological studies have taken advantage of this feature to examine the evolution of plant strategies.
NIRS QUANTIFIES FUNCTIONAL TRAIT VARIABILITY AND SUMMARIZES PLANT ECOLOGICAL STRATEGIES

A key goal of trait-based ecology is to determine the physiological mechanisms of plant adaptation to the environment through the measurement of multiple traits related to resource-use, growth, development, and physiology. Recent efforts based on analyzing interspecific trait diversity have revealed functional tradeoffs at both local and global scales (Messier et al., 2016), which suggests that plant diversity is shaped by universal constraints. For instance, Diaz et al. (2016) recently analyzed more than 45,000 plant species and demonstrated that their diversity falls along two main phenotypic dimensions: one related to plant size, which affects competitive ability and dispersal; the other related to leaf anatomy, chemical composition, and longevity. This second phenotypic dimension, called the leaf economics spectrum (Wright et al., 2004), trades off traits positively related to nutrient retention—such as leaf dry matter content (LDMC), leaf nitrogen content (LNC), and leaf life span—with traits positively related to carbon acquisition—such as specific leaf area (SLA) and leaf photosynthetic rate. Importantly, the same trade-off has been observed within species (Vasseur et al., 2012; Anderegg et al., 2018; Sartori et al., 2019).

Different theories have been proposed to categorize plant phenotypic diversity into ecological strategies related to plant adaptation to the environment. Among these theories, Grime (1974) proposed that the quantitative variation in plant strategies is expected to result from their adaptation to contrasting levels of resource availability and disturbance. Following this hypothesis, plant strategies can thus be classified through a combination of three main axes, competitors (C), stress-tolerators (S), and ruderalists (R; Grime, 1977, 1988). The “CSR” model suggests that the evolution of plant strategies is driven by trade-offs between resource capture and conservation, space occupancy, longevity, and dispersal. For instance, C-type plants invest resources into the growth of large organs to outcompete neighbors, S-type plants invest resources to conserve nutrients and protect tissues from stress damages, while R-type plants invest resources into rapid reproduction and propagule dispersal in highly disturbed environments. The CSR strategies are often depicted in a triangle with the primary types occupying the corners and intermediate forms, composed of a combination of types (e.g., “SR” and “CS”), arrayed within the triangle. The quantitative variations between CSR strategies are expected to result from plant adaptation to contrasting levels of abiotic stresses and disturbance. CSR variation has also been reported within species—notably in A. thaliana—and explained by evolutionary adaptation to the environment (May et al., 2017; Vasseur et al., 2018b). However, measuring through destructive methods, the numerous traits that enable the quantification of ecological strategies within—and a fortiori across—species remains an obstacle for the large-scale analysis of plant populations, which therefore limits our ability to temporally follow the relationships between plant traits, strategies, and the environment.

Using convolutional neural network (CNN; Box 2, Supplementary Material, Supplementary Table S2) to analyze our database of spectra and traits in A. thaliana, we show that most leaf traits were accurately predicted (Table 1). For instance, only leaf relative water content (RWC) and the leaf isotopic ratio of nitrogen (δ15N) had validation $R^2$ below 0.65 (Table 1). Yet, previous studies showed that δ15N can be predicted with NIRS (Kleinebecker et al., 2009). Here, correlations between measured and predicted values were the highest for leaf traits associated with the leaf economics spectrum (SLA, LDMC, and LNC, all $r^2 > 0.85$; Table 1). Importantly, for these traits, the correlations calculated from the predicted data were the same as those calculated from the direct measurements ($p > 0.05$; Figures 1A–C). Previous studies showed that SLA can be accurately measured with NIRS from the level of individual leaves to the level of the tree canopies (Curran, 1989; Lymburner et al., 2000; Asner and Martin, 2008; Asner et al., 2009; Jacquemoud et al., 2009; Kokaly et al., 2009; Ecarnot et al., 2013; Singh et al., 2015; Serbin et al., 2016). Other LES traits have been shown to be well predicted by NIRS (Ecarnot et al., 2013; Kattenborn et al., 2017, 2019). In addition, LNC, another LES trait, can also be predicted using light reflectance at the individual leaf and at canopy levels (Sims and Gamon, 2002). Other traits related to resource-use and conservation can be predicted with spectroscopy, such as leaf age and photosynthetic capacity (Doughty et al., 2011; Chavana-Bryant et al., 2017). Thus, NIRS can provide estimates of integrated properties, such as trait covariations, whole-plant traits, and
TABLE 1 | Prediction accuracy for functional traits.

