Computational and Structural Biotechnology Journal 27 (2025) 1741–1753

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Integrating spatial mapping and metabolomics: A novel platform for bioactive compound discovery and saline land reclamation

Tushar Andriyas [a,b], Nisa Leksungnoen [c], Pichaya Pongchaidacha [a,b], Arashaporn Uthairangsee [a,b], Suwimon Uthairatsamee [c], Peerapat Doomnil [d], Yongkriat Ku-Or [c], Chatchai Ngernsaengsaruay [e], Sanyogita Andriyas [f], Arerut Yarnvudhi [c], Rossarin Tansawat [a,b,g,*,1]

[a] Department of Food and Pharmaceutical Chemistry, Faculty of Pharmaceutical Science, Chulalongkorn University, Bangkok, Thailand
[b] Center of Excellence in Metabolomics for Life Sciences, Chulalongkorn University, Bangkok, Thailand
[c] Department of Forest Biology, Faculty of Forestry, Kasetsart University, Bangkok, Thailand
[d] Forest Resource Management Office No.4 (Tak), Pa Mamuang sub-district, Muang Tak, Tak 63000, Thailand
[e] Department of Botany, Faculty of Science, Kasetsart University, Bangkok, Thailand
[f] Department of Irrigation and Drainage Engineering, Vaugh Institute of Agriculture Engineering and Technology, Sam Higginbottom University of Agriculture, Technology, and Sciences, Prayagraj, Uttar Pradesh, India
[g] Thailand Metabolomics Association, Bangkok, Thailand

ARTICLE INFO

Keywords:
Species Distribution Model
Saline land
Metabolomics
Buchanania siamensis
Secondary metabolites
Bioactive

ABSTRACT

Saline lands pose significant environmental and agricultural challenges due to high soil salinity, which disrupts water uptake and ionic balances, limiting conventional crop productivity. Yet, certain endemic plants thrive under these conditions and may offer untapped bioactive compounds. This study proposes a novel platform that integrates species distribution modeling (SDM) and advanced metabolomics to screen for bioactive secondary metabolites, using *Buchanania siamensis*, a rare native species, as a case study. An ensemble SDM model incorporating environmental and soil parameters identified salinity as a critical factor influencing the species' distribution. Leaf samples were collected from naturally growing trees at both saline (SS) and non-saline (NS) sites. LC-QTOF metabolomic analysis annotated a total of 1106 metabolites across the leaf samples, with 175 found to be significantly different between the groups. Among them, 108 metabolites exhibited higher abundance in the SS group. Additionally, antioxidant assays including DPPH, FRAP, and total phenolic content tests, were conducted. Data were further analyzed using O-PLSR models to identify key metabolites most relevant to antioxidant properties. The results indicated that afzelin was the key metabolite responsible for the antioxidant properties of *B. siamensis*, with significantly higher levels in SS compared to NS samples ($p < 0.05$), as determined by peak area. By leveraging this multidisciplinary approach, we propose a framework to support both bioactive

Abbreviations: ANOVA, Analysis of variance; AUC, Area under the curve; *B. siamensis,*, *Buchanania siamensis*; BD, Bulk density; BIEN, Botanical information and ecology network; COR, Correlation coefficient; DPPH, 2,2-Diphenyl-1-picrylhydrazyl; EC, Electrical conductivity; ET, Evapotranspiration; FC, Folin-ciocalteu; FDR, False discovery rate; FRAP, Ferric reducing antioxidant power assay; GAE, Gallic acid equivalent; GAM, Generalized additive model; GBIF, Global biodiversity information facility; GLM, Generalized linear model; ISRIC, International soil reference and information centre; LC-MS,, Liquid chromatography–mass spectrometry; LC-QTOF, Liquid chromatography-quadrupole time-of-flight; MaxEnt, Maximum entropy; NS, Non-saline samples; O-PLSR, Orthogonal partial least squares regression; OK, Ordinary kriging; OPLS-DA, Orthogonal partial least squares-discriminant analysis; PDSI, Palmer drought severity index; PTFE, Polytetrafluoroethylene; QC, Quality control; QTOF, Quadrupole time-of-flight; RF, Random forest; RK, Regression-kriging; RMSE, Root-mean-square error; ROS, Reactive oxygen species; SDM, Species distribution modeling; SM, Soil moisture; SS, Saline samples; SVM, Support vector machine; Tmax, Maximum temperature; Tmin, Minimum temperature; TPC, Total phenolic content; TSS, True skill statistic; UPLC, Ultra-performance liquid chromatography; VIF, Variance inflation factor; VIP, Variable importance in projection; VPD, Vapor pressure deficit.

* Correspondence to: Department of Food and Pharmaceutical Chemistry, Faculty of Pharmaceutical Sciences, Chulalongkorn University, 254 Phayathai Rd., Pathumwan, Bangkok 10330, Thailand.
E-mail address: rossarin.t@pharm.chula.ac.th (R. Tansawat).
[1] ORCID iD: 0000–0001-5486–0231

https://doi.org/10.1016/j.csbj.2025.04.035
Received 30 March 2025; Received in revised form 23 April 2025; Accepted 24 April 2025
Available online 30 April 2025

compound discovery and saline land reclamation, offering potential environmental and pharmaceutical benefits. This integrated platform may support pharmaceutical research, particularly in drug discovery efforts.

1. Introduction

Global salt-affected areas are estimated to be nearly around a billion hectares [1], with a rapid expansion predicted in the coming decades [2]. Soil salinity reduces the global soil organic carbon stocks by hindering microbial activity and plant growth, leading to higher carbon emission [3,4]. Thailand has experienced similar expansion in salt-affected areas due to extensive deforestation [5] or long-term salt mining in the past, resulting in the surface up-welling of salt in parent rock material [6]. Specifically, the salinity levels in northeastern Thailand have been measured as high as 40–60 dS/m [6], under which most species would be unable to survive.

High salinity levels affect plant survival and soil viability by creating imbalances in water and nutrient availability, lowering plant root and leaf biomass, photosynthesis rate, and reducing root length [7]. This is done through an overproduction of reactive oxygen species (ROS), leading to oxidative damage of cellular macromolecules such as proteins, lipids, and DNA [8]. To counteract such stresses, plants utilize a non-enzymatic ROS detoxification system which utilizes carotenoids, phenolic compounds, and flavonoids [8], tending to modulate the levels of these bioactive compounds or secondary metabolites [9]. This can increase the accumulation of secondary metabolites such as alkaloids [10], total phenolics, flavonoids, carotenoid contents, and antioxidant activity in medicinal plants [11,12], that play an important role in enhancing plant defense mechanisms against oxidative stress [13]. An advanced approach is needed to effectively screen and predict the distribution of plant species in saline environments, particularly those with potential bioactive properties.

Species distribution modeling (SDM) is a spatial technique that utilizes environmental data and species occurrence records to predict the geographic range of species. SDMs are used to explore the relationship between species occurrences and environmental and soil variables. The modeling has been widely used in biodiversity assessments, conservation efforts, evolutionary biology, epidemiology, biology of global change, and wildlife management [14]. Integrated frameworks like openModeller [15], BIOENSEMBLES [16], and ModeEco [16] offer multiple SDM fitting routines. By integrating SDM, researchers can identify areas where endemic plants adapted to saline conditions are likely to thrive, providing a targeted and efficient method for exploring biodiversity in challenging environments.

Buchanania siamensis Miq. is a medium-to-large tree and belongs to the Anacardiaceae family (mango), which consists of a diverse group of flowering plants, with an estimated 82 genera and over 700 species globally [17]. *B. siamensis* is a key yet lesser-known species from this genera, sharing its genetic lineage with cashew nut (*Anacardium occidentale L.*) [18]. The conservation status of*s*, according to the International Union for Conservation of Nature (IUCN), is unknown or not readily available, with databases such as GBIF [19] indicating that this species is endemic to Southeast Asia [20]. It is a rare species, native to northeastern, central, eastern, and western Thailand, particularly in deciduous forests. Its rarity can be attributed to habitat loss and degradation from deforestation, land conversion for agriculture, and urban development, which have gradually contributed to the reduction of its natural habitat [21]. The species is frequently found in open deciduous forests in alluvial plains at altitudes between 50 and 200 m [21] and can grow in well-drained sandy loam or loamy soil types in slightly acidic to neutral environments.

The species has several ethnobotanical uses, as indicated by past studies related to the biological activity of *B. siamensis*. Its young leaves are considered edible, having a sour taste and are a food source for humans [22], while the fruits are consumed by birds and seeds propagated far from the mother plant. Locals use various parts of the tree in traditional medicine to treat ailments like leucorrhoea [23], food poisoning, fever, gingivitis, and other infections. The presence of various bioactive compounds [24] and in cosmeceutical products [20] have been reported in *B. siamensis*. A high accumulation of phenolic content and antioxidants in the stem and leaf extracts indicate its potential medicinal usage in patients suffering from diseases such as cancer [25].

Even though phenolics, flavonoids and antioxidant activity have been previously reported in the leaves of *B. siamensis* [20], its comprehensive metabolomic profiling under the influence of salinity remains unexplored. The species has also been identified as a potential candidate for afforestation initiatives in salt-affected landscapes [21], making it a candidate to the study the effects of soil salinity on its secondary metabolite composition. Hence, utilizing native medicinal plants like *B. siamensis* for bioremediation of degraded habitats [26] would not only improve vegetation cover but also support conservation efforts by promoting the survival and propagation of these regionally significant species.

Metabolomics is an indispensable tool in plant science research [27,28]. Through metabolomics, a comprehensive profiling of secondary metabolites produced and modulated as an adaptive response to abiotic stressors, such as salinity, can be analyzed. Additionally, it has been used to identify metabolite biomarkers that are largely independent of genetic and environmental variations [29].Combining SDM with high-throughput metabolomics enables the identification and characterization of small molecules produced by these plants, particularly metabolites with potential bioactive properties. This integrative approach bridges the gap between ecological modeling and chemical profiling, offering a powerful platform for both bioactive discovery and saline land reclamation.

This study aimed to add value to saline land by exploring the potential pharmaceutical applications of plants that can be grown in these areas. We therefore present a platform that combines species distribution modeling (SDM) and untargeted liquid chromatography–mass spectrometry (LC-MS) metabolomic analysis to identify potential planting sites in degraded lands, as well as screening of bioactive compounds from rare plant species. By incorporating high-technology platforms, particularly spatial mapping, this approach can assist in the bioactive discovery process and enhance the efficiency of compound identification. Additionally, untargeted metabolomics enables faster and more comprehensive screening of bioactive compounds compared to traditional methods, allowing the identification of a wide range of compounds within a shorter time frame and thereby accelerating the drug discovery process. The integration of these technologies not only supports land reclamation and pharmaceutical development but also contributes to the economic growth of our country. To the best of our knowledge, this is the first report dealing with the impacts of salinity on the foliar bioactive compounds and antioxidant activities of *B. siamensis* growing naturally on degraded lands. This approach may support the discovery of valuable secondary metabolites, while also contributing to conservation efforts of this species and providing potential economic benefits to local communities. The case of *B. siamensis* demonstrates how environmental factors like salinity can be leveraged to enhance the production of specific bioactive compounds, a model that can be extended to other such rare species thriving under abiotic stress.

