5

A Workflow for Metabolomics of Forest Tree Biotic Stress Response and Applications for Management

Anna O. Conrad[1], *Caterina Villari*[2], *and Pierluigi Bonello*[3]

[1] *USDA Forest Service, Northern Research Station, West Lafayette, IN, USA*
[2] *Warnell School of Forestry & Natural Resources, University of Georgia, Athens, USA*
[3] *Department of Plant Pathology, The Ohio State University, Columbus, USA*

5.1 Introduction

Trees are exposed to a variety of stressors over their lifetimes. To survive and thrive under diverse conditions, they have evolved a suite of defenses, which confer resistance or tolerance to most biotic and abiotic stressors. Even though disease (or insect pest infestation) is the exception and not the rule, pests and pathogens have had and continue to have a profound impact on the health of forests globally. This is especially true for invasive and/or non-native pests and pathogens and, more recently, for native pests and pathogens whose range or virulence has changed because of climate change (1–3). Any estimate of the impact of pests and pathogens on forests is likely conservative and an underestimate of the actual impact on forests and the communities that rely on them. The incidence of non-native insects (and pathogens) continues to increase every year (4), and increased tree mortality within just the United States is predicted in 63% of forestland due to non-native insect pests and pathogens (5). Over the last 100 years, invasive pests and pathogens have functionally eradicated or decimated multiple forest tree species in local populations or across species ranges. Examples of impacted forest tree species include, but are not limited to, chestnut (chestnut blight), pines (white pine blister rust and mountain pine beetle), ash (emerald ash borer and ash dieback), and tanoak, oaks, and larch (sudden oak death and Ramorum blight).

The impacts of pests and pathogens are widespread, from economic losses to loss of forest ecosystem services and negative outcomes for human health. For example, emerald ash borer, which was first identified in the early 2000s,

Monitoring Forest Damage with Mass Spectrometry-Based Metabolomics Methods, First Edition. Edited by Carla António.
© 2024 John Wiley & Sons, Inc. Published 2024 by John Wiley & Sons, Inc.

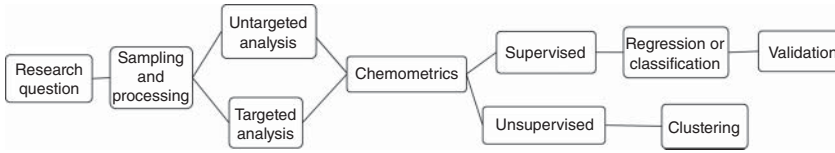

Figure 5.1 A workflow for metabolomics studies of forest tree biotic stress interactions. Typical research questions focus on identifying resistant trees or detecting diseased or pest-infected trees based on metabolomics data.

is estimated to have killed tens of millions of ash trees in the eastern United States (6). In areas of emerald ash borer infestation, increased human mortality from cardiovascular illness has been reported (7). Impacts have also been felt in the housing industry, where emerald ash borer infection negatively impacted home values in one Midwestern United States city (8). These studies do not even include more direct impacts, such as costs of forest management for the treatment, removal, or replacement of susceptible ash trees (9). Wide-reaching impacts are not exclusive to the emerald ash borer. Similar impacts have been observed or estimated for other pathogens and pests of trees across wildlands and urban landscapes (10, 11).

To counter biotic threats, trees possess a suite of defenses that act directly and/or indirectly on the attacking pest or pathogen, and range from physical defenses, including bark and lignin, to chemical defenses, such as phenolics and terpenoids. These defenses can be general or specific for a given organism and can be produced constitutively (pre-infection) or induced following infection. The reason why invasive and non-native pests and pathogens have been particularly problematic is because affected tree species lack a co-evolutionary history with them, so the pests and pathogens have no co-evolved natural enemies. Thus, the suite of defenses that the novel host possesses is unable to prevent infection or minimize adverse impacts to tree health, ultimately, resulting in higher-than-normal levels of mortality of impacted host tree species, often across larger spatial scales.

In this chapter, we focus on the role of metabolite defenses in forest tree responses to biotic stress. Within this context, we describe a workflow (Figure 5.1) for use in metabolomics studies of forest tree-pest/pathogen interactions. Applications for the workflow as well as considerations for applying it across scales, from individual trees to landscapes, are also discussed. Finally, case studies from multiple tree-pathogen and pest systems are presented, including oak (sudden oak death and European oak leafroller), ash (emerald ash borer and ash dieback), and pine (Diplodia tip blight and mountain pine beetle).

5.2 Methods

5.2.1 Research Question

The research question(s) will dictate the experimental approach and statistical methods employed in a study of forest tree biotic stress response. Questions can direct the focus of study to all metabolites produced in a biological system (i.e. the metabolome of a host tree), or to a sub-set (i.e. metabolite profiling). Metabolite profiling is of relevance to studies of forest tree biotic stress response because metabolite profiling aims to identify metabolites correlated with a specific response, such as infection or disease resistance (12, 13). In some situations, a more rapid method of assessing the metabolome of a host tree is needed. This is referred to as metabolic fingerprinting (12). In contrast to metabolite profiling, individual metabolites are not identified or quantified. For this reason, metabolic fingerprints can be used for rapid classification (12). Key features of the metabolome, such as metabolites associated with specific treatments, might not be resolvable using the latter approach, so follow-up studies will be required for more precise identification and quantification.