| Variable                  | Calibration | Validation |
|---------------------------|-------------|------------|
|                           | n           | SD        | $R^2$ | RMSE  | Bias  | Slope | RPD  |
| LDMC (mg g$^{-1}$)         | 2,932       | 52.73     | 0.86  | 16.10 | 0.38  | 1.06  | 3.28 |
| SLA (mm$^2$ mg$^{-1}$)     | 3,423       | 20.90     | 0.85  | 7.47  | 0.14  | 1.01  | 2.80 |
| LNC (%)                   | 1,961       | 2.18      | 0.93  | 0.53  | −0.06 | 0.97  | 4.12 |
| Leaf thickness (μm)        | 4,143       | 178.08    | 0.89  | 69.49 | 2.79  | 1.02  | 2.56 |
| RWC (%)                   | 1,421       | 22.06     | 0.17  | 4.52  | 0.40  | 1.27  | 4.88 |
| LCC (%)                   | 1,960       | 4.78      | 0.65  | 1.17  | 0.03  | 0.86  | 4.10 |
| δ$^{13}$C                  | 1,222       | 1.59      | 0.83  | 0.62  | −0.04 | 0.95  | 2.56 |
| δ$^{15}$N                 | 1,223       | 3.76      | 0.28  | 1.83  | −0.13 | 0.82  | 2.06 |
| Plant lifespan (days)      | 1,403       | 10.55     | 0.17  | 8.01  | −1.31 | 0.86  | 1.32 |
| Plant growth rate (mg d$^{-1}$) | 701     | 0.01      | 0.53  | 0.00  | 0.00  | 0.96  | 1.94 |
| C score (%)               | 2,905       | 10.25     | 0.88  | 3.28  | −0.02 | 1.03  | 3.13 |
| S score (%)               | 2,905       | 11.64     | 0.75  | 2.57  | 0.19  | 1.11  | 4.53 |
| R score (%)               | 2,905       | 17.03     | 0.87  | 4.79  | 0.33  | 0.99  | 3.55 |

LDMC, leaf dry matter content; SLA, specific leaf area; LNC, leaf nitrogen content; RWC, relative water content; LCC, leaf carbon content; δ$^{13}$C, fraction of $^{13}$C isotope; and δ$^{15}$N, fraction of $^{15}$N isotope. CSR scores were estimated from leaf traits by the algorithm from Pierce et al. (2017). n is the total number of leaves used for modeling from our database that are associated with both trait and spectra measurements. All predictions have been obtained from convolutional neural network (CNN) models (see Supplementary Material for details). SD, standard deviation; RMSE, root mean square deviation; and RPD, relative percent difference.

MEASURING PLANT RESPONSES TO THE ENVIRONMENT WITH NIRS

Large-scale comparisons of ecological strategies have been performed with large databases of trait values measured on many species under various conditions, from lab benches to greenhouse, common garden, and field conditions (Kattge et al., 2020). Although these trait databases are used to interpret plant adaptation to the environment, they surprisingly contain very little information about the response of the measured plant properties (demographics, growth rate, and traits) to the environment (Salguero-Gómez et al., 2018). Indeed, comparative approaches generally focus on interspecific variation, considering a mean trait value per species and neglecting intraspecific variability and phenotypic plasticity (but see Albert et al., 2010, 2011). For instance, CSR strategies, which should reflect environmental specialization and specific stress resistance, still remain largely unconnected to the plant evolutionary responses to biotic and abiotic stresses (Takou et al., 2018).

Spectral measurements are widely used to design screening protocols for plant drought responses (Shepherd and Walsh, 2007; Barradas et al., 2021; Burnett et al., 2021). For example, Cabrera-Bosquet et al. (2011) used spectra to accurately predict genotypic differences in the kernel and leaf water content in maize grown under different water treatments. In addition, spectral measurement is a promising method for detecting the severity of damage caused by pathogens, especially for those leaving no visible signs (Spinelli et al., 2004; Sabatier and Rutherford, 2013). Indeed, healthy plants interact (absorb, reflect, emit, transmit, and fluoresce) with electromagnetic radiation in a manner different from that of infected or damaged plants (Li et al., 2014).