2. Materials and methods

2.1. Spatial mapping

2.1.1. Spatial distribution modeling

The spatial distribution of *B. siamensis* was determined using occurrence data from several public databases, including the Global Biodiversity Information Facility (GBIF) and the Botanical Information and Ecology Network (BIEN) [30,31], supplemented with prior field observations in the northeastern part of Thailand. The rarity of this species' occurrence was evident through predominant occurrences concentrated in Southeast Asia, reflecting its ecological preference for tropical and subtropical climates. Specifically, a total of 136 unique occurrences (seen as black dots in Fig. 1) within Thailand were determined through the databases and *in situ* verifications.

2.1.2. Soil and climate data

The soil rasters used to predict the soil salinity levels in Thailand were obtained from the SoilGrids dataset provided by the International Soil Reference and Information Centre or ISRIC [32], located at https://files.isric.org/soilgrids/latest/data/. This dataset includes several soil attributes at different depths such as soil organic carbon content, pH, sand, silt, and clay fractions, bulk density (BD), and cation-exchange capacity. Raster grids for soil properties that included BD, pH, sand, silt, and clay percentages, and nitrogen content were downloaded at depths of 30–60 cms (i.e., at the root-zone of the species). The data were downloaded, processed, and analyzed using R statistical software [33]. Elevation raster data for Thailand (obtained using the *raster* package in R) was also integrated into the prediction model alongside soil attributes to understand its influence on salinity distribution.

The climate data rasters used in the SDM were downloaded from the TerraClimate global climate dataset [34] located at https://climate.northwestknowledge.net/TERRACLIMATE. TerraClimate provides a global dataset with a spatial resolution of 0.040 degrees and a monthly temporal resolution, covering the period from 1958 to 2020. Spatial rasters of several primary and secondary climate variables were downloaded, such as maximum temperature (Tmax), minimum temperature (Tmin), vapor pressure deficit (VPD), wind speed, downward solar radiation, volumetric soil moisture (SM), total rainfall, evapotranspiration (ET), and palmer drought severity index (PDSI, which indicates drought severity based on precipitation and temperature data). The data were accessed from the public domain and extracted and analyzed using the R statistical software [33].

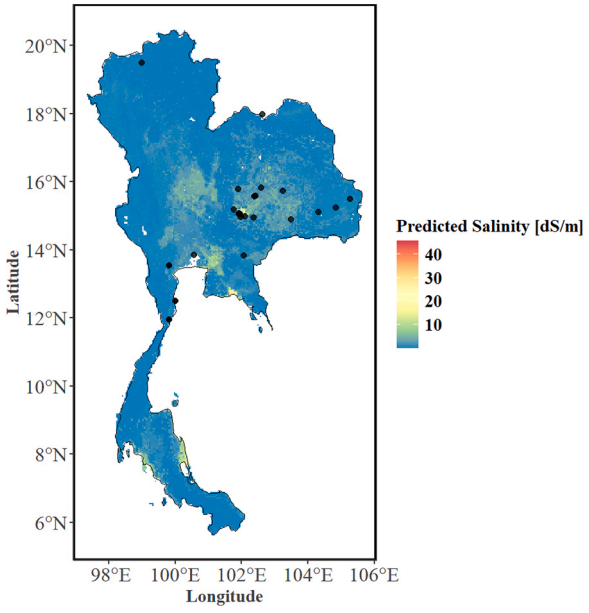

Fig. 1. Distribution of *B. siamensis* in various parts of Thailand where the species is natively found. The locations (indicated by black dots) are overlaid on the salinity distribution as predicted by the best model (random forest). The locations were obtained either from the GBIF database or were identified individually using a local survey.

2.1.3. Prediction of salinity raster

The stacked raster grids of elevation, BD, pH, sand, silt, and clay percentages, and nitrogen content were used to predict the variations in salinity around Thailand. The salinity values were spatially predicted using various regression-kriging (RK) models using rasters of both soil and auxiliary variables (elevation). Regression algorithms that included Ordinary Kriging (OK), Generalized Linear Model (GLM), Generalized Additive Model (GAM), Random Forest (RF) [35], and their combinations with Ordinary Kriging of residuals (GLM + OK, GAM + OK, RF + OK), were compared through a k-fold cross-validation using a 70:30 train-test split. The best fit model was chosen to predict the salinity raster based on the performance estimated using Root-Mean-Square Error (RMSE).

2.1.4. Species distribution model

The *sdm* package [36] was used to model the species distribution, given its diverse modeling approaches, model ensembles, and comprehensive evaluation, with an object-oriented design for error handling and a user-friendly GUI in R [33]. The package contains several methods to predict the spatial distribution of species, that include generalized additive model (GAM), generalized linear model (GLM), Random Forest (RF), Support Vector Machine (SVM), Maximum Entropy (MaxEnt) algorithms, among others. The climate data rasters, in addition to the salinity raster that was constructed using the soil variables, were used to generate a distribution map of the most probable locations of *B. siamensis* around Thailand. An ensemble prediction was then used to enhance the prediction accuracy by combining the strengths of the selected algorithms, indicating species' habitat suitability and distribution patterns [37]. The performance of SDMs was quantified using several metrics that included area under the curve (AUC), correlation coefficient (COR), true skill statistic (TSS), and deviance.

2.1.5. Determination of salinity levels to characterize leaf metabolome

Soil samples were collected from naturally growing trees in saline and non-saline locations around the Kam Tale So district, Nakhon Ratchasima province, Thailand, during February 2023. This region is the native habitat of this species and experiences severe salinity stress, making it ideal for this study. Soil samples were collected from the root zone (at a depth of 0–30 cm) of trees concurrently with the leaf collection for the estimation of site-specific salinity levels. Approximately 500 g of soil was placed in labeled plastic bags to prevent contamination and moisture loss during transport to the laboratory. Subsequently, the soil samples were air-dried at room temperature to a constant weight at Kasetsart University. The dried soil was then sieved through a 2 mm mesh and salinity was measured using the soil extract (5:1 ratio, $EC_{1:5}$). The saturation pasted electrical conductivity (ECe) was then calculated using the regression equation reported by Leksungnoen [6] using an EC meter (Model Edge HI 2030 with Electrode HI 763100, Hanna Instruments Inc., Woonsocket, RI, USA).

These values were used to characterize the variations in leaf metabolome of *B. siamensis* as a function of salinity. Salinity values between 0 and 2 dS/m were categorized as non-saline (NS for Non-saline Samples), while any values above this range were deemed as saline (SS for Saline Samples). Subsequently, the qualitative metabolomic profile of the samples was determined using LC-MS and multivariate data analysis was used to assess the effect of soil salinity on the leaf metabolome.

2.2. Metabolomics analysis

2.2.1. Leaf samples

Leaves of *B. siamensis* were collected from naturally growing trees in saline (n = 17) and non-saline locations (n = 13). The leaf samples from each tree were stored in plastic bags in an ice box and sent to Chulalongkorn University, Faculty of Pharmacy, Bangkok, Thailand within 24 h of collection for further processing.

The extraction was performed using the method outlined in Veeramohan et al. [38]. Briefly, fresh leaves of *B. siamensis* were ground in liquid nitrogen with a mortar and pestle. Approximately 100 mg of the powder was placed in a 15-mL falcon tube, soaked in 5-mL of ice-cold methanol, vortexed and incubated overnight in wrist-action shaker (Rotamax 120, Heidolph, Schwabach, Germany) at 20°C. The extraction mixture was centrifuged at 4°C and 6000 rpm for 10 min (Centrifuge 5810 R, Eppendorf, Hamburg, Germany). The supernatant was filtered through a 0.22 μm polytetrafluoroethylene (PTFE) syringe filter prior to analysis via liquid chromatography–mass spectrometry (LC-MS). To prevent degradation, the samples were stored at −80°C until the analysis.

2.2.2. LC-MS analysis

A Bruker Ultra-Performance Liquid Chromatography (UPLC) system coupled with a Quadrupole Time-of-Flight (QTOF) compact mass spectrometry (Bruker, Bremen, German) was used for the qualitative analysis of *B. siamensis*. Sample separation was performed on an Intensity Solo C18 column (Bruker, 100 mm × 2.1 mm, 1.8 μm; Bruker, Bremen, German). Chromatographic separation was carried out using a gradient elution with 2.5 mM ammonium formate buffer in water (solvent A) and 2.4 mM ammonium formate buffer in methanol (solvent B), at a flow rate of 4 mL/min and an injection volume of 4 μL. The gradient condition initiated with 99.9 % solvent A (0.0–2.1 min), then gradually decreased to 75.0 % solvent A (2.1–10.0 min). It was further reduced to 20.0 % solvent A (10.0–12.0 min), followed by a continual decline to 10.0 % solvent A (12.0–21.0 min). The proportion was then reduced to 0.1 % of solvent A (21.0–23.0 min). Subsequently, the gradient was increased back to 99.9 % solvent A (23.0–24.0 min) and maintained at this proportion from 24.0 to 26.1 minutes. The mass spectrum analysis was conducted in positive ion mode using an electrospray ionization (ESI) source with the following parameters: spray voltage of 500 V, a capillary voltage of 4500 V, and a mass range of m/z 50–1000. Nitrogen gas was used as both the nebulizer and collision gas, with a drying gas flow rate of 4.0 L/min at 200°C. The collision energy ranged from 20 to 30 eV. Five technical replicates were analyzed in this study to ensure that the identified features originated from the analyzed samples rather than from artifacts resulting from low reproducibility [39]. Analytical performance was evaluated using a quality control (QC) sample prepared by pooling 100 μL of supernatant from each sample and running it between every 10 sample injections.

2.2.3. Metabolite identification

The raw chromatograms and mass spectra obtained from LC-MS were processed using the MS-DIAL software (version 5.4.241004) [40], to detect peak heights significantly above the noise floor. After preprocessing steps that included baseline correction, peak alignment, and normalization, the peak intensities were normalized across samples using statistical analysis. These were compared with databases using mass spectral data and retention times. Compounds were annotated if their mass spectra exhibited at least 70 % similarity to entries in the NIST database (https://webbook.nist.gov/chemistry/) and other publicly available mass spectrometry databases, including MassBank (http://www.massbank.jp), GNPS, MONA, and METLIN (http://metlin.scripps.edu/index.php).