5.2.2 Sample Selection and Processing

The life cycles of forest trees and of the pathogens and pests that attack them need to be considered when selecting samples, both in determining optimal tissues to sample (e.g. leaves, stem, or roots) and when to sample (i.e. timing corresponding both with host and pathogen/pest cycles). Metabolites are known to change with the season (tree phenology), age, and position on the tree (e.g. sun versus shade leaves) (14–17), so considerations for standardizing sample collection, regardless of whether the focus is on constitutive or induced metabolites, is necessary.

Under optimal conditions, samples are harvested and flash frozen (in liquid nitrogen) and then stored in a low-temperature freezer. This minimizes sample degradation. However, where liquid nitrogen is not available, or its use is unfeasible (e.g. in rugged field conditions), dry ice can be used to preserve samples until returning to the laboratory.

Tissue homogenization is the next step, and the best approach will depend on the tissue. Woody tissues can be difficult to homogenize. Lyophilizing (freeze drying) samples is sometimes useful or needed. Homogenization can be accomplished by hand (with mortar and pestle and liquid nitrogen to maintain frozen tissue) or with a ball mill. When sampling bark or wood from mature trees, using a cordless hand drill to sample tissues is useful for pre-grinding samples and making subsequent homogenization easier. Various extraction methods

can be employed depending on the metabolites of interest, although methanol is widely used as a solvent for extraction of biotic stress-related metabolites; for further examples and protocols, see (18–20).

5.2.3 Analytical Methods

The analytical method used will depend on the research question and which metabolites or metabolite classes are of interest to the system. For example, phenolic compounds are well known for their association with biotic stress response in forest trees (21, 22). The analytical approach can be classified into two categories: untargeted and targeted. Sedio et al. (23) provide a protocol for high-throughput, untargeted metabolomics analysis. A targeted approach focuses on either a class of metabolites or specific, individual metabolites.

For studies of metabolites in forest trees subjected to biotic stress, liquid chromatography mass spectrometry (LC–MS) or gas chromatography mass spectrometry (GC–MS)-based approaches are the norm for separating out individual compounds for subsequent identification and quantification; see (24) for a comprehensive review of the topic. GC–MS is suitable for volatile compounds or compounds that can be derivatized, like terpenes (25). MS libraries can aid in compound identification, although libraries are more well-developed for GC–MS versus LC–MS (24). Internal standards are used for quantification and normalization, and to ensure technical accuracy, such as extraction efficiency and run-to-run consistency (26). External standards are also useful for quantifying individual compounds and for confirming the presence of specific metabolites in samples based on retention time and spectral properties (26).

5.2.4 Chemometrics

Chemometrics is the multivariate analysis of chemical data and covers the analysis of different types of metabolomics data from spectral data to MS data sets (27, 28). While univariate approaches (e.g. analysis of variance [ANOVA]), are useful for evaluating treatment effects on individual metabolites, chemometrics is useful for simultaneous analysis of multiple metabolites. Approaches that aim to reduce redundancy in the data, a result of co-linearity, are of particular importance when developing models to predict disease-resistant trees or to discriminate between healthy and infected trees based on metabolite levels. Otherwise, models are at risk of being overfit, which can lead to poor performance when models are applied to new datasets (29).

There are two approaches for chemometric analysis, namely, supervised and unsupervised. Supervised analysis considers a priori groupings and labels the data, while unsupervised analysis does not consider any grouping or

treatment. Unsupervised analysis (e.g. principal component analysis [PCA]) can be useful for identifying natural groupings or clustering of samples and metabolites associated with differences between treatments or groups without considering the group or treatment (28). While both approaches can be used to identify metabolites that vary between treatments or groups, supervised analysis is especially well suited for identifying metabolites most important for distinguishing between groups or identifying metabolites most influenced by treatment effects (30). See Table 5.1 for a summary of common chemometric methods applied to forest tree biotic stress studies.

For studies focused on predicting or modeling specific traits, such as those focused on distinguishing between trees that differ in disease severity or identifying resistant trees, availability of robust and reliable phenotypic data (paired with metabolomics data) is essential. Many forest tree responses to biotic stressors are quantitative in nature. So, symptoms fall along a continuum with trees ranging from resistant (no to limited symptoms) to susceptible (severe symptoms often resulting in death), and progression of symptoms occurs over months to years. For these reasons, accurate phenotyping can be a challenge. Therefore, evaluation of multiple disease or pest severity metrics, both qualitative and quantitative, might be needed.

5.2.5 Pre-processing

Once thresholds for detection are specified, metabolite peaks are identified, and raw data is extracted from analytical software (e.g. XCMS (43)), metabolite data needs to be prepared for analysis (pre-processing). For metabolite profiling studies, functional data analysis (FDA) can be used to identify potential outliers (44). FDA considers the depth of the curve, such that the depth is inversely related to the sample outlyingness. Outliers can also be identified from multivariate analyses, for instance clustering analysis (28). Once outliers are identified and trimmed, standardization and normalization of data are needed. Internal standards are useful for normalizing data (26). Mean centering and scaling, and other types of transformation, including log transformation, can also be used.