To further explore the potential of NIRS as a predictive tool of plant stress level, we used experimental data included in our database (Supplementary Material) from 30 genotypes of Arabidopsis thaliana subjected to water deficit combined with either high or low (freezing) temperatures (Estarague et al., 2021). All plant individuals were measured for leaf NIRS in the course of the treatment, and survival was visually recorded after the treatment. Both measured and CNN-predicted survival rates varied from 14 to 80% depending on the genotype, with an estimated accuracy of survival prediction of 91% in an external validation dataset (Table 2; Figure 2A). Importantly, spectral measurements...
TABLE 2 | Prediction accuracy for five plant categories.

| Category                        | Calibration accuracy (%) | Validation accuracy (%) |
|---------------------------------|--------------------------|-------------------------|
| Survival (2)                    | 0.988                    | 0.915                   |
| Genotypes (10)                  | 0.831                    | 0.640                   |
| Indoor/Outdoor (2)              | 0.998                    | 1.000                   |
| CSR categories (11)             | 0.980                    | 0.700                   |
| Treatment (2)                   | 0.955                    | 0.714                   |

Plant survival has two categories (dead or alive), which were measured according to the protocol described in Estarague et al. (2021). Genotypes have 10 categories corresponding to the 10 natural accessions used here. Indoor/outdoor represents whether a plant has been grown in a greenhouse or growth chamber (indoor) or in a common garden (outdoor) across all the experiments included in the database used here. CSR categories are the intermediate CSR classes estimated from leaf traits by the algorithm from Pierce et al. (2017), such as R/SR, S/SC, RS, and C/CSR (see Supplementary Material). Treatment has two categories (control and water stress) from the dedicated experiments included in the database (see Supplementary Material). All predictions have been obtained from CNN models.

FIGURE 1 | Predictions of the leaf economics spectrum and CSR strategies. Log_{10} relationships between specific leaf area (SLA, mm² mg⁻¹) and leaf nitrogen content (LNC, %; A): between leaf nitrogen content (LNC, %) and leaf dry matter content (LDMC, mg g⁻¹); B: Only predicted values in the validation dataset (1/4 of the whole dataset, n=123) were plotted here. Observed trait values are colored in blue and predicted trait values are colored in red. Regression lines have been estimated by standard major axis (SMA). P is the p value of the SMA test of slope difference between observed and predicted relationships. (C) 3D representation of the leaf economics spectrum between observed and predicted trait values in the validation dataset (n=123). (D) CSR triangle between observed and predicted trait values in the validation dataset (n=699) depicting the variation of plant ecological strategies between competitive ability (C), stress-tolerance (S), and ruderalism (R). CSR scores (%) have been measured from leaf traits following the method from Pierce et al. (2017) (see Supplementary Material). Only measurements performed on fully expanded but non-senescing leaves, and only under non stressing conditions, were used here.

were taken during the treatment, before individuals started showing visible signs of death (Estarague et al., 2021). This suggests that NIRS is a powerful tool to estimate stress effects leading to plant death early on, even before any visible sign of adverse effects.

Convolutional neural network models were able to accurately predict the environmental treatment in which plants were grown (control vs. water stress; prediction accuracy = 71%, Table 2). More surprisingly, CNN models reached 100% accuracy to predict if a plant was grown indoor (growth chamber and greenhouse) vs. outdoor (common garden; Figure 2B; Table 2). This result not only demonstrates the capacity of NIRS and deep learning to characterize the environmental conditions in which plants are cultivated but also suggests that plants grown indoor and outdoor have very contrasted spectral signatures. In turn, these questions our ability to draw conclusions about plant adaptation in natural conditions from experiments led in controlled conditions (growth chamber and greenhouse).
METABOLOMICS AS A NEW PHENOTYPIC DIMENSION: FUTURE PERSPECTIVES FOR THE CHARACTERIZATION OF PLANT ECOLOGICAL STRATEGIES

A broader screening of the metabolic pathways involved in the physiological adaptation of plants to contrasting environments is a promising avenue for ecology in the future. So far, large comparative approaches remain limited by the type and availability of traits collected from the literature and organized into shared databases (Kattge et al., 2020). This constraint reduces our ability to fully understand the drivers of phenotypic diversity, as well as to identify new and ecologically meaningful axes of phenotypic variation. In this perspective, NIRS allows us to detect a large variety of commonly measured chemical compounds such as phosphorus (P)—a key element of the leaf economics spectrum—and base cations [calcium (Ca), potassium (K), and magnesium (Mg)], and other micronutrients (Cozzolino et al., 2001; Ortiz-Monasterio et al., 2007; Galvez-Sola et al., 2015; Ercioglu et al., 2018; de Oliveira et al., 2019; Yu et al., 2019; Prananto et al., 2021). This opens new avenues to link resource-use strategies with plant elemental composition, fluxes, stoichiometry, and beyond, with nutrient cycling in ecosystems (Ustin et al., 2004). In addition, studies have shown that not only LNC but also chlorophyll a and b can be predicted using reflectance and transmittance of light from individual leaves and at canopy level (Sims and Gamon, 2002).