2.3. Antioxidant activity

2.3.1. Chemicals

2,2-Diphenyl-1-picrylhydrazyl (DPPH), iron(III) chloride hexahydrate, iron(II) sulfate heptahydrate, 2,4,6-tri(2-pyridyl)-s-triazine (TPTZ), and gallic acid were sourced from Sigma Aldrich (St. Louis, MO, USA). Hydrochloric acid, acetic acid, methanol and Folin–Ciocalteu's reagent were obtained from Merck KGaA (Darmstadt, Germany). Sodium acetate and sodium carbonate were purchased from Qrec chemical (Auckland, New Zealand). All chemicals used in the experiment were of analytical grade.

2.3.2. Sample preparation

After grinding the leaves of *B. siamensis* to a fine powder, 0.1 gm was extracted with 5 mL cold methanol. The mixture was vortexed and incubated overnight in an incubator shaker (Rotamax 120, Heidolph instruments, Bavaria, Germany) at 20°C. Subsequently, the mixture was centrifuged (Centrifuge 5810 R, Eppendorf, Leipzig, Germany) at 4°C and 6000 rpm for 10 mins and the supernatant obtained was then used to determine the total phenolic content (TPC) and analyze for the antioxidant activity through DPPH and ferric reducing antioxidant power (FRAP) assays.

2.3.3. DPPH radical scavenging activity

The DPPH radical scavenging activity was determined based on the modified method of Jayasekera et al. [41]. For each leaf extract, 25 μL of extract was mixed with 250 μL of freshly prepared 0.2 mM DPPH in 95 % ethanol. The mixture was incubated in the dark at room temperature for 30 mins. Absorbance was measured at 550 nm using a CLARIOstar microplate reader (CLARIOstar, BMG labtech, Ortenberg, Germany), with ascorbic acid as the standard. The percentage of DPPH free radical inhibition was determined using the formula:

$$\text{Inhibition}(\%) = \frac{A\ control - A\ test}{A\ control} \text{x}\ 100$$

where "*A control*" represents the absorbance of the control and "*A test*" refers to the absorbance of the reaction mixture containing the sample. All tests were performed on five replicates, and the average values were calculated.

2.3.4. FRAP assay

The FRAP assay was performed using the modified method proposed by Jayasekera et al. [41]. The FRAP reagent was freshly prepared by mixing 300 mM acetate buffer (pH 3.6), 10 mM TPTZ in 40 mM HCl, and 20 mM $FeCl_3 \cdot 6H_2O$ solution in a ratio of 10:1:1. In each well of a 96-well plate, 8.5 μL of sample, 250 μL of reagent, and 25 μL of DI water were added. Following a 30 min incubation period, the absorbance was measured at a wavelength of 595 nm using a microplate reader (CLARIOstar, BMG labtech Ortenberg, Germany). All analyses were performed on five replicates and expressed in μM Fe^{2+} per g of dry extract weight.

2.3.5. Quantification of TPC

TPC was determined using the Folin-Ciocalteu (FC) method [42]. A 12.5 μL aliquot of each sample was mixed with 250 μL of 2 % sodium carbonate solution and left to stand for 5 mins at room temperature. Then, 12.5 μL of 50 % Folin-Ciocalteu reagent was added. The mixture was incubated in the dark at room temperature for 30 mins and the absorbance was measured at a wavelength of 650 nm using a CLARIOstar microplate reader (CLARIOstar, BMG labtech, Ortenberg, Germany). Measurements were performed on five replicates. Results were linearized using a calibration curve prepared with an aqueous gallic acid solution (10–100 μg/mL) and expressed as mg gallic acid equivalents (GAE) per g of dry extract weight.

2.4. Statistical analysis

The statistical analysis workflow of the foliar metabolome had the following steps to filter significant metabolites between saline and non-saline samples as well as metabolite(s) most relevant to the antioxidant properties. From the initial dataset of metabolic features post-MS-DIAL processing, the orthogonal partial least squares-discriminant analysis (OPLS-DA) modeling was used to determine the significantly discriminating metabolites (based on a Variable Importance in Projection (VIP) score > 1 and $p < 0.05$). Data were further analyzed by three orthogonal partial least squares regression (O-PLSR) models, independently constructed for the three antioxidant activity assays (DPPH, FRAP, and TPC).

3. Results

3.1. Salinity prediction model

Fig. 1 depicts the distribution of *B. siamensis* in various parts of Thailand, with species locations marked by black dots overlaid on a map of predicted soil salinity as obtained by the best spatial interpolation model (RF algorithm). The performance metrics of the various algorithms used to interpolate salinity data are listed in Table 1. The results indicate that OK had the lowest as well as inconsistent performance as estimated by the highest average RMSE (11.531 dS/m) and a high standard deviation (8.769 dS/m). In contrast, RF had the lowest average RMSE (1.732 dS/m) with a standard deviation of 2.389 dS/m, implying the best modeling robustness and accuracy. GAM performed better than GLM, while RF combined with Ordinary Kriging of residuals (RF OK) had a slight reduction in performance compared to RF alone. GLM combined with Ordinary Kriging of residuals (GLM OK) had a marginal improvement over a standalone GLM, while GAM combined with Ordinary Kriging of residuals (GAM OK) resulted in the second-best performance in terms of an average RMSE of 1.964 dS/m and a standard deviation of 2.449 dS/m. Overall, RF had the best overall performance and was used to generate the salinity raster (Fig. 1) for further use in the SDM prediction of the spatial distribution of *B. siamensis*.

3.2. Species distribution modeling

The salinity raster was combined with other environmental variables to generate the most probable spatial distribution for the species using SDM. Replication was incorporated through sub-sampling and bootstrapping with 30 % test samples per model and three replicates per model. The performance of SDMs using RF, MaxEnt, and SVM had the best goodness of fit metrics (listed in Table 2) amongst the available methods in the SDM package and were used to create an ensemble model. RF had the highest AUC, COR, TSS, and the lowest deviance, indicating excellent classification accuracy. Fig. 2 illustrates the predicted spatial distribution of *B. siamensis* using a weighted ensemble model of the RF, SVM, and MaxEnt algorithms.

The environmental conditions at various locations of species presence were determined using the Terraclimate and ISRIC databases. The variables of interest included Tmax, Tmin, VPD, wind speed, downward solar radiation, SM, total rainfall, ET, Palmer drought index, soil bulk density, soil pH, sand-silt-clay percentage, and nitrogen content. Collinear variables were excluded using *vif* (variance inflation factor), which reduced the set to wind speed, VPD, Tmax, salinity, rainfall, and PI, which were used to predict species presence using an ensemble average of RF, SVM, and MaxEnt models, as illustrated in Fig. 2. The rarity of this species as well as occurrence in saline areas can be observed, as indicated by SDM, which indicates its potential in reclamation efforts. Additionally, the VIP scores for the environmental variables indicated salinity as among the most important variables influencing the presence of this species (Fig. 3).

Table 1
Performance of various spatial interpolation schemes for salinity prediction in Thailand. The rows indicate average and standard deviation results for a 10-fold cross-validation. The best performing model (RF) was used to generate salinity predictions for Thailand.

Model	OK	RF	GLM	GAM	RF OK	GLM OK	GAM OK
Average RMSE (dS/m)	11.531	1.732	8.442	2.317	2.434	7.828	1.964
Standard Dev. (dS/m)	8.769	2.389	4.916	2.418	2.251	5.090	2.449

Table 2
Performance metrics of various species distribution modeling (SDM) schemes, including AUC (area under the curve), COR (correlation), TSS (true skill statistic), and deviance, determined through cross-validation.

Method	AUC	COR	TSS	Deviance
RF	0.98	0.92	0.93	0.26
MaxEnt	0.98	0.9	0.9	0.52
SVM	0.97	0.9	0.89	0.36

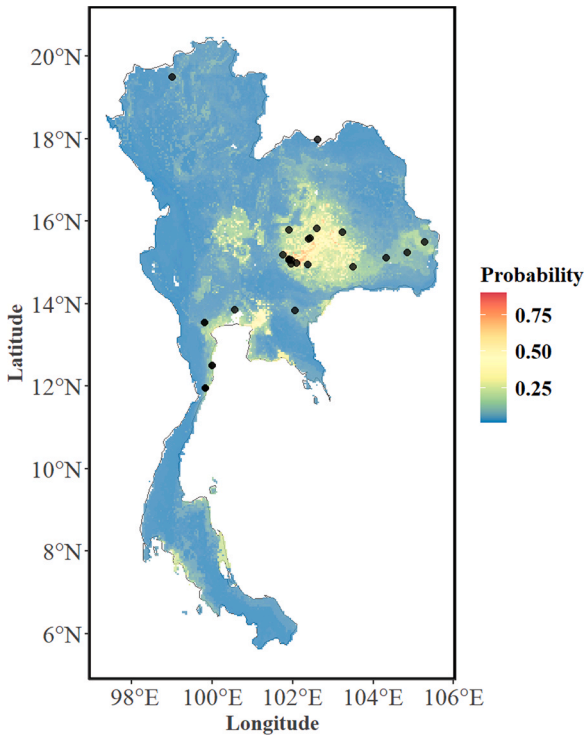

Fig. 2. The predicted probability of the spatial distribution of *B. siamensis* as predicted by a weighted ensemble of RF, SVM, and MaxEnt. Varying probabilities of species presence can be seen, with higher probabilities indicated in red. The points overlaid indicate the presence of the species as per the species database and personal observations.

3.3. Metabolomic analysis of B. siamensis leaf samples

As per the previous modeling, cultivation of *B. siamensis* in saline conditions could enhance the production of valuable secondary metabolites, supporting both its conservation and commercial utilization. Therefore, a foliar LC-QTOF metabolmic analysis was used to determine the qualitative foliar secondary metabolite profile of trees found in non-saline and saline locations. From over 10,000 features, a total of 1106 were annotated as metabolites across the leaf samples, as detailed in

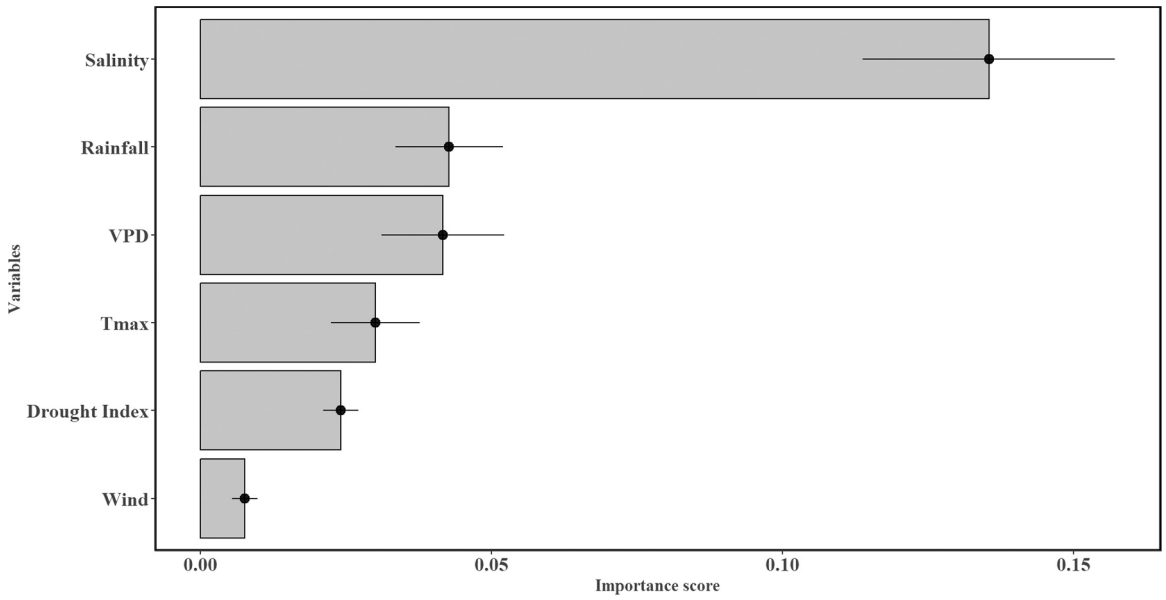

Fig. 3. Variable importance of the best performing species distribution model (RF) used to determine the spatial distribution of *B. siamensis*.