5.2.6 Classification

Classification of metabolite data is conducted using supervised or unsupervised approaches. Supervised methods are commonly used for examining forest tree biotic stress responses (see Table 5.1), although unsupervised approaches can be useful for examining natural clustering of samples, without influence of the grouping or treatment. Regardless of whether a priori groupings are considered, classification can be accomplished using multivariate statistical methods or machine learning. Machine learning algorithms learn

Table 5.1 Commonly used chemometric methods in forest tree biotic stress studies.

Method	Type	Supervised or unsupervised	Variable selection	Application	Selected references
Principal component analysis	Classification	Unsupervised, supervised	Yes	Infection, resistance	(26, 31–34)
Support vector machine	Classification	Supervised	No	Resistance	(35, 36)
Random forest	Classification	Supervised	Yes	Resistance, infection	(35, 37)
Linear discriminant analysis	Classification	Unsupervised, supervised	Yes	Resistance	(38, 39)
Partial least squares discriminant analysis	Classification	Supervised	Yes	Infection, resistance	(32, 35, 37, 40)
Soft independent modeling of class analogy	Classification	Supervised	Yes	Resistance	(36, 41, 42)
Logistic regression	Regression	N/A	Yes	Resistance	(38, 39)
Partial least squares regression	Regression	N/A	Yes	Resistance	(36)

from experience and aim to generalize trends or patterns in the data (45–47). Examples of classification methods include PCA (can be supervised or unsupervised), partial least squares discriminant analysis (PLS–DA), linear discriminant analysis (LDA), support vector machine (SVM), and random forest (RF) (Table 5.1).

5.2.7 Regression

Regression is useful for identifying associations between data, where the response variable is quantitative, including continuous and binomial response variables. Regression can also be used to make predictions based on known data. Common regression methods include linear regression (continuous response variable), logistic regression (binomial response variable), and partial least squares regression (PLSR) (Table 5.1). Regression approaches, such as backward stepwise logistic regression, can also be used to identify the most important variables (see Section 5.2.8) (38, 39).

5.2.8 Variable Selection

Reducing the number of variables, especially in metabolomics data sets with co-linearity, or a high number of variables relative to sample size can be essential to building accurate and robust models and preventing overfitting (28,48). In addition, variable selection can be useful for teasing apart biological treatment effects and identifying metabolites most associated with biotic stress response. Various approaches can be used and include classification and regression methods (see Table 5.1). For example, sparse PLS–DA (a variation of PLS–DA) combines classification with variable selection to identify and include only the most important features (49). Even with unsupervised approaches, like PCA, loading values can be used to identify metabolites most important for separating treatment groups. Random forest can also be used for variable selection (e.g. variable selection using random forests [VSURF]) or variable importance score (VIP) (50, 51).

5.2.9 Validation

For classification and regression models, including machine learning approaches, validation is needed to assess model quality and transferability (28). Model performance (e.g. accuracy, precision, and area under the receiver operator characteristic curve) is ideally assessed using external validation (separate data set) or cross-validation. The latter approach is especially useful when sample sizes are limited; see (28) for further information on validation in chemometrics.

5.2.10 Available Tools for Analyzing Metabolomics Data

Various registered software and open-source software packages are available for exploring and analysis of multivariate data, including those tailored specifically to metabolomics data analysis or analysis of complex omics data sets (e.g. Metaboanalyst, mixOmics, mdatools) (30,52–54). Some of these tools can be used directly in a web browser (e.g. Metaboanalyst) (52). In addition, Conrad et al. developed a workflow for the analysis of spectral data for use in disease prediction (35, 55) and resistance screening (36).

5.3 Application

The workflow presented in Section 5.2 (see Figure 5.1 for overview) is customizable depending on the research question. Consequently, the specific methods used within each step will vary depending on the research question and the tree-pathogen/pest system. Two primary applications were tested: (i) identifying infected trees; and (ii) identifying resistant trees. The workflow, however, is broadly applicable and may be useful beyond metabolomics studies. This is because many of the methods used for chemometrics (e.g. PLS–DA) are appropriate for use with other omics datasets (30).

5.3.1 Detection of Diseased and Pest-Infected Trees

Disease and pest management relies on identifying infected trees. This often occurs after symptoms develop, and unfortunately, after the pest or pathogen has become established in an area. Thus, early detection, including identifying infected trees before symptoms develop, is key. This is because the lag time between initial infection and symptom development provides the pathogen or pest time to spread without intervention. Examples include identifying infected but asymptomatic American beech trees with *Litylenchus crenatae mccannii*, a non-native nematode associated with beech leaf disease (35), and distinguishing between symptomatic and asymptomatic Norway spruce infected with the fungus *Herterobasidion* spp. (56).