Using quantitative measurements of 67 metabolites with GC–MS and LC–MS (Supplementary Material), we investigated whether NIRS can estimate variations in foliar content of sugars (e.g., glucose and fructose), hormones (e.g., salicylic acid, auxin, and abscisic acid), and secondary metabolites (e.g., phenolic compounds and glucosinolates). Our results show that prediction accuracy (estimated in an external dataset; Supplementary Material) was highly variable between metabolites. For instance, validation $r^2$ ranged from 0% for the poorest predictions (see examples in Table 3) to 85% for the highest (dihydro caffeoyl glucuronide; Table 3). For sugars, the best predictions were obtained for fructose, cellobiose, mannose, and raffinose (Table 3). Among hormones, only auxin (IAA) and jasmonic acid (JA) were satisfactorily predicted by NIRS (Table 3), while other hormones were very poorly predicted (for instance, $r^2<0.10$). Glucosinolates are a class of metabolites produced by the Brassicaceae family, which are involved in plant defense against herbivores (Ratzka et al., 2002). Many of them showed relatively high prediction accuracy (e.g., glucoraphenin and neoglucobrassicin with $r^2>0.70$; Table 3), which paves the way for predicting plant responses to herbivore attack on many individuals at low cost. Finally, many other secondary metabolites showed substantial prediction accuracy (e.g., $r^2>50$; Table 3), although prediction accuracy was very variable between metabolites. More studies are needed to fully explore the potential of NIRS and deep learning to predict leaf chemistry and metabolisms. However, applying NIRS—coupled with deep learning computation—for high-throughput phenotyping of plants from cellular level to whole-plant level is perhaps the most exciting perspective of this approach.
| Variable                  | Calibration validation |
|---------------------------|------------------------|
|                           | $SD$ | $R^2$ | RMSE | Bias | Slope | RPD  |
| **Sugars**                |      |       |      |      |       |      |
| Glucose                   | 6764.56 | 0.14 | 1621.88 | −4.49 | 0.95 | 4.17 |
| Fructose                  | 10240.92 | 0.56 | 1316.93 | 352.08 | 1.17 | 7.78 |
| Sucrose                   | 11380.72 | 0.00 | 2086.69 | 538.48 | −12.55 | 5.45 |
| Fucose                    | 28.65 | 0.03 | 1.90 | 0.37 | 0.75 | 15.04 |
| Isomaltose                | 26.02 | 0.16 | 6.58 | 1.44 | 1.41 | 3.95 |
| Cellobiose                | 157.51 | 0.39 | 73.21 | 19.87 | 1.85 | 2.15 |
| Arabinose                 | 37.57 | 0.00 | 51.42 | 9.39 | 100.65 | 0.73 |
| Galactose                 | 293.66 | 0.16 | 304.29 | 82.21 | 1.11 | 0.97 |
| Inositol                  | 911.06 | 0.31 | 158.28 | 23.17 | 1.29 | 6.29 |
| Maltose                   | 58.40 | 0.02 | 57.31 | 19.37 | 0.86 | 1.02 |
| Mannose                   | 219.79 | 0.42 | 35.78 | 12.77 | 2.19 | 6.14 |
| Raffinose                 | 644.65 | 0.57 | 457.00 | 112.77 | 1.12 | 1.41 |
| Rhamnose                  | 68.56 | 0.02 | 95.56 | 17.09 | −1150.74 | 0.72 |
| Ribose                    | 32.35 | 0.00 | 42.17 | 13.41 | 138.61 | 0.77 |
| Palatinose                | 236.89 | 0.00 | 294.60 | 36.80 | −5.60 | 0.80 |
| Melizitose                | 15.62 | 0.38 | 7.47 | 1.31 | 1.26 | 2.09 |