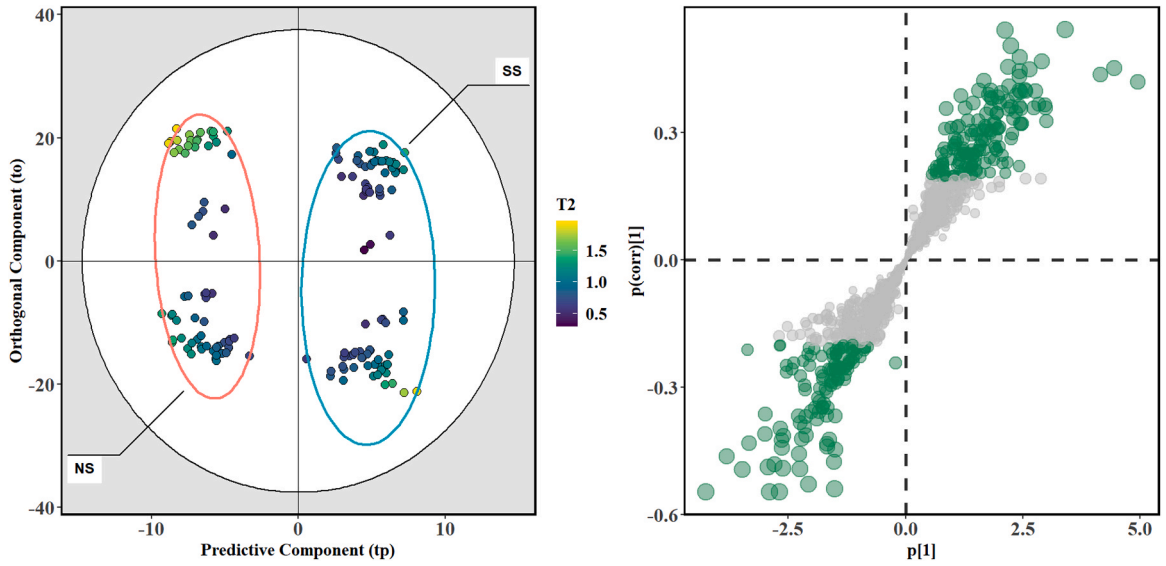

Fig. 4. Cross-validated OPLS-DA scores plot (left panel) and S-plot to classify saline (SS) and non-saline (NS) samples. The scores plot shows the distribution of samples across the predictive and orthogonal components, with the black ellipse obtained through Hotelling's T^2 test (at a significance of 99 %), while the two colored ellipses contain 99 % of the sample points in the two salinity classes. The color gradient of points represents the T^2 statistic, which signifies the confidence of classification for each sample. The S-plot indicates metabolites deemed significant in contributing to the discrimination between the two categories, based on $p < 0.05$ and VIP > 1. The size of points indicates the VIP scores of the respective metabolite with the significant metabolites highlighted in green.

Supplementary Table S1. Among these 1106 identified metabolites, 175 were found to be significantly different between the saline (SS) and non-saline (NS) samples, based on False Discovery Rate (FDR)-adjusted p-values ($p < 0.05$), with 108 metabolites showing higher abundance in the SS group (see Supplementary Table S2).

Fig. 4 presents the results of OPLS-DA scores plot in the left panel to determine the separation between the sample groups, with the 99 % confidence ellipse generated using Hotelling's T^2 test. The right panel contains an S-plot to isolate metabolites that were significant in differentiating between the saline and non-saline leaf samples. It can be seen that the SS and NS leaf samples are very well-separated. Metabolites positioned at the extremes of this plot as well as having a VIP scores greater than 1 and $p < 0.05$ are plotted in green circles and were identified as highly influential, both in terms of their contribution to class separation as well as strong correlation with the predictive component (Supplementary Table S3). Notably, metabolites found in *B. siamensis* leaves with higher abundance in saline samples included brucine, hispiduloside, and corylin ($p < 0.0001$, VIP scores > 2.0), as well as liquiritin, mollugin, canthin-6-one, and baicalin ($p < 0.001$, VIP scores > 1.5), suggesting their potential role in defense mechanisms against salinity-induced stress and contributing to plant survival under saline conditions.

3.4. Antioxidant activity

Antioxidant activities of leaf samples from trees in non-saline (NS) and saline (SS) soils were compared using boxplots through three different assays that included DPPH radical scavenging activity and

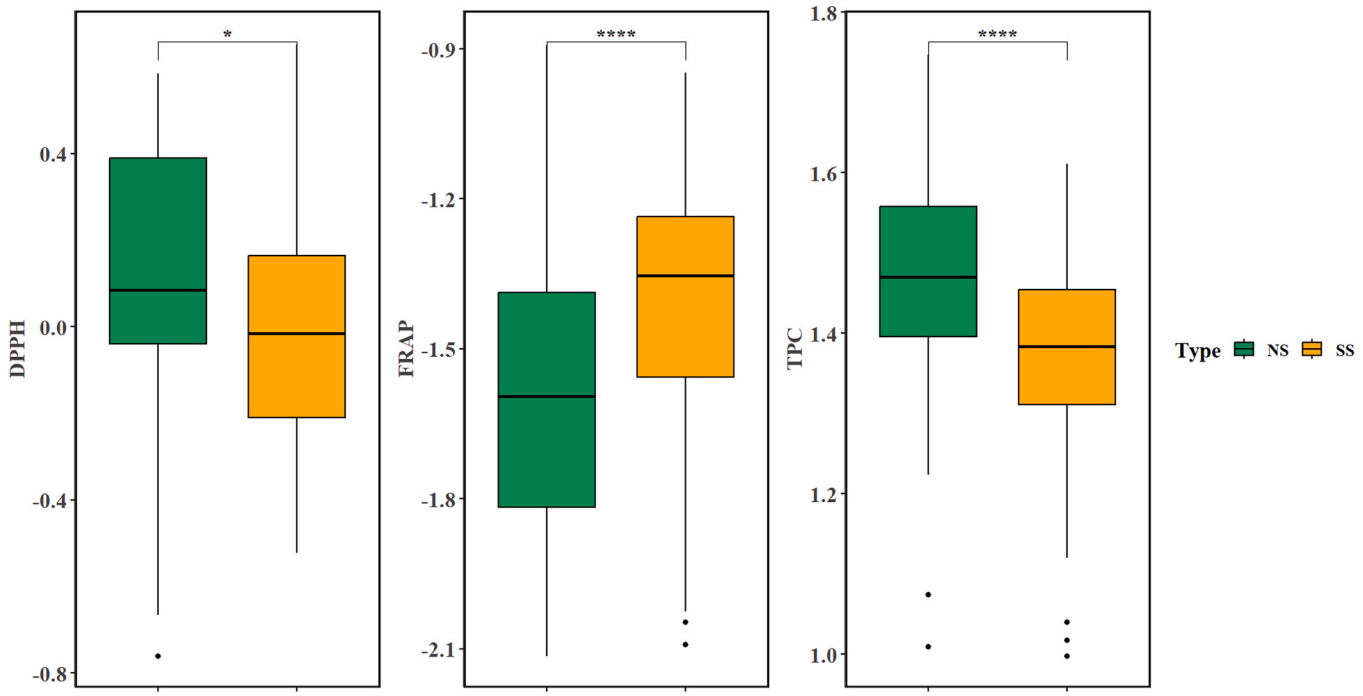

Fig. 5. Boxplots comparing the scaled antioxidant activity (DPPH, FRAP) and TPC of leaf samples collected from plants growing in non-saline (NS) and saline (SS) soils. Statistical significance is indicated by asterisks with $^{*}p < 0.05$ and $^{****}p < 0.0001$.

FRAP, along with TPC, as shown in Fig. 5. For the FRAP assay, the NS group exhibited significantly lower reducing power compared to the SS group ($p < 0.0001$), while DPPH and TPC were significantly lower under saline conditions. FRAP activity ranged between 2.6 to about 56.4 mg μM Fe^{2+}/g dry weight of leaf (about 22-fold change). The percentage inhibition of DPPH activity ranged from 33.1 % to 91.4 %, (about 3-fold change), with TPC ranging from 135.4 to about 718.8 mg GAE/g dry weight of leaf (about 5-fold change).

A further search for key metabolites associated with antioxidant activities was conducted by regressing the metabolomics profile against the levels of the three antioxidant assays to construct three respective O-PLSR models (Fig. 6A–C). All the models were found to be highly significant ($p < 0.0001$) as indicated by the cross-validated analysis of variance (ANOVA) diagnostic on the residuals of the regression between cross-validated scores of the predictive O-PLS component and the response in terms of the antioxidant activities. Additionally, the three O-PLSR models had similar and reasonably good predictive abilities (DPPH: R2X = 0.46, Q2Y = 0.37, FRAP: R2X = 0.47, Q2Y = 0.37, and TPC: R2X = 0.43, Q2Y = 0.28).