5.3.2 Identifying Resistant Trees

The workflow can also be used to identify disease or pest-resistant trees. Use of resistant trees is a viable option for disease/pest management, restoration of ecologically important forest tree species, and as part of improvement programs for economically important species (57, 58). Identifying resistant trees using metabolomics data is the primary application of this workflow to date. Examples include identifying coast live oak resistant to sudden oak

death (38, 39, 41), European ash resistant to ash dieback (42), and Austrian pine resistant to Diplodia tip blight (36). Constitutive or induced metabolites can be analyzed. If the aim is to develop a non-invasive screen for resistance (without the need for natural or artificial inoculations), then the focus should be on constitutive metabolites (e.g. (38, 41)).

5.3.3 Landscape-Level Applications

Translating metabolomics tools from individual trees or tree populations to the landscape scale is a topic of interest not only for identifying infected or resistant trees, but more broadly within the fields of high-throughput plant phenotyping (59) and digital forestry. High-throughput and landscape-level assessment of forest tree biotic stress responses can be limited with MS-based approaches, unless applied only to a subset of individuals within a species or population. However, more sensitive MS-based approaches can be integrated with more high-throughput remote sensing platforms for application at larger scales, such as aerial hyperspectral sensing, which is capable of picking up on physiological differences in trees (e.g. water stress associated with pathogen infection) as well as foliar biochemical variation (60). Use of proximal and remote sensing (with spectral sensors) has already been shown to be effective at identifying oak wilt-infected trees at multiple scales (61, 62).

5.4 Case Studies

The following case studies demonstrate the use of metabolomics to examine and predict pest and pathogen infection and resistance within forest trees.

5.4.1 Ash

Ash dieback is caused by the invasive fungal pathogen *Hymenoscyphus fraxineus* and is currently impacting European ash (*Fraxinus excelsior*) across Europe (63). It is reported that some trees are resistant to the disease, and genotypes with resistance have been identified (63, 64). Untargeted metabolomics analysis of leaves from high- and low-susceptibility individuals was performed by Sambles et al. (37) to identify metabolites associated with disease resistance. Partial least squares discriminate analysis was used along with variable influence on projection and RF to determine important metabolites. Putative iridium glycosides were found to differ between resistant and susceptible trees. Nemesio-Gorriz et al. (40) also identified metabolites associated with resistance (tolerance) to ash dieback. Using PLS–DA combined with variable selection, the authors identified that higher levels of fraxetin and escalation were associated with lower disease susceptibility. These compounds

inhibited the growth of *H. fraxineus in vitro*. Finally, constitutive bark spectral profiles of resistant and susceptible European ash grown across six countries could be distinguished using soft independent modeling of class analogy (SIMCA) (42).

Emerald ash borer (*Agrilus planipennis*) is an invasive, non-native insect pest that has destroyed North American ash populations across the eastern United States and Canada (reviewed in (65)). While resistant (lingering) ash have been identified (66), most studies to date have examined metabolite differences between susceptible North American ash species (e.g. green ash, *Fraxinus pennsylvanica* and white ash, *Fraxinus americana*) and resistant Manchurian ash (*Fraxinus mandshurica*). For example, Qazi et al. (67) used untargeted HPLC–MS/MS to examine differences in constitutive foliar metabolite profiles of ash species with differing susceptibility to the pest. Using PCA, the authors identified metabolites whose levels increased in resistant Manchurian ash. Similarly, Chakraborty et al. (26) compared phloem phenolic profiles of resistant Manchurian ash and susceptible black ash (*Fraxinus nigra*) following infection with emerald ash borer and drought stress. Using a multi-faceted approach of PCA, hierarchical cluster analysis (HCA), and ANOVA, the authors discovered that pinoresinol A was induced by emerald ash borer infection and that levels of six other phenolic compounds decreased following infection.

5.4.2 Oak

Since its discovery in the western United States in the 1990s, sudden oak death, caused by the non-native pathogen *Phytophthora ramorum*, has led to the extensive mortality of tanoak (*Notholithocarpus densiflorus*) and true oaks in the red oak group (*Quercus* section *Lobatae*) (68). The pathogen is also found in the United Kingdom, where impacts are felt in particular on plantation-grown larch (69). In the United States, multiple studies have examined the role of metabolites, particularly phenolic compounds, in coast live oak (*Quercus agrifolia*) defense and resistance to *P. ramorum* (38,39,41,70,71). McPherson et al. (39) and Conrad et al. (38) identified phenolic biomarkers of resistance in coast live oak previously exposed to the pathogen and in naïve trees, respectively. To do this, sparse linear discriminant analysis (SLDA) and logistic regression were used to identify biomarkers and predict resistance based on constitutive levels of the identified compounds. In naïve trees, biomarkers of resistance included myrictrin and three incompletely characterized flavonoids (38).