| Melibiose                 | 200.00 | 0.09 | 264.69 | 47.47 | 0.69 | 0.76 |
| Trehalose                 | 176.00 | 0.00 | 148.34 | 23.78 | −1.69 | 1.20 |
| **Hormones**              |      |       |      |      |       |      |
| Xylose                    | 36.75 | 0.13 | 7.09 | 1.54 | 1.32 | 5.04 |
| ABA                       | 12.54 | 0.06 | 11.25 | 1.43 | 0.57 | 1.12 |
| IAA                       | 21.37 | 0.26 | 18.16 | 1.84 | 0.95 | 1.18 |
| JA                        | 337.70 | 0.29 | 197.91 | 31.53 | 1.03 | 1.71 |
| SA                        | 799.00 | 0.00 | 495.41 | 147.44 | −10.54 | 1.61 |
| CMLX                      | 7277.61 | 0.02 | 8086.67 | 2421.27 | 63.66 | 0.90 |
| **Glucosinolates**        |      |       |      |      |       |      |
| Glucoalysin               | 28.79 | 0.10 | 27.76 | 3.95 | 1.05 | 1.04 |
| Glucobrassicin            | 1462.69 | 0.15 | 914.32 | 210.01 | 0.76 | 1.60 |
| Glucoerucin               | 12.22 | 0.39 | 5.88 | 0.51 | 0.86 | 2.08 |
| Glucoraphanin             | 5005.90 | 0.00 | 4703.53 | 2123.30 | 0.43 | 1.06 |
| Glucosinurin              | 94.96 | 0.00 | 91.73 | 12.46 | 0.62 | 1.03 |
| Glucoraphanin             | 1865.98 | 0.00 | 166.48 | 221.14 | 0.22 | 1.12 |
| Glucoraphanin             | 1.78 | 0.74 | 0.62 | 0.07 | 0.91 | 2.88 |
| Epigallocatechin          | 210.86 | 0.27 | 163.05 | 2.91 | 0.83 | 1.29 |
| Progolitrin               | 666.26 | 0.01 | 564.65 | 135.83 | 0.38 | 1.18 |
| Epiprogoitrin             | 6316.22 | 0.09 | 5944.42 | 1814.64 | 0.74 | 1.06 |
| Isobutyl                  | 473.57 | 0.03 | 356.50 | 56.56 | 0.67 | 1.33 |
| Glucosinibin              | 10.35 | 0.00 | 7.96 | 1.28 | 2.52 | 1.30 |
| Sinigrin                  | 4445.20 | 0.07 | 4259.39 | 1571.86 | 1.04 | 1.04 |
| Hexyl                     | 49.96 | 0.00 | 45.61 | 12.28 | 0.53 | 1.10 |
| Butyl                     | 5.49 | 0.51 | 3.20 | −0.24 | 1.07 | 1.72 |
| Neoglucobrassicin Peak1   | 265.97 | 0.73 | 273.80 | 59.08 | 1.86 | 0.97 |
| Neoglucobrassicin Peak2   | 1051.25 | 0.06 | 254.92 | 24.16 | 0.41 | 4.12 |
| X3MTP                     | 47.48 | 0.51 | 9.63 | 0.36 | 1.41 | 4.93 |
| X5MTP                     | 20.76 | 0.61 | 11.56 | 1.14 | 1.40 | 1.80 |
| X6MSH                     | 51.83 | 0.22 | 48.64 | 9.55 | 1.00 | 1.07 |
| X7MTH                     | 261.68 | 0.18 | 277.93 | 88.23 | 1.19 | 0.94 |
| X7MTH                     | 244.30 | 0.36 | 224.81 | 36.56 | 1.04 | 1.09 |
| X8MTO                     | 2013.33 | 0.31 | 1528.42 | 169.92 | 0.87 | 1.32 |
| Apigenin rutinoside       | 1278.38 | 0.17 | 1053.50 | 176.17 | 0.85 | 1.21 |
| **Other secondary metabolites** | | | | | | |
| Caffeic Acid              | 30.01 | 0.32 | 0.96 | −0.20 | 0.74 | 31.31 |
| Chlorogenic Acid          | 29.55 | 0.66 | 16.29 | 1.38 | 1.09 | 1.81 |
| Citrin                    | 2647.54 | 0.44 | 1894.98 | 169.09 | 1.08 | 1.40 |
| Cyanidin rhamnoside       | 1431.34 | 0.53 | 842.46 | −56.16 | 0.81 | 1.70 |
| Cyanidin sophorosid glucose | 674.85 | 0.31 | 387.08 | 88.61 | 1.04 | 1.74 |
| Dihydro caffeoyl glucuronide | 27.05 | 0.85 | 8.96 | 0.01 | 1.12 | 3.02 |
| Fumarat                   | 294.76 | 0.10 | 174.41 | 18.17 | 0.68 | 1.69 |
| Kaempferol glucosyl       | 989.20 | 0.14 | 518.91 | 97.70 | 0.69 | 1.91 |
| Rhamnosyl glucoside       |      |       |      |      |       |      |