In all the figures, metabolites located in the upper right quadrant indicate stronger positive correlation with higher levels of respective antioxidant activity as measured by DPPH, FRAP, or TPC. On the contrary, metabolites in the lower left quadrant indicate negative correlation or higher levels of the metabolite associated with decreased activity. Additionally, the magnitude of variance (or variations either along the positive or *x-axis*) indicates the contribution of each metabolite to the activity. Metabolites located farther from the origin have higher contribution to the observed variance and are important for the activity being analyzed. The results showed that afzelin was a key metabolite responsible for antioxidant properties of *B. siamensis* leaves, as it was consistently located in the upper right quadrant of the S-plot across DPPH (Fig. 6A), FRAP (Fig. 6B), and TPC (Fig. 6C) assays. Additionally, its levels were significantly elevated in the SS group compared to the NS group, as indicated by the average peak area ($p < 0.05$; Fig. 7)

4. Discussion

4.1. Salinity prediction through spatial mapping

Land degradation, from issues such as soil salinization, is a growing environmental challenge that can negatively impact both agricultural productivity and ecosystem vitality. Bioremediation strategies, such as planting of salt-tolerant species in saline areas can be used to mitigate such issues. Through this study, we propose a platform that combines the power of spatial mapping techniques, such as interpolation and species distribution modeling (SDM) with untargeted metabolomics, to screen for suitable species for reclaiming such degraded environments. As a case study, we analyzed the potential of *B. siamensis*, a rare, native species for planting in salt-affected areas, based on its frequent occurrence in such environments of northeastern Thailand. The SDM predictions of the most probable presence locations of the species were based on significant environmental and soil variables, including salinity.

Several schemes were used to predict the salinity raster in Thailand through a k-fold cross-validation using soil variables, with a 70:30 train-test split was used to validate each model and reduce overfitting. RF was the best performing model in predicting the soil salinity across Thailand as evidenced by a low testing RMSE and standard deviation (see Table 1). The superior performance of the RF algorithm in handling complex, high-dimensional data can be attributed to the model's ability to capture any nonlinear interactions between soil variables [43], as heterogeneous environments can induce variations in salinity influenced by multiple interrelated factors. RF has also been reported to perform well with an overall accuracy of 97 % in detecting salinization using remote sensing data [44].

SDM was then used to determine the most probable locations for *B. siamensis* after excluding collinear variables through *vif* to avoid redundancy and to improve SDM prediction accuracy [45]. Previous studies have demonstrated that optimizing SDM inputs by reducing collinearity can significantly reduce model bias, improving statistical significance and prediction accuracy [46]. From among the models used, RF, SVM, and MaxEnt were identified as the best performing classifiers (see Table 2), with RF performing the best among these (as

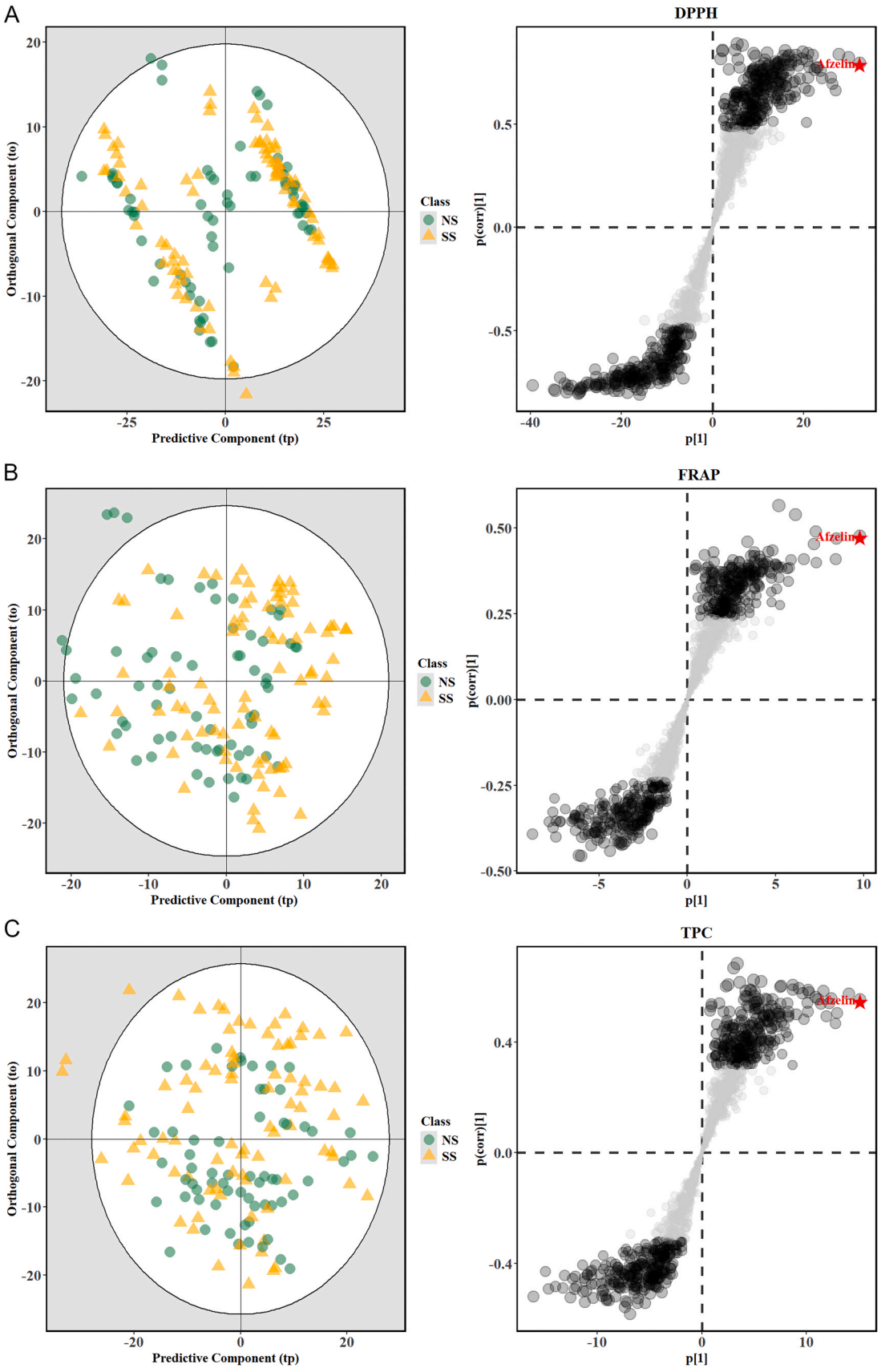

Fig. 6. A-C. Cross-validated O-PLS score plots (left panel) and S-plots (right panel) obtained from the regression of metabolomic profiles with the respective antioxidant activity: **(A)** DPPH, **(B)** FRAP, and **(C)** total phenolic content (TPC). The ellipse obtained through Hotelling's T^2 test (at a significance of 99 %). Significant metabolites ($p < 0.05$, VIP > 1) are highlighted in black and the location of Afzelin as a key metabolite is annotated by a red star.

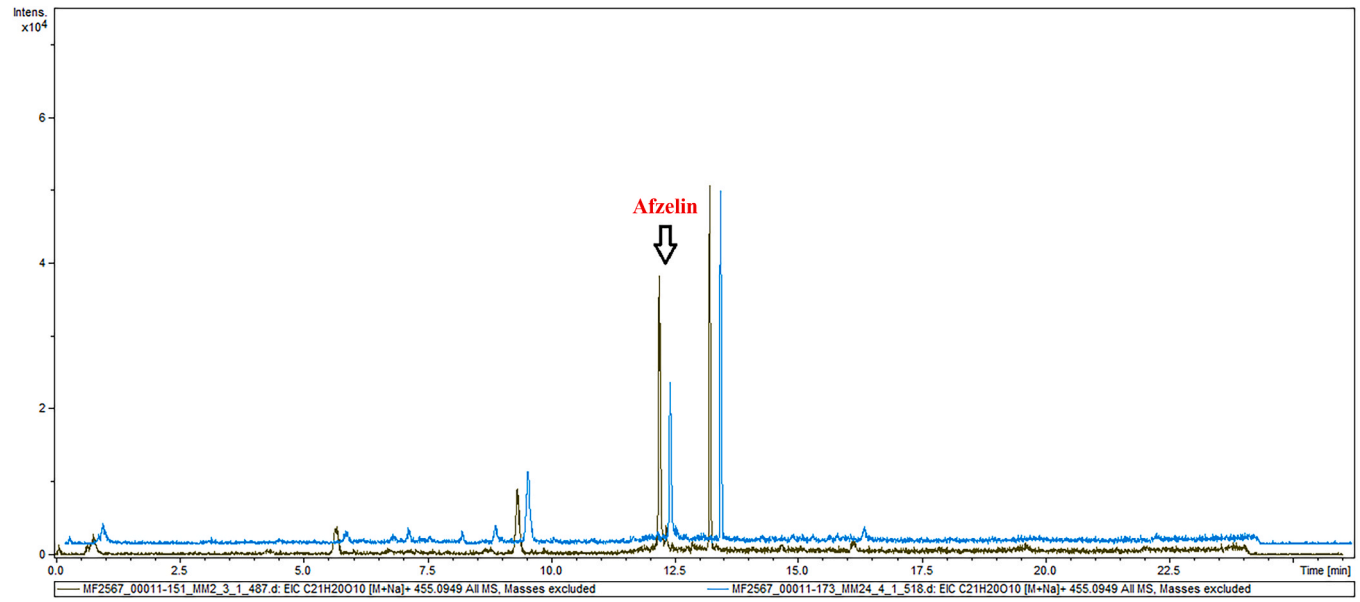

Fig. 7. Representative chromatograms of Afzelin from LC-MS metabolomic analysis of *B. siamensis* leaf samples of trees growing in non-saline (blue) and saline (black) soils.

indicated by the highest classification metrics, including AUC, COR, and TSS), demonstrating its ability to handle complex environmental data and nonlinear interactions [47].

Salinity has long been recognized as a critical factor determining species tolerance and establishment in arid and semi-arid environments [48]. A high VIP score for salinity (Fig. 3) in the best performing SDM RF classifier is indicative of *B. siamensis* being able to exhibit unique physiological adaptations that enable its survival in such stressful environments [49]. This finding aligns with previous research highlighting salinity as a key environmental stressor that influences plant physiology and ecology [50]. Plants inhabiting saline environments often exhibit specialized adaptations, such as osmotic regulation and ion transport mechanisms, enabling them to tolerate high salt concentrations [49]. Similarly, variables which exhibited a high VIP score like VPD and Tmax are critical in influencing evapotranspiration rates and water availability, further affecting species viability. VPD, which is indicative of atmospheric dryness, can directly affect transpiration rates, leading to increased water loss and stress, particularly in arid regions [51]. Additionally, a combined effect of high VPD and Tmax can severely impact plant vitality, especially in drought-prone areas [52].

The predicted distribution of SDM was further refined through an ensemble model, which combined the outputs of the three best performing models RF, MaxEnt, and SVM. Ensemble methods have been widely recognized for improving model performance by leveraging the strengths of individual algorithms, thus reducing error, bias, and variance in predictions relative to the individual models [53]. The frequent occurrence of *B. siamensis* in saline areas in the ensemble model suggests its suitability in saline conditions, aiding in the selection of appropriate species for bioremediation efforts of such degraded ecosystems [53].