The European oak leafroller (*Tortrix viridana*) causes severe defoliation of pedunculate oak (*Quercus robur*) in Europe. Differences in *Q. robur* susceptibility to *T. viridana* have been observed and two classes of trees with differing susceptibility have been defined: T-oaks (resistant, less defoliation) and S-oaks (susceptible and greater defoliation) (31). Using multivariate data analysis, specifically PCA and PLSR, Kersten et al. (31) identified metabolic differences

between resistant and susceptible trees. The authors found that resistant trees contained higher levels of galloylated flavonol glycosides compared to susceptible trees. Similarly, Bertić et al. (32) identified metabolic biomarkers to distinguish *T. viridana* resistant and susceptible *Q. robur* using PCA and orthogonal partial least squares discriminant analysis (OPLS–DA). One of the goals of this study was to identify biomarkers for use in assisting tree nurseries in seedling selection. For that reason, biomarkers were used to develop prediction models for the selection of resistant trees.

5.4.3 Pine

Diplodia tip blight is caused by the pathogen *Diplodia sapinea* and affects two-needle pine species, including Austrian pine (*Pinus nigra*). The pathogen is widely distributed and can be especially problematic in urban areas or where stressed pine trees are found (i.e. disease symptoms become more pronounced with drought stress) (72). Wallis et al. (73) identified associations between Austrian pine susceptibility to *D. sapinea* and metabolites, including lignin (negatively correlated with susceptibility) and pinosylvin and pinosylvin monomethyl ether (positively correlated) using correlation analysis. Using PLSR, SIMCA, and SVM, constitutive metabolite profiles from the mid-infrared spectrum were used to distinguish between trees that differed in disease susceptibility and to predict differences in pine susceptibility to *D. sapinea* (36).

Mountain pine beetle (*Dendroctonus ponderosae*) is native to western North America. Its range and devastation, particularly of lodgepole pine (*Pinus contorta*), has increased due to climate change (74). Six et al. (33) examined differences in terpenoids among two groups of whitebark pine, *Pinus albicaulis*: survivors (trees that showed no signs of beetle attack) and generals (trees that were mature but below the size threshold for beetle attack). Permutational multivariate analyses of variance (PERMANOVA) and PCA were used to examine differences in terpenoids between groups and across sites. Concentrations of terpenoids differed between generals and survivors and were mostly higher in survivors. A similar approach of PCA and PERMANOVA was used to examine the impacts of mountain pine beetle fungal symbionts on lodgepole pine metabolites (34). Potential biomarkers of resistance were identified using multivariate empirical Bayes (MEBA) time-series analysis and included β-phellandrene, β-myrcene, β-pinene, and 3-carene.

5.5 Conclusions and Future Perspectives

As forest trees continue to be threatened by new and existing pests and pathogens, alone and in combination with climate change and other stressors, new solutions for disease and pest management will be needed. One solution

is to use metabolites associated with tree biotic stress response as biomarkers for pathogen and pest infection and/or for differentiating between resistant and susceptible trees. The workflow presented here has been successfully tested in multiple tree-pathogen and pest systems to distinguish between resistant and susceptible trees and for disease detection. However, additional approaches, beyond those highlighted in this chapter, will likely need to be considered and evaluated on a case-by-case basis. Finally, translating results and metabolite-based predictive models from controlled studies to real-world scenarios, either in the forest or in managed environments (e.g. nurseries and tree plantations) remains a challenge. However, advances in digital forestry and high-throughput phenotyping will likely facilitate translation and application at larger spatial scales.

Acknowledgments

The authors thank Carrie Fearer for providing helpful comments on an earlier draft of this chapter. The authors declare no conflicts of interest or competing interests. The use of trade names is for the information and convenience of the reader and does not imply official endorsement or approval by the USDA or the Forest Service of any product to the exclusion of others that may be suitable. This work was supported in part by the US Department of Agriculture, Forest Service.

References

1 Ramsfield TD, Bentz BJ, Faccoli M, Jactel H, Brockerhoff EG. Forest health in a changing world: effects of globalization and climate change on forest insect and pathogen impacts. *Forestry*. 2016; 89: 245–252.
2 Simler-Williamson AB, Rizzo DM, Cobb RC. Interacting effects of global change on forest pest and pathogen dynamics. *Annu Rev Ecol Evol Syst*. 2019; 50: 381–403.
3 Dukes JS, Pontius J, Orwig D, Garnas JR, Rodger VL, Brazee N, et al. Responses of insect pests, pathogens, and invasive plant species to climate change in the forests of northeastern North America: what can we predict? *Can J For Res*. 2009; 39: 231–248.
4 Aukema JE, McCullough DG, Von Holle B, Liebhold AM, Britton K, Frankel SJ. Historical accumulation of nonindigenous forest pests in the continental United States. *Bioscience*. 2010; 60: 886–897.
5 Lovett GM, Weiss M, Liebhold AM, Holmes TP, Leung B, Lambert KF, et al. Nonnative forest insects and pathogens in the United States: impacts and policy options. *Ecol Appl*. 2016; 26: 1437–1455.