(Continued)
CONCLUSION

In this paper, we argue that NIRS coupled with recent advances in deep learning approaches is a promising method to broadly capture various information about plant functioning, ecological strategies, response to environment, and metabolism. In particular, NIRS affords considerable time and cost savings (spectrum acquisition lasts only a few seconds), and without using hazardous chemicals. In addition, samples can be analyzed in neither their natural form without destruction nor any special sample preparation. Thus, NIRS makes it possible to create extensive databases of traits at different temporal, spatial, and taxonomic scales and facilitate the adoption of phenomics into ecology. It might provide a reliable tool for the characterization of plant populations across geographical ranges, specifically if combined with other omics approaches and deep learning computation. Of course, developing calibration equations takes time, but selecting a suitable subset of samples to use in the calibration equation and validating the calibration equation take only a matter of hours in addition to standard laboratory work to chemically analyze the subset. Clearly, NIRS is more suited for larger data sets than those containing only a few samples. As calibration equations keep available for future studies, the time and financial cost of calibrations will decrease. Thus, adopting NIRS in trait-based ecology would literally multiply the number of species, genotypes, and environments potentially measurable, a key point to link functional trait variation to plant physiology and adaptation.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

FV led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

FUNDING

This work was supported by INRAE, CNRS, the French Agency for Research (ANR grant ANR-17-CE02-0018-01; “AraBreed” to FV, DV, and CV), Région Occitanie (FEDER FSE IEJ 2014-2020; Project PHENOPSIS 2.0), and the European Research Council (ERC;"CONSTRAINTS": grant ERC-StG-2014-639706-CONSTRAINTS to CV). Metabolite analytics were funded by the DeutscheForschungsgemeinschaft (DFG, German Research foundation) – Projektnummer INST 37/696-1 FUGG.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.836488/full#supplementary-material

REFERENCES

1001 Genomes Consortium (2016). 1135 sequenced natural inbred lines reveal the global pattern of polymorphism in Arabidopsis thaliana. Cell 166, 481–491. doi: 10.1016/j.cell.2016.05.063

Albert, C. H., Grassein, F., Schurr, F. M., Vieilledent, G., and Violle, C. (2011). When and how should intraspecific variability be considered in trait-based plant ecology? Perspect. Plant Ecol. Evol. Syst. 13, 217–225. doi: 10.1016/j.ppees.2011.04.003

Albert, C. H., Thuiller, W., Voccioz, N. G., Soudant, A., Boucher, F., Saccone, P., et al. (2010). Intraspecific functional variability: extent, structure and
sources of variation. *J. Ecol.*, 98, 604–613. doi: 10.1111/1365-2745.2010.01651.x

Anderegg, L. D. L., Berner, L. T., Badgley, G., Sethi, M. L., Law, B. E., and HilleRisLambers, J. (2018). Within-species patterns challenge our understanding of the leaf economics spectrum. *Ecol. Lett.*, 21, 734–744. doi: 10.1111/eol.12945

Andrés, S., Giráldez, F. J., López, S., Mantecon, À. R., and Calleja, A. (2005). Nutritional evaluation of herbage from permanent meadows by near-infrared reflectance spectroscopy: 1. Prediction of chemical composition and vitamin digestibility. *J. Sci. Food Agric.*, 85, 1564–1571. doi: 10.1002/jsfa.2138

Arslan, M., Xiaobo, Z., Xuetao, H., Tahir, H. E., Shi, J., Khan, M. R., et al. (2018). Near infrared spectroscopy coupled with chemometric algorithms for predicting chemical components in black goji berries (*Lycium ruthenicum Murr.*). *J. Near Infr. Spectrosc.*, 26, 275–286. doi: 10.1017/S0967703318975997

Asner, G. P., and Martin, R. E. (2008). Spectral and chemical analysis of tropical forests: scaling from leaf to canopy levels. *Remote Sens. Environ.*, 112, 3958–3970. doi: 10.1016/j.rse.2008.07.003

Asner, G. P., Martin, R. E., Ford, A. J., Metcalfe, D. J., and Liddell, M. J. (2009). Leaf chemical and spectral diversity in Australian tropical forests. *Ecol. Appl.*, 19, 236–253. doi: 10.1890/08-0023.1

Barradas, A., Correia, P. M. P., Silva, S., Mariano, P., Pires, M. C., Matos, A. R., et al. (2021). Comparing machine learning methods for classifying plant drought stress from leaf reflectance spectra in *Arabidopsis thaliana*. *NATO Adv. Sci. Inst. Ser. E Appl. Sci.* 61:3692. doi: 10.3920/app11146392

Beale, D. J., Kay, J., and Ahmed, W. (2016). “Beyond metabolomics: a review of multi-omics-based approaches” in *Microbial Metabolomics: Applications in Clinical, Environmental, and Industrial Microbiology*. eds. D. J. Beale, K. A. Kouremenos and E. A. Palombo (Cham: Springer International Publishing), 289–312.