4.2. Metabolite profiling of B. siamensis leaves through metabolomic analysis

To analyze these adaptations, a qualitative foliar metabolomic approach was used to identify features that might be associated with complex metabolic adaptations related to salinity stress [54]. Metabolomic profiling has emerged as a powerful tool for screening of interesting metabolites with potential therapeutic applications [55]. The untargeted screening of the foliar metabolome identified 1106 metabolites, with 108 metabolites showing higher abundance in the SS group ($p < 0.05$). Among other significant metabolites, brucine, hispiduloside, and corylin demonstrated strong statistical significance ($p < 0.0001$) and high VIP scores (VIP > 2.0), highlighting their higher abundance in the foliar metabolome of *B. siamensis* growing in saline soils.

Under salinity stress, plants exhibit complex metabolic adaptations that enhance their tolerance to ionic and osmotic imbalances. The overall metabololite profiles reflected the presence of certain metabolites that have osmoprotective tendencies and stabilizing cellular structures. Elevated levels of stress-responsive alkaloids like brucine highlighted metabolic mechanisms that could be involved with stress mitigation under salinity. Brucine, an alkaloid commonly derived from *Strychnos* species, has been reported to have anti-cancer, anti-inflammatory, analgesic activities and effects related to the cardiovascular system [56]. Its elevated presence in saline conditions is consistent with reports of increased alkaloid production under environmental stress as a protective mechanism [57].

Hispiduloside is a flavonoid glycoside composed of a flavonoid core linked to a sugar moiety. It has been identified in various plant species, particularly those belonging to the *Clerodendrum* genus [58]. Its aglycone form, hispidulin, has been reported to exhibit potential anticancer activity against hepatocellular carcinoma, renal cell carcinoma, and pancreatic cancer, as demonstrated in mouse xenograft models [59]. Additionally, animal studies have indicated that hispidulin can suppress pro-inflammatory cytokines and modulate microbial activity, contributing to its neuroprotective effects against excitotoxic compound-induced neurotoxicity [60].

Corylin is an isoflavonoid compound found in Fabaceae species, particularly *Psoralea corylifolia* L. Numerous studies have highlighted its anti-inflammatory properties, as demonstrated in both *in vitro* and *in vivo* models. It has been shown to inhibit the production of prostaglandin E2 and nitric oxide, as well as reduce levels of pro-inflammatory cytokines [61]. Additionally, it exhibits protective effects against oxidative stress-related complications in diabetes, as evidenced by studies in mouse models [62]. Furthermore, corylin has been reported to regulate genes involved in obesity, leading to reduced fat accumulation and lowering body weight [63].

In addition, several metabolites, including liquiritin, mollugin, canthin-6-one, and baicalin, were also found to have a higher presence in the SS group. Liquiritin is a flavone derivative, primarily derived from licorice roots of *Glycyrrhiza spp.*, with antioxidant, anti-inflammatory

and hepatoprotective properties [64]. Its role in adaptive metabolic adjustments enhances antioxidant responses in saline environments, highlighting its significance as a key metabolite in plants subjected to abiotic stress [65]. Mollugin, a naphthohydroquinone compound, is primarily found in *Rubia cordifolia*. Several studies have reported its potential anticancer activity [66], as it can inhibit pro-inflammatory cytokines [67], as well as its neuroprotective and anti-inflammatory effects [68,69]. Canthin-6-one is an alkaloid predominantly found in the *Rutaceae* and *Simaroubaceae* families. This compound has demonstrated *in vitro* antimicrobial activity [70] and anti-inflammatory [71]. Baicalin is a flavonoid compound found in *Oroxylum indicum* and *Scutellaria baicalensis*. Animal studies have demonstrated its potential in alleviating inflammatory disorders such as pulmonary fibrosis, arthritis, and obesity [72]. Additionally, it has been reported to enhance the production of short-chain fatty acids, which influence gut microbiota composition by increasing beneficial microbes and reducing harmful bacteria [73].

4.3. Antioxidant activity of B. siamensis leaves

Antioxidant activity is a primary test for plant bioactive effects because oxidative stress contributes to cellular damage and various diseases. Thus, assessing antioxidant activity can indicate potential health benefits, including anti-inflammatory, anti-aging, and neuroprotective effects. It also links to disease prevention, as oxidative stress plays a role in cancer, cardiovascular, and neurodegenerative diseases. Many antioxidants have additional effects that are antimicrobial and anticancer in nature, making them useful in pharmaceuticals and nutraceuticals.

The DPPH assay measures the free radical scavenging potential, with a higher DPPH activity indicating stronger antioxidant capacity, reflecting a sample's ability to neutralize DPPH radicals by donating hydrogen atoms or electrons [74]. The FRAP assay estimates the reducing power of antioxidants based on their ability to reduce the ferric (Fe^{3+}) to ferrous (Fe^{2+}) form, where a higher FRAP value indicates stronger antioxidant activity [75]. TPC (Total Phenolic Content) quantifies the antioxidant phenolic compounds, such as flavonoids and tannins, present in the sample that have the ability to donate hydrogen atoms and quench free radicals, assisting in plant defense against stress by neutralizing ROS [76].

In this study, a significant reduction in the %DPPH inhibition under saline conditions suggests that salinity stress could decrease the plant's ability to scavenge free radicals. Conversely, for the FRAP assay, the NS group exhibited significantly lower reducing power compared to the SS group ($p < 0.0001$), indicating that salinity stress substantially elevated the plant's ferric-reducing capacity or antioxidant defense system under saline conditions. In addition, significant reductions in TPC in saline-stressed plants was observed compared to non-saline plants ($p < 0.0001$). The decrease in TPC may indicate that phenolic compounds had already been utilized in antioxidant defense mechanisms, leading to lower measured values. Additionally, other non-phenolic compounds could contribute to the antioxidant activity detected by the FRAP assay. These compounds could maintain high activity despite the lower phenolic content and may correspond to the distinct reaction mechanisms of antioxidant assays, which interact uniquely with the compounds present in various organs of a plant [77], including the leaves of *B. siamensis*.

The varying antioxidant capacities between saline and non-saline extracts highlights the species' resilience and adaptability to harsh conditions. Salinity induced changes in antioxidant activity can be species-specific [78]. The antioxidant activity (as quantified through DPPH, FRAP, and TPC) corresponded with the salinity levels in *Moringa oleifera* [79]. However, elevated antioxidant activity was reported at moderate to high salinity stress in Artichoke [80], *Salvadora persica* [81], sweet marjoram [82], and *Halophila johnsonii* [83]. On the other hand, reduced antioxidant activity was reported for *Mesembryanthemum edule* L. [84] and wheat [85]. Although maximum TPC was observed at lower levels of salinity in *G. maritima*, the individual phenolics were higher at higher levels of salinity [78]. The DPPH was found to be significantly lower under saline conditions, previously noted by Hassan et al. [2], who concluded that exposure to salinity could reduce %DPPH by around 50 %. Additionally, a higher antioxidant capacity measured by FRAP systems with increasing salinity has been reported previously [79]. Furthermore, the lowering in TPC with increased salinity have also been reported in *Acalypha wilkesiana* [86], *Halophila johnsonii* [83] and bean [87].

4.4. Key metabolite associated with antioxidant activity

O-PLSR analysis was used to determine the relationship between metabolite profiles and associated bioactivities, offering a targeted approach for key metabolite identification [88]. In the present study, afzelin was identified as the key metabolite contributing to the antioxidant properties of *B. siamensis* leaves, with significantly higher levels in plants sampled from saline locations ($p < 0.05$). Afzelin is a flavonoid glycoside composed of a polyphenolic core and a sugar moiety, also known as kaempferol-3-O-rhamnoside. This compound is widely distributed across various plant species, including *Quercus* spp. and *Crataegus* spp. [89]. It exhibits multiple bioactive properties, including antioxidant, anti-inflammatory, anticancer, antidiabetic, antibacterial, and neuroprotective effects [90]. Its antioxidant activity has been assessed using DPPH and FRAP assays, demonstrating its ability to scavenge reactive oxygen species (ROS) and chelate metal ions, thereby inhibiting free radical generation [89]. Additionally, afzelin has been reported to exert anti-inflammatory effects by suppressing cytokine production in inflammatory pathways, as shown in *in vitro* studies [91]. Furthermore, *in vivo* studies indicate that afzelin possesses neuroprotective properties, contributing to improved cognitive function in animal models [92].

Given its promising bioactive properties, future research should focus on optimizing the extraction process of afzelin from *B. siamensis* leaves to enhance yield and purity. Additionally, comprehensive studies on its safety, pharmacokinetics, and efficacy are essential to establish its therapeutic potential. Cultivating *B. siamensis* in saline-prone areas could thus provide bioremediation of degraded lands, transforming unproductive, highly saline soils to obtain valuable agricultural produce. This strategy would not only enhance ecological sustainability but also offer economic benefits by utilizing otherwise unproductive land for bioactive compound production. Ultimately, well-designed preclinical and clinical trials will be crucial in paving the way for afzelin's application in improving human health and pharmaceutical development, contributing to both environmental and economic value.

5. Conclusions

In this study, we present an approach that integrates large-scale spatial mapping, utilizing species distribution modeling, with small molecule untargeted metabolomics to screen medicinal plant species for reclamation of saline and degraded areas devoid of vegetation. Applied to *B. siamensis*, a rare native species with pharmaceutical applications

and frequently found in salt-affected areas of Thailand, the ensemble SDM model highlighted salinity as a critical variable influencing the distribution of this species. By using an untargeted metabolomics approach, a detailed list of plant-derived small bioactive molecules was acquired. Furthermore, antioxidant assays were performed, and computational analyses were employed to correlate these datasets, identifying key potential bioactive metabolites.

By leveraging this multidisciplinary approach, we developed a platform that simultaneously advances bioactive discovery and saline land reclamation, demonstrating dual environmental and pharmaceutical benefits. This integrative framework holds significant potential for application in the pharmaceutical industry, particularly in drug discovery initiatives. The proposed screening platform can be scaled to identify other rare, native species with ecological adaptability and bioactive benefits for a sustainable land restoration. Further refinement of the SDM with additional soil data and targeted metabolomics for metabolites such as afzelin, brucine, hispiduloside, and corylin could enhance species selection for sustainable land management and create economic opportunities during ecosystem restoration.

CRediT authorship contribution statement

Tushar Andriyas: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Resources, Validation, Visualization, Writing – original draft. **Nisa Leksungnoen:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Pichaya Pongchaidacha:** Formal analysis, Investigation, Methodology, Validation, Writing – original draft. **Arashaporn Uthairangsee:** Formal analysis, Investigation, Methodology, Writing – original draft. **Suwimon Uthairatsamee:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Resources, Writing – original draft. **Peerapat Doomnil:** Investigation, Validation, Visualization. **Yongkriat Ku-Or:** Investigation, Validation, Visualization. **Chatchai Ngernsaengsaruay:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Writing – original draft. **Sanyogita Andriyas:** Formal analysis, Software. **Arerut Yarnvudhi:** Investigation, Validation, Visualization. **Rossarin Tansawat:** Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The author is an Editorial Board Member/Editor-in-Chief/Associate Editor/Guest Editor for this journal and was not involved in the editorial review or the decision to publish this article.