6 USDA APHIS. Emerald ash borer beetle [Internet]. Available from (https://www.aphis.usda.gov/aphis/resources/pests-diseases/hungry-pests/the-threat/emerald-ash-borer/emerald-ash-borer-beetle), accessed 30 March 2023.
7 Donovan GH, Butry DT, Michael YL, Prestemon JP, Liebhold AM, Gatziolis D, et al. The relationship between trees and human health: evidence from the spread of the emerald ash borer. *Am J Prev Med*. 2013; 44: 139–145.
8 Li X, Holmes TP, Boyle KJ, Crocker EV, Nelson CD. Hedonic analysis of forest pest invasion: the case of emerald ash borer. *Forests*. 2019; 10: 1–15.
9 Kovacs KF, Haight RG, McCullough DG, Mercader RJ, Siegert NW, Liebhold AM. Cost of potential emerald ash borer damage in U.S. communities, 2009–2019. *Ecol Econ*. 2010; 69: 569–578.
10 Kovacs K, Holmes TP, Englin JE, Alexander J. The dynamic response of housing values to a forest invasive disease: evidence from a sudden oak death infestation. *Environ Resour Econ*. 2011; 49: 445–471.
11 Fei S, Morin RS, Oswalt CM, Liebhold AM. Biomass losses resulting from insect and disease invasions in US forests. *Proc Natl Acad Sci USA*. 2019; 116: 17371–17376.
12 Fiehn O. Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comp Funct Genomics*. 2001; 2: 155–168.
13 Fiehn O. Metabolomics – the link between genotypes and phenotypes. *Plant Mol Biol*. 2002; 48: 155–171.
14 Salminen JP, Roslin T, Karonen M, Sinkkonen J, Pihlaja K, Pulkkinen P. Seasonal variation in the content of hydrolyzable tannins, flavonoid glycosides, and proanthocyanidins in oak leaves. *J Chem Ecol*. 2004; 30: 1693–1711.
15 Yarnes CT, Boecklen WJ, Salminen J-P. No simple sum: seasonal variation in tannin phenotypes and leaf-miners in hybrid oaks. *Chemoecology*. 2008; 18: 39–51.
16 Lichtenthaler HK, Ač A, Marek MV, Kalina J, Urban O. Differences in pigment composition, photosynthetic rates and chlorophyll fluorescence images of sun and shade leaves of four tree species. *Plant Physiol Biochem*. 2007; 45: 577–588.
17 Conrad AO. Metabolomics of *Quercus* spp. to understand and predict resistance to *Phytophthora ramorum*. PhD Thesis. Ohio State University; 2015. http://rave.ohiolink.edu/etdc/view?acc_num=osu1429716653
18 Buche G, Colas C, Fougère L, Giordanengo T, Destandau E. Untargeted UHPLC-Q-TOF-HRMS based determination of discrimating compounds for oak species *Quercus robur* L. and *Quercus petraea* Liebl. identification. *Phytochem Anal*. 2021; 32: 660–671.

19 Bonello P, Blodgett JT. *Pinus nigra-Sphaeropsis sapinea* as a model pathosystem to investigate local and systemic effects of fungal infection of pines. *Physiol Mol Plant Pathol*. 2003; 63: 249–261.

20 Di Masi S, De Benedetto GE, Malitesta C, Saponari M, Citti C, Cannazza G, Ciccarella G. HPLC-MS/MS method applied to an untargeted metabolomics approach for the diagnosis of "olive quick decline syndrome". *Anal Bioanal Chem*. 2022; 414: 465–473.

21 Appel HM. Phenolics in ecological interactions – the importance of oxidation. *J Chem Ecol*. 1993; 19: 1521–1552.

22 Witzell J, Martin JA. Phenolic metabolites in the resistance of northern forest trees to pathogens – past experiences and future prospects. *Can J For Res*. 2008; 38: 2711–2727.

23 Sedio BE, Boya PCA, Rojas Echeverri JC. A protocol for high-throughput, untargeted forest community metabolomics using mass spectrometry molecular networks. *APPS*. 2018; 6(3): e1033.

24 Moritz T, Johansson AI. Plant metabolomics. In: Griffiths WJ, editor. *Metabolomics, Metabonomics and Metabolite Profiling*. Cambridge (UK): Royal Society of Chemistry; 2008. p. 254–272.

25 Lundborg L, Sampedro L, Borg-Karlson AK, Zas R. Effects of methyl jasmonate on the concentration of volatile terpenes in tissues of maritime pine and Monterey pine and its relation to pine weevil feeding. *Trees-Struct Funct*. 2019; 33: 53–62.

26 Chakraborty S, Whitehill JGA, Hill AL, Opiyo SO, Cipollini D, Herms DA, et al. Effects of water availability on emerald ash borer larval performance and phloem phenolics of Manchurian and black ash. *Plant Cell Environ*. 2014; 37: 1009–1021.

27 Paul A, de Boves Harrington P. Chemometric applications in metabolomic studies using chromatography-mass spectrometry. *Trends Anal Chem*. 2021; 135: 116165.

28 Wehrens R. *Chemometrics with R: Multivariate Data Analysis in the Natural Sciences and Life Sciences*. Heidelberg, New York: Springer; 2011.

29 Sperschneider J. Tansley insight machine learning in plant–pathogen interactions: empowering biological predictions from field scale to genome scale. *New Phytol*. 2020; 228: 35–41.

30 Rohart F, Gautier B, Singh A, Lê Cao KA. mixOmics: an R package for 'omics feature selection and multiple data integration. *PLoS Comput Biol*. 2017; 13: 1–19.