Biancolillo, A., and Marini, F. (2018). Chemometric methods for spectroscopy-based pharmaceutical analysis. *Front. Chem.*, 6:378. doi: 10.3389/fchem.2018.00376

Birth, G. S., and Hecht, H. G. (1987). “The physics of near-infrared reflectance” in *Near-Infrared Technology in the Agricultural and Food Industries*. ed. M. N. Saint-Paul (American Association of Cereal Chemists, Inc.), 1–15.

Burruto, A. C., Serbin, S. P., Davidson, K. J., Ely, K. S., and Rogers, A. (2021). Detection of the metabolic response to drought stress using hyperspectral reflectance. *J. Exp. Bot.*, 72, 6474–6489. doi: 10.1093/jxb/erab255

Cabrera-Bosquet, L., Sánchez, C., Rosales, A., Palacios-Rojas, N., and Araus, J. L. (2011). Near-infrared reflectance spectroscopy (NIRS) assessment of 818O and nitrogen and ash contents for improved yield potential and drought adaptation in maize. *J. Agric. Food Chem.*, 59, 467–474. doi: 10.1021/jf103395z

Chan, E. K. F., Rowe, H. C., Hansen, B. G., and Kliebenstein, D. J. (2010). Leaf aging of *Arabidopsis thaliana*: a latitudinal gradient or a center-margins differentiation of ecological strategies in *Arabidopsis thaliana* bioRxiv [Preprint]. doi: 10.1121/2021.10.416205

Exposito-Alonso, M. (2020). Seasonal timing adaptation across the geographic range of *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* 117, 9665–9667. doi: 10.1073/pnas.1921798117

Exposito-Alonso, M., Exposito-Alonso, M., Gómez-Rodriguez, R., Barragán, C., Capovilla, G., Chae, E., et al. (2019). Natural selection on the *Arabidopsis thaliana* genome in present and future climates. *Nature* 573, 126–129. doi: 10.1038/s41586-019-1520-3

Faqurghar, G. D., Ehleringer, J. R., and Hubick, K. T. (1989). Carbon isotope discrimination and photosynthesis. *Annu. Rev. Plant Biol.*, 40, 503–537. doi: 10.1146/annurev.pp.40.060189.002443

Foley, W. J., McAllwee, A., Lawler, I., Aragones, L., Woolnough, A. P., and Berding, N. (1998). Ecological applications of near infrared reflectance spectroscopy—a tool for rapid, cost-effective prediction of the composition of plant and animal tissues and aspects of animal performance. *Oecologia* 116, 293–305. doi: 10.1007/s004420050591

Fu, P., Meacham-Hensold, K., Guan, K., Wu, I., and Bernacchi, C. (2020). Estimating photosynthetic traits from reflectance spectra: a synthesis of spectral indices, numerical inversion, and partial least square regression. *Plant Cell Environ.*, 43, 1241–1258. doi: 10.1111/pce.13718

Galvez-Sola, L., García-Sánchez, F., Pérez-Pérez, J. G., Gimeno, V., Navarro, J. M., Moral, R., et al. (2015). Rapid estimation of nutritional elements on citrus leaves by near infrared reflectance spectroscopy. *Front. Plant Sci.*, 6:571. doi: 10.3389/fpls.2015.00571

Garnier, E., Navas, M.-L., and Grigulis, K. (2016). *Plant Functional Diversity: Organism Traits, Community Structure, and Ecosystem Properties*. United Kingdom: Oxford University Press.