Acknowledgements

This research is funded by Thailand Science Research and Innovation Fund Chulalongkorn University [Grant No. HEA_FF_68_295_3300_018]. This research project is supported by the Second Century Fund (C2F), Chulalongkorn. We sincerely thank Prof. Oliver Fiehn from the West Coast Metabolomics Center, University of California, Davis, for his invaluable support in providing a letter of recommendation for our funding application and for reviewing the manuscript. We also extend our gratitude to Emeritus Prof. Daren Cornforth from the Department of Nutrition, Dietetics, and Food Sciences at Utah State University, USA, for his valuable assistance in English language editing and proofreading the manuscript. Finally, we thank Dr. Pantamith Rattanakrajang for his graphical assistance.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2025.04.035.

Data Availability

This article contains supporting information in the Supplementary Tables.

References

[1] Gupta SR, Dagar JC. Tree plantations in saline environments: ecosystem services, carbon sequestration and climate change mitigation. Agrofor Manag Water Saline Soils Poor-Qual Waters 2016:181–95.
[2] Hassan AHA, Alkhalifah DHM, Al Yousef SA, Beemster GTS, Mousa ASM, Hozzein WN, et al. Salinity stress enhances the antioxidant capacity of bacillus and planococcus species isolated from saline lake environment. Front Microbiol 2020; 11:561816.
[3] Setia R, Gottschalk P, Smith P, Marschner P, Baldock J, Setia D, et al. Soil salinity decreases global soil organic carbon stocks. Sci Total Environ 2013;465:267–72.
[4] Wong VNL, Murphy BW, Koen TB, Greene RSB, Dalal RC. Soil organic carbon stocks in saline and sodic landscapes. Soil Res 2008;46:378–89.
[5] Leksungnoen N, Uthairatsamee S, Takuathung CN. Germination test on native salt tolerant seeds (Buchanania siamensis Miq.) collected from natural saline and non-saline soil. Thai J 2016;35:1–14.
[6] Leksungnoen N. Reclaiming saline areas in Khorat Basin (Northeast Thailand): soil properties, species distribution, and germination of potential tolerant species. Arid L Res Manag 2017;31. https://doi.org/10.1080/15324982.2017.1305467.
[7] Balasubramaniam T, Shen G, Esmaeili N, Zhang H. Plants' response mechanisms to salinity stress. Plants 2023;12:2253.
[8] Gill SS, Tuteja N. Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. Plant Physiol Biochem 2010;48:909–30.
[9] Ravishankar Ramakrishna A. GA. Influence of abiotic stress signals on secondary metabolites in plants. Plant Signal Behav 2011;6:1720–31. https://doi.org/10.4161/psb.6.11.17613.
[10] Hmidi D, Abdelly C, Athar H-R, Ashraf M, Messedi D. Effect of salinity on osmotic adjustment, proline accumulation and possible role of ornithine-$δ $-aminotransferase in proline biosynthesis in Cakile maritima. Physiol Mol Biol Plants 2018;24:1017–33.
[11] Alam MA, Juraimi AS, Rafii MY, Hamid AA, Aslani F, Alam MZ. Effects of salinity and salinity-induced augmented bioactive compounds in purslane (Portulaca oleracea L.) for possible economical use. Food Chem 2015;169:439–47.
[12] Yang L, Wen K-S, Ruan X, Zhao Y-X, Wei F, Wang Q. Response of plant secondary metabolites to environmental factors. Molecules 2018;23. https://doi.org/10.3390/molecules23040762.
[13] Wahid A, Ghazanfar A. Possible involvement of some secondary metabolites in salt tolerance of sugarcane. J Plant Physiol 2006;163:723–30.
[14] Araújo MB, New M. Ensemble forecasting of species distributions. Trends Ecol \ Evol 2007;22:42–7.
[15] de Souza Muñoz ME, De Giovanni R, de Siqueira MF, Sutton T, Brewer P, Pereira RS, et al. openModeller: a generic approach to species' potential distribution modelling. Geoinformatica 2011;15:111–35.
[16] Diniz-Filho JAF, Mauricio Bini L, Fernando Rangel T, Loyola RD, Hof C, Nogués-Bravo D, et al. Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. Ecography (Cop) 2009;32:897–906.
[17] Pell SK. Molecular systematics of the cashew family (Anacardiaceae). Louisiana State University and Agricultural \& Mechanical College; 2004.
[18] Daly DC, Harley MM, Mart\'\inez-Habibe MC. Weeks A. Flowering plants. Eudicots: Sapindales, Cucurbitales, Myrtaceae. Burseraceae. Berlin: Springer Berlin-Heidelberg; 2011. p. 76–104.
[19] Chamberlain SA, Boettiger C. R Python, Ruby Clients GBIF Species Occur data 2017.
[20] Petchsomrit A, Chanthathamrongsiri N, Manmuan S, Leelakanok N, Wangpradit N, Vongsak B, et al. Green extraction of Buchanania siamensis and water-based formulations. Sustain Chem Pharm 2022;30:100883.
[21] Leksungnoen Yongkriat K-O, Ngernsaengsaruay N, Andriyas C. T. Seed longevity of Buchanania siamensis in reclaiming salt-affected areas in Thailand. Biodiversitas J Biol Divers 2020;21.
[22] Prompanya C, Petchsomrit A, Vongsak B. Antioxidant evaluation and HPLC analysis of Buchanania lanzan and Buchanania siamensis leaf extracts. J Res Pharm 2023;27.
[23] Chuakul W, Saralamp P, Boonpleng A. Medicinal plants used in the Kutchum district, Yasothon Province, Thailand. Thai J Phytopharm 2002;9:22–49.
[24] Schulze-Kaysers N, Feuereisen MM, Schieber A. Phenolic compounds in edible species of the Anacardiaceae family–a review. RSC Adv 2015;5:73301–14.
[25] Yodsaoue O, Kamonwannasit S, Frangkrathok N. Antioxidant and anticancer activities of Buchanania siamensis Miq. stem and leaf extracts. J Sci Technol Ubon Ratchathani Univ Spec Issue Sept 2017;2017:13–6.
[26] Sarkar AK, Sadhukhan S. Bioremediation of salt-affected soil through plant-based strategies. Adv Bioremediation Phytoremediat Sustain Soil Manag Princ Monit Remediat 2022:81–100.