31 Kersten B, Ghirardo A, Schnitzler J-P, Kanawati B, Schmitt-Kopplin P, Fladung M, et al. Integrated transcriptomics and metabolomics decipher differences in the resistance of pedunculate oak to the herbivore *Tortrix viridana* L. *BMC Genomics*. 2013; 14: 737.

32 Bertić M, Schroeder H, Kersten B, Fladung M, Orgel F, Buegger F, et al. European oak chemical diversity – from ecotypes to herbivore resistance. *New Phytol*. 2021; 232: 818–834.

33 Six DL, Trowbridge A, Howe M, Perkins D, Berglund E, Brown P, et al. Growth, chemistry, and genetic profiles of whitebark pine forests affected by climate-driven mountain pine beetle outbreaks. *Front For Global Change*. 2021; 4: 1–22.

34 Zaman R, May C, Ullah A, Erbilgin N. Bark beetles utilize ophiostomatoid fungi to circumvent host tree defenses. *Meta*. 2023; 13: 239.

35 Fearer CJ, Conrad AO, Marra RE, Georskey C, Villari C, Slot J, et al. A combined approach for early in-field detection of beech leaf disease using near-infrared spectroscopy and machine learning. *Front For Global Change*. 2022; 5: 1–16.

36 Conrad AO, Villari C, Sherwood P, Bonello P. Phenotyping Austrian pine for resistance using Fourier-transform infrared spectroscopy. *AUF*. 2020; 46: 276–286.

37 Sambles CM, Salmon DL, Florance H, Howard TP, Smirnoff N, Nielsen LR, et al. Data descriptor: ash leaf metabolomes reveal differences between trees tolerant and susceptible to ash dieback disease. *Sci Data*. 2017; 4: 1–13.

38 Conrad AO, McPherson BA, Wood DL, Madden LV, Bonello P. Constitutive phenolic biomarkers identify naïve *Quercus agrifolia* resistant to *Phytophthora ramorum*, the causal agent of sudden oak death. *Tree Physiol*. 2017; 37: 1686–1696.

39 McPherson BA, Mori SR, Opiyo SO, Conrad AO, Wood DL, Bonello P. Association between resistance to an introduced invasive pathogen and phenolic compounds that may serve as biomarkers in native oaks. *For Ecol Manag*. 2014; 312: 154–160.

40 Nemesio-Gorriz M, Menezes RC, Paetz C, Hammerbacher A, Steenackers M, Schamp K, et al. Canditate metabolites for ash dieback tolerance in *Fraxinus excelsior*. *J Exp Bot*. 2020; 71: 6074–6083.

41 Conrad AO, Rodriguez-Saona LE, McPherson BA, Wood DL, Bonello P. Identification of *Quercus agrifolia* (coast live oak) resistant to the invasive pathogen *Phytophthora ramorum* in native stands using Fourier-transform infrared (FT-IR) spectroscopy. *Front Plant Sci*. 2014; 5: 521.

42 Villari C, Dowkiw A, Enderle R, Ghasemkhani M, Kirisits T, Kjær ED, et al. Advanced spectroscopy-based phenotyping offers a potential solution to the ash dieback epidemic. *Sci Rep*. 2018; 8: 17448.

43 Benton HP, Ivanisevic J, Mahieu NG, Kurczy ME, Johnson CH, Franco L, et al. Autonomous metabolomics for rapid metabolite identification in global profiling. *Anal Chem*. 2015; 87: 884–891.

44 Febrero-Bande M, de la Fuente M. Statistical computing in functional data analysis: the R package fda.usc. *J Stat Softw*. 2012; 51: 1–28.

45 Singh A, Ganapathysubramanian B, Singh AK, Sarkar S. Machine learning for high-throughput stress phenotyping in plants. *Trends Plant Sci*. 2016; 21: 110–124.

46 Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. *Science*. 2015; 349: 255–260.

47 Liakos KG, Busato P, Moshou D, Pearson S, Bochtis D. Machine learning in agriculture: a review. *Sensors (Switzerland)*. 2018; 18: 1–29.

48 Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. *PLoS One*. 2019; 14: 1–20.

49 Lê Cao K-A, Boitard S, Besse P. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinf*. 2011; 12: 253.

50 Genuer R, Poggi J-M, Tuleau-Malot C. VSURF: an R package for variable selection using random forests. *R J*. 2015; 7: 19–33.

51 Stock NL, Doran MC, Bonners RF, March RE. Liquid chromatography/mass spectrometry for the detection of ash tree metabolites following emerald ash borer infestation. *Rapid Commun Mass Spectrom*. 2018; 32: 385–392.

52 Xia J, Psychogios N, Young N, Wishart DS. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Res*. 2009; 37: W652–W660.

53 Xia J, Sinelnikov IV, Han B, Wishart DS. MetaboAnalyst 3.0—making metabolomics more meaningful. *Nucleic Acids Res*. 2015; 43: W251–W257.