Gitelis, A. A., and Merzlyak, M. N. (1997). Remote estimation of chlorophyll content in higher plant leaves. *Int. J. Remote Sens.*, 18, 2691–2697. doi: 10.1080/0143116972175358

Grime, J. P. (1974). Vegetation classification by reference to strategies. *Nature* 245, 26–31. doi: 10.1038/250026a0

Grime, J. P. (1977). Evidence for the existence of three primary strategies in plants and its relevance to ecological and evolutionary theory. *Ann. Nat.*, 111, 1169–1194. doi: 10.1086/283244

Grime, J. P. (1988). Plant strategies and the dynamics and structure of plant-communities. *Nature* 336:630. doi: 10.1038/336630a0

Jacquemoud, S., Verhoeven, S., Baret, F., Bacour, C., Zarco-Tejada, P. J., Asner, G. P., et al. (2009). PROSPECT+SAIL models: A review of use for vegetation characterization. *Remote Sens. Environ.*, 113, S56–S66. doi: 10.1016/j.rse.2008.01.026

Kattenborn, T., Fassnacht, F. E., Pierce, S., Lopatin, J., Grime, J. P., and Schmidtlein, S. (2017). Linking plant strategies and plant traits derived by reflectance transfer modelling. *J. Veg. Sci.*, 28, 717–727. doi: 10.1111/jvs.12525

Kattenborn, T., Fassnacht, F. E., and Schmidtlein, S. (2019). Differentiating plant functional types using reflectance: which traits make the difference? *Remote Sens. Ecol. Conserv.*, 5, 5–19. doi: 10.1002/rsa.2086

Katteg, J., Bönisch, G., Díaz, S., and Lavoerl, S. (2020). TRY plant trait database—enhanced coverage and open access. *Global Chang. Biol.*, 26, 119–188. doi: 10.1111/gcb.14904

Keddy, P. A. (1992). A pragmatic approach to functional ecology. *Funct. Ecol.*, 6, 621–626. doi: 10.2307/2389954

Kleinebecker, T., Schmidt, S. R., Fritz, C., Smolders, A. J. P., and Holzel, N. (2009). Prediction of 613C and 815N in plant tissues with near-infrared
scaling allometry. *Ecol. Lett.* 15, 1149–1157. doi: 10.1111/j.1461-0248.2012.01839.x

Violle, C., Navas, M. L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I., et al. (2007). Let the concept of trait be functional! *Oikos* 116, 882–892. doi: 10.1111/j.0030-1299.2007.15559.x

Violle, C., Reich, P. B., Pacala, S. W., Enquist, B. J., and Kattge, J. (2014). The emergence and promise of functional biogeography. *Proc. Natl. Acad. Sci.* 111, 13690–13696. doi: 10.1073/pnas.1415442111

Wadoux, A. M. J.-C., Heuvelink, G. B. M., de Bruin, S., and Brus, D. J. (2021). Spatial cross-validation is not the right way to evaluate map accuracy. *Ecol. Model.* 457:109692. doi: 10.1016/j.ecolmodel.2021.109692

Wójcicki, K. (2015). Application of NIR spectroscopy for whisky identification and determination the content of ethanol. *Curr. Trends Commod. Sci.* 123, 123–133.

Wold, S., Martens, H., and Wold, H. (1983). “The multivariate calibration problem in chemistry solved by the PLS method,” in *Matrix Pencils* (Berlin Heidelberg: Springer), 286–293.

Wright, I. J., Reich, P. B., Westoby, M., Ackerly, D. D., Baruch, Z., Bongers, F., et al. (2004). The worldwide leaf economics spectrum. *Nature* 428, 821–827. doi: 10.1038/nature02403

Wu, J., Chavana-Bryant, C., Prohaska, N., Serbin, S. P., Guan, K., Albert, L. P., et al. (2016). Convergence in relationships between leaf traits, spectra and age across diverse canopy environments and two contrasting tropical forests. *New Phytol.* 214, 1033–1048. doi: 10.1111/nph.14051

Wu, S., Tohge, T., Cuadros-Inostroza, Á., Tong, H., Tenenboim, H., Kooke, R., et al. (2018). Mapping the Arabidopsis metabolic landscape by untargeted metabolomics at different environmental conditions. *Mol. Plant* 11, 118–134. doi: 10.1016/j.molp.2017.08.012

Yu, E., Zhao, R., Cai, Y., Huang, J., Li, C., Li, C., et al. (2019). Determination of manganese content in cottonseed meal using near-infrared spectrometry and multivariate calibration. *J. Cotton Res.* 2:12. doi: 10.1186/s42397-019-0030-5

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Vasseur, Cornet, Beurier, Messier, Rouan, Bresson, Ecarnot, Stahl, Heumos, Gérard, Reijnen, Tillard, Lacombe, Emanuel, Floret, Estarague, Przybylska, Sartori, Gillespie, Baron, Kazakou, Vile and Violle. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.