[27] Tugizimana F, Piater L, Dubery I. Plant metabolomics: A new frontier in phytochemical analysis. S Afr J Sci 2013;109:1–11.
[28] Kapoore RV, Vaidyanathan S. Towards quantitative mass spectrometry-based metabolomics in microbial and mammalian systems. Philos Trans R Soc A Math Phys Eng Sci 2016;374:20150363.
[29] Steinfath M, Strehmel N, Peters R, Schauer N, Groth D, Hummel J, et al. Discovering plant metabolic biomarkers for phenotype prediction using an untargeted approach. Plant Biotechnol J 2010;8:900–11.
[30] Maitner BS, Boyle B, Casler N, Condit R, Donoghue J, Durán SM, et al. The bien r package: A tool to access the Botanical Information and Ecology Network (BIEN) database. Methods Ecol Evol 2018;9:373–9.
[31] Chamberlain S, Oldoni D, Waller J. rgbif: Interface Glob Biodivers Inf Facil API 2022.
[32] Dent D. International Soil Reference and Information Centre (ISRIC). Encycl Soil Sci. CRC Press Boca Raton, FL, USA; 2017. p. 1232–6.
[33] R. Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria 2023.
[34] Abatzoglou JT, Dobrowski SZ, Parks SA, Hegewisch KC. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. Sci Data 2018;5:1–12.
[35] Bishop TFA, McBratney AB. A comparison of prediction methods for the creation of field-extent soil property maps. Geoderma 2001;103:149–60.
[36] Naimi B, Araújo MB. sdm: a reproducible and extensible R platform for species distribution modelling. Ecography (Cop) 2016;39:368–75.
[37] Cushman SA, Kilshaw K, Campbell RD, Kaszta Z, Gaywood M, Macdonald DW. Comparing the performance of global, geographically weighted and ecologically weighted species distribution models for Scottish wildcats using GLM and Random Forest predictive modeling. Ecol Model 2024;492:110691.
[38] Veeramohan R, Zamani AI, Azizan KA, Goh HH, Aizat WM, Razak MFA, et al. Comparative metabolomics analysis reveals alkaloid repertoires in young and mature Mitragyna speciosa (Korth.) Havil. Leaves. PLoS One 2023;18:1–24. https://doi.org/10.1371/journal.pone.0283147.
[39] Schiffman C, Petrick L, Perttula K, Yano Y, Carlsson H, Whitehead T, et al. Filtering procedures for untargeted LC-MS metabolomics data. BMC Bioinforma 2019;20:1–10.
[40] Tsugawa H, Cajka T, Kind T, Ma Y, Higgins B, Ikeda K, et al. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. Nat Methods 2015;12(6):523.
[41] Jayasekera S, Molan AL, Garg M, Moughan PJ. Variation in antioxidant potential and total polyphenol content of fresh and fully-fermented Sri Lankan tea. Food Chem 2011;125:536–41.
[42] Lawag IL, Nolden ES, Schaper AAM, Lim LY, Locher C. A modified folin-ciocalteu assay for the determination of total phenolics content in honey. Appl Sci 2023;13:2135.
[43] Xiao C, Ji Q, Chen J, Zhang F, Li Y, Fan J, et al. Prediction of soil salinity parameters using machine learning models in an arid region of northwest China. Comput Electron Agric 2023;204:107512.
[44] Sahbeni G. Comparative Study of Machine-Learning-Based Classifiers for Soil Salinization Prediction using Sentinel-1 SAR and Sentinel-2 MSI Data. 2022 10th Int. Conf Agro-geoinformatics 2022:1–4.
[45] Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography (Cop) 2013;36:27–46.
[46] De Marco P, Nóbrega CC. Evaluating collinearity effects on species distribution models: An approach based on virtual species simulation. PLoS One 2018;13:e0202403.
[47] Valavi R, Elith J, Lahoz-Monfort JJ, Guillera-Arroita G. Modelling species presence-only data with random forests. Ecography (Cop) 2021;44:1731–42.
[48] Uçarl i C. Effects of salinity on seed germination and early seedling stage. Abiotic Stress Plants 2020;211.
[49] Hualpa-Ramirez E, Carrasco-Lozano EC, Madrid-Espinoza J, Tejos R, Ruiz-Lara S, Stange C, et al. Stress salinity in plants: New strategies to cope with in the foreseeable scenario. Plant Physiol Biochem 2024:108507.
[50] Lamsaadi N, Farssi O, El Moukhtari A, Farissi M. Different approaches to improve the tolerance of aromatic and medicinal plants to salt stressed conditions. J Appl Res Med Aroma Plants 2024:100532.
[51] Xu Z, Tian Y, Liu Z, Xia X. Comprehensive effects of atmosphere and soil drying on stomatal behavior of different plant types. Water 2023;15:1675.
[52] Danylenko I, Le Dantec V, Fanise P, Boujnah D, Cheheb H, Gascoin S, et al. Use of the Photochemical Reflectance Index to determine water stress in semi-arid climate conditions. EGU Gen Assem Conf Abstr 2023. EGU–13621.
[53] Harris J, Pirtle JL, Laman EA, Siple MC, Thorson JT. An ensemble approach to species distribution modelling reconciles systematic differences in estimates of habitat utilization and range area. J Appl Ecol 2024;61:351–64.
[54] Summer LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al. Proposed minimum reporting standards for chemical analysis. Metabolomics 2007;3:211–21.
[55] Stuart KA, Welsh K, Walker MC, Edrada-Ebel R. Metabolomic tools used in marine natural product drug discovery. Expert Opin Drug Discov 2020;15:499–522.
[56] Lu L, Huang R, Wu Y, Jin J-M, Chen H-Z, Zhang L-J, et al. Brucine: a review of phytochemistry, pharmacology, and toxicology. Front Pharm 2020;11:377.
[57] Al-Huqail AA, Ali EF. Effect of jasmonic acid on alkaloids content and salinity tolerance of Catharanthus roseus based on morpho-physiological evaluation. South Afr J Bot 2021;141:440–6.
[58] Tangsongcharoen T, Issaravanich S, Palanuvej C, Ruangrungsi N. Quantitative Analysis of Hispidulin Content in Clerodendrum petasites Roots Distributed in Thailand. Pharm J 2019;11:1093–9. https://doi.org/10.5530/pj.2019.11.171.
[59] Liu K, Zhao F, Yan J, Xia Z, Jiang D, Ma P. Hispidulin: A promising flavonoid with diverse anti-cancer properties. Life Sci 2020;259:118395. https://doi.org/10.1016/j.lfs.2020.118395.
[60] Lin TY, Lu CW, Wang SJ, Huang SK. Protective effect of hispidulin on kainic acid-induced seizures and neurotoxicity in rats. Eur J Pharm 2015;755:6–15. https://doi.org/10.1016/j.ejphar.2015.02.041.
[61] Hung Y-L, Fang S-H, Wang S-C, Cheng W-C, Liu P-L, Su C-C, et al. Corylin protects LPS-induced sepsis and attenuates LPS-induced inflammatory response. Sci Rep 2017;7:46299. https://doi.org/10.1038/srep46299.
[62] Seo E, Lee E-K, Lee CS, Chun K-H, Lee M-Y, Jun H-S. Psoralea corylifolia L. seed extract ameliorates streptozotocin-induced diabetes in mice by inhibition of oxidative stress. Oxid Med Cell Longev 2014;2014:897296. https://doi.org/10.1155/2014/897296.
[63] Chen C-C, Kuo C-H, Leu Y-L, Wang S-H. Corylin reduces obesity and insulin resistance and promotes adipose tissue browning through SIRT-1 and β3-AR activation. Pharm Res 2021;164:105291. https://doi.org/10.1016/j.phrs.2020.105291.
[64] Tang T-J, Wang X, Wang L, Chen M, Cheng J, Zuo M-Y, et al. Liquiritin inhibits H2O2-induced oxidative stress injury in H9c2 cells via the AMPK/SIRT1/NF-$κ$B signaling pathway. J Food Biochem 2022;46:e14351.
[65] Akram W, Yasin NA, Shah AA, Khan WU, Li G, Ahmad A, et al. Exogenous application of liquiritin alleviated salt stress and improved growth of Chinese kale plants. Sci Hortic (Amst) 2022;294:110762.
[66] Wang Z, Li MY, Mi C, Wang KS, Ma J, Jin X. Mollugin Has an Anti-Cancer Therapeutic Effect by Inhibiting TNF-α-Induced NF-κB Activation. Int J Mol Sci 2017;18. https://doi.org/10.3390/ijms18081619.
[67] Lu Y, Liu R, Sun C, Pan Y. An effective high-speed countercurrent chromatographic method for preparative isolation and purification of mollugin directly from the ethanol extract of the Chinese medicinal plant Rubia cordifolia. J Sep Sci 2007;30:1313–7. https://doi.org/10.1002/jssc.200600440.
[68] Jeong G-S, Lee D-S, Kim D-C, Jahng Y, Son J-K, Lee S-H, et al. Neuroprotective and anti-inflammatory effects of mollugin via up-regulation of heme oxygenase-1 in mouse hippocampal and microglial cells. Eur J Pharm 2011;654:226–34. https://doi.org/10.1016/j.ejphar.2010.12.027.
[69] Kim K-J, Lee JS, Kwak M-K, Choi HG, Yong CS, Kim J-A, et al. Anti-inflammatory action of mollugin and its synthetic derivatives in HT-29 human colonic epithelial cells is mediated through inhibition of NF-κB activation. Eur J Pharm 2009;622:52–7. https://doi.org/10.1016/j.ejphar.2009.09.008.
[70] Farouil L, Sylvestre M, Fournet A, Cebrián-Torrejón G. Review on canthin-6-one alkaloids: Distribution, chemical aspects and biological activities. Eur J Med Chem Rep 2022;5:100049. https://doi.org/10.1016/j.ejmcr.2022.100049.
[71] Cho S-K, Jeong M, Jang DS, Choi J-H. Anti-inflammatory Effects of Canthin-6-one Alkaloids from Ailanthus altissima. Planta Med 2018;84:527–35. https://doi.org/10.1055/s-0043-123349.
[72] Hu Z, Guan Y, Hu W, Xu Z, Ishfaq M. An overview of pharmacological activities of baicalin and its aglycone baicalein: New insights into molecular mechanisms and signaling pathways. Iran J Basic Med Sci 2022;25:14–26. https://doi.org/10.22038/IJBMS.2022.60380.13381.
[73] Wang X, Xie L, Long J, Liu K, Lu J, Liang Y, et al. Therapeutic effect of baicalin on inflammatory bowel disease: A review. J Ethnopharmacol 2022;283:114749. https://doi.org/10.1016/j.jep.2021.114749.
[74] Gulcin \.Ilhami, Alwasel SH. DPPH radical scavenging assay. Processes 2023;11:2248.
[75] Tanvir EM, Afroz R, Chowdhury MAZ, Khalil MI, Hossain MS, Rahman MA, et al. Honey has a protective effect against chlorpyrifos-induced toxicity on lipid peroxidation, diagnostic markers and hepatic histoarchitecture. Eur J Integr Med 2015;7:525–33.
[76] de Abreu IN, Mazzafera P. Effect of water and temperature stress on the content of active constituents of Hypericum brasiliense Choisy. Plant Physiol Biochem 2005;43:241–8.
[77] Xia P, Ahmmed MK. others. Exploring efficient extraction methods: Bioactive compounds and antioxidant properties from New Zealand damson plums. Food Biosci 2023;55:103057.
[78] Pungin A, Lartseva L, Loskutnikova V, Shakhov V, Popova E, Skrypnik L, et al. Effect of salinity stress on phenolic compounds and antioxidant activity in halophytes Spergularia marina (L.) Griseb. and Glaux maritima L. cultured in vitro. Plants 2023;12:1905.
[79] Azeem M, Pirjan K, Qasim M, Mahmood A, Javed T, Muhammad H, et al. Salinity stress improves antioxidant potential by modulating physio-biochemical responses in Moringa oleifera Lam. Sci Rep 2023;13:2895.
[80] Rezazadeh A., Ghasemnezhad A., Barani M., Telmadarrehei T. Effect of salinity on phenolic composition and antioxidant activity of artichoke (Cynara scolymus L.) leaves. 2012.
[81] Sharma V, Ramawat KG. Salinity-induced modulation of growth and antioxidant activity in the callus cultures of miswak (Salvadora persica). 3 Biotech 2013;3:11–7.
[82] Baätour O, Tarchoun I, Nasri N, Kaddour R, Harrathi J, Drawi E, et al. Effect of growth stages on phenolics content and antioxidant activities of shoots in sweet marjoram (Origanum majorana L.) varieties under salt stress. Afr J Biotechnol 2012;11:16486–93.
[83] Gavin NM, Durako MJ. Localization and antioxidant capacity of flavonoids in Halophila johnsonii in response to experimental light and salinity variation. J Exp Mar Bio Ecol 2012;416:32–40.
[84] Falleh H, Jalleli I, Ksouri R, Boulaaba M, Guyot S, Magné C, et al. Effect of salt treatment on phenolic compounds and antioxidant activity of two Mesembryanthemum edule provenances. Plant Physiol Biochem 2012;52:1–8.

[85] Rao A, Ahmad SD, Sabir SM, Awan SI, Hameed A, Abbas SR, et al. Detection of saline tolerant wheat cultivars (Triticum aestivum L.) using lipid peroxidation, antioxidant defense system, glycinebetaine and proline contents. J Anim Plant Sci 2013;23:1742–8.
[86] Odjegba VJ, Alokolaro AA. Simulated drought and salinity modulates the production of phytochemicals in Acalypha wilkesiana. J Plant Stud 2013;2:105.
[87] Telesiński A, Nowak J, Smolik B, Dubowska A, Skrzypiec N. Effect of soil salinity on activity of antioxidant enzymes and content of ascorbic acid and phenols in bean [Phaseolus vulgaris L.] plants. J Elem 2008;13.
[88] Marini F. Orthogonal PLS (O-PLS) and related algorithms. J Chemom 2020;34.
[89] Kciuk M, Garg N, Dhankhar S, Saini M, Mujwar S, Devi S, et al. Exploring the Comprehensive Neuroprotective and Anticancer Potential of Afzelin. Pharm (Basel) 2024;17.
[90] Mediani A., Hamezah H.S., Rohani E., Kamal N., Perumal V., Salim F., et al. Afzelin: Advances on Resources, Biosynthesis Pathway, Bioavailability, Bioactivity, and Pharmacology, 2023, p. 1–45.
[91] Kim Hee J, Kim M, Kim Min J, Lee M, Seo Jun S, Park Young K. Afzelin suppresses proinflammatory responses in particulate matter-exposed human keratinocytes. Int J Mol Med 2019;43:2516–22.
[92] Oh S-Y, Jang MJ, Choi Y-H, Hwang H, Rhim H, Lee B, et al. Central administration of afzelin extracted from Ribes fasciculatum improves cognitive and memory function in a mouse model of dementia. Sci Rep 2021;11:9182.