54 Kucheryavskiy S. mdatools – R package for chemometrics. *Chemom Intell Lab Syst*. 2020; 198: 103937.

55 Conrad AO, Li W, Lee D-Y, Wang G-L, Rodriguez-Saona L, Bonello P. Machine learning-based presymptomatic detection of rice sheath blight using spectral profiles. *Plant Phenomics*. 2020; 2020: 1–10.

56 Mukrimin M, Conrad AO, Kovalchuk A, Julkunen-Tiitto R, Bonello P, Asiegbu FO. Fourier-transform infrared (FT-IR) spectroscopy analysis discriminates asymptomatic and symptomatic Norway spruce trees. *Plant Sci*. 2019; 289: 110247.

57 Naidoo S, Slippers B, Plett JM, Coles D, Oates CN. The road to resistance in forest trees. *Front Plant Sci*. 2019; 10: 273.

58 Sniezko RA, Koch J. Breeding trees resistant to insects and diseases: putting theory into application. *Biol Invasions*. 2017; 19: 3377–3400.

59 Hall RD, D'Auria JC, Silva Ferreira AC, Gibon Y, Kruszka D, Mishra P, et al. High-throughput plant phenotyping: a role for metabolomics? *Trends Plant Sci*. 2022; 27: 549–563.

60 Chlus A, Townsend PA. Characterizing seasonal variation in foliar biochemistry with airborne imaging spectroscopy. *Remote Sens Environ*. 2022; 275: 113023.

61 Fallon B, Yang A, Lapadat C, Armour I, Juzwik J, Montgomery RA, et al. Spectral differentiation of oak wilt from foliar fungal disease and drought is correlated with physiological changes. *Tree Physiol*. 2020; 40: 377–390.

62 Sapes G, Lapadat C, Schweiger AK, Juzwik J, Montgomery R, Gholizadeh H, et al. Remote sensing of environment canopy spectral reflectance detects oak wilt at the landscape scale using phylogenetic discrimination. *Remote Sens Environ*. 2022; 273: 112961.

63 Enderle R, Stenlid J, Vasaitis R. An overview of ash (*Fraxinus* spp.) and the ash dieback disease in Europe. *CAB Rev Perspect Agric Vet Sci Nutr Nat Resour*. 2019; 14: 025.

64 Cleary MR, Andersson PF, Broberg A, Elfstrand M, Daniel G, Stenlid J. Genotypes of *Fraxinus excelsior* with different susceptibility to the ash dieback pathogen *Hymenoscyphus pseudoalbidus* and their response to the phytotoxin viridiol – a metabolomic and microscopic study. *Phytochemistry*. 2014; 102: 115–125.

65 Villari C, Herms DA, Whitehill JGA, Cipollini D, Bonello P. Progress and gaps in understanding mechanisms of ash tree resistance to emerald ash borer, a model for wood-boring insects that kill angiosperms. *New Phytol*. 2016; 209: 63–79.

66 Koch JL, Carey DW, Mason ME, Poland TM, Knight KS. Intraspecific variation in *Fraxinus pennsylvanica* responses to emerald ash borer (*Agrilus planipennis*). *New For*. 2015; 46: 995–1011.

67 Qazi SS, Lombardo DA, Abou-Zaid MM. A metabolomic and HPLC-MS/MS analysis of the foliar phenolics, flavonoids and coumarins of the *Fraxinus* species resistant and susceptible to emerald ash borer. *Molecules*. 2018; 23: 1–19.

68 Rizzo DM, Garbelotto M, Davidson JM, Slaughter GW, Koike ST. *Phytophthora ramorum* as the cause of extensive mortality of *Quercus* spp. and *Lithocarpus densiflorus* in California. *Plant Dis*. 2002; 86: 205–214.

69 Grunwald NJ, LeBoldus JM, Hamelin RC. Ecology and evolution of the sudden oak death pathogen *Phytophthora ramorum*. *Annu Rev Phytopathol*. 2019; 57: 301–321.

70 Ockels FS, Eyles A, McPherson BA, Wood DL, Bonello P. Phenolic chemistry of coast live oak response to *Phytophthora ramorum* infection. *J Chem Ecol*. 2007; 33: 1721–1732.

71 Nagle AM, McPherson BA, Wood DL, Garbelotto M, Bonello P. Relationship between field resistance to *Phytophthora ramorum* and constitutive phenolic chemistry of coast live oak. *For Pathol*. 2011; 41: 464–469.

72 Sherwood P, Villari C, Capretti P, Bonello P. Mechanisms of induced susceptibility to *Diplodia* tip blight in drought-stressed Austrian pine. *Tree Physiol*. 2015; 35: 549–562.

73 Wallis C, Eyles A, Chorbadjian RA, Riedl K, Schwartz S, Hansen R, et al. Differential effects of nutrient availability on the secondary metabolism of Austrian pine (*Pinus nigra*) phloem and resistance to *Diplodia pinea*. *For Pathol*. 2011; 41: 52–58.

74 Cudmore TJ, Björklund N, Carroll AL, Staffan Lindgren B. Climate change and range expansion of an aggressive bark beetle: evidence of higher beetle reproduction in naïve host tree populations. *J Appl Ecol*. 2010; 47: 1036–1043.