MOLECULAR ECOLOGY

Molecular Ecology (2016) doi: 10.1111/mec.13525

Exploring natural variation of *Pinus pinaster* Aiton using metabolomics: Is it possible to identify the region of origin of a pine from its metabolites?

MÓNICA MEIJÓN,*[1] ISABEL FEITO,* MICHAL ORAVEC,† CAROLINA DELATORRE,* WOLFRAM WECKWERTH,‡§ JUAN MAJADA¶ and LUIS VALLEDOR†**

*Regional Institute for Research and Agro-Food Development in Asturias, Experimental Station "La Mata", 33820 Grado, Spain, †Czechglobe, Academy of Sciences of the Czech Republic, Bělidla 986/4a, 603 00, Brno, Czech Republic, ‡Department of Ecogenomics and Systems Biology, Faculty of Life Sciences, University of Vienna, Althanstrasse 14, 1090 Vienna, §Vienna Metabolomics Center, University of Vienna, Universitätsring 1, 1010 Vienna, ¶Forest and Wood Technology Research Centre, Experimental Station "La Mata", 33820 Grado, **Plant Physiology, University of Oviedo, Catedrático Rodrigo Uría, 33006 Oviedo*

Abstract

Natural variation of the metabolome of *Pinus pinaster* was studied to improve understanding of its role in the adaptation process and phenotypic diversity. The metabolomes of needles and the apical and basal section of buds were analysed in ten provenances of *P. pinaster*, selected from France, Spain and Morocco, grown in a common garden for 5 years. The employment of complementary mass spectrometry techniques (GC-MS and LC-Orbitrap-MS) together with bioinformatics tools allowed the reliable quantification of 2403 molecular masses. The analysis of the metabolome showed that differences were maintained across provenances and that the metabolites characteristic of each organ are mainly related to amino acid metabolism, while provenances were distinguishable essentially through secondary metabolism when organs were analysed independently. Integrative analyses of metabolome, environmental and growth data provided a comprehensive picture of adaptation plasticity in conifers. These analyses defined two major groups of plants, distinguished by secondary metabolism: that is, either Atlantic or Mediterranean provenance. Needles were the most sensitive organ, where strong correlations were found between flavonoids and the water regime of the geographic origin of the provenance. The data obtained point to genome specialization aimed at maximizing the drought stress resistance of trees depending on their origin.

Keywords: abiotic stress, adaptation, conifers, flavonoids, GC-MS, terpenoids, UPLC-Orbitrap-MS

Received 6 August 2015; revision received 12 November 2015; accepted 10 December 2015

Introduction

Maritime pine (*Pinus pinaster* Aiton) is one of the most common forest species in the Mediterranean area, playing an important role in its ecosystems and forestry industries. Currently most *P. pinaster* distribution areas correspond to plantations, as this species is used for afforestation due to its fast growth and the quality of its timber. However, natural populations of this species are distributed throughout the Mediterranean basin, with six major groups: Continental France, Corsica, northwest Spain, southeast Spain, Morocco and Tunisia (Bucci *et al.* 2007). Contrasting temperatures, irradiances, soil composition and water availability occur across the geographic areas of these six major groups,

Correspondence: Mónica Meijón, Fax: (+34) 985104777; E-mail: meijonmonica@uniovi.es and Luis Valledor, Fax: (+34) 985104777; E-mail: valledorluis@uniovi.es
[1]Present address: Plant Physiology, University of Oviedo, Catedrático Rodrigo Uría, 33006 Oviedo, Asturias, Spain

this variation being reflected in molecular markers and quantitative traits that correspond with the different ecological conditions where each population is found (González-Martínez *et al.* 2002).

In this current context of global change, it has become essential to analyse and understand the adaptation mechanisms of plants to improve management practices, especially the key species that maintain our ecosystems and forests (West *et al.* 2009). The identification of several components of multigene regulatory networks that regulate complex processes, such as adaptation, is difficult using traditional genetic screening, which typically only identifies single regulatory elements. Recent new approaches using natural variation in plants, such as genomewide association studies (GWAS), have become an obvious choice for studying the genetics of natural variation and traits of agricultural importance (Huang *et al.* 2010; Meijón *et al.* 2014). Despite being useful in model species, this approach is difficult to apply in unsequenced trees, such as *P. pinaster*, mainly due to their large genome, and the difficulties to establish common garden experiments with forest species, implying long-term studies, high time-consumption and costs. However, current genomic tools are allowing genome-wide studies in nonmodel organism, and recent theoretical advances can help to design research strategies that combine genomics and field experiments to examine the genetics of local adaptation (Savolainen *et al.* 2013). Furthermore, these approaches only provide information about genome variations, and phenotype plasticity is a complex process that involves not only the genome, but also the proteome and metabolome. Metabolites change over the course of evolutionary time: new protein-coding genes and gene families emerge and diversify, ultimately affecting the metabolome and the organism's phenotype and interactions with its environment (Cañas *et al.* 2015). For instance, for many temperate and boreal tree species, cold acclimation exhibits genetic variation in timing and extent along latitudinal gradients, reflecting local adaptation. Working with Sitka spruce, Dauwe *et al.* (2012) have documented substantial variation in the expression of genes with putative roles in seasonal temperature-related adaptation between populations along latitudinal gradients. Many of these transcripts appear to be involved in different aspects of both primary metabolism (e.g. carbohydrate and lipid metabolism) and secondary metabolism (e.g. anthocyanin metabolism), suggesting that changes in metabolism during cold acclimation may reflect genetic adaptation to local climate conditions. Metabolomics techniques are an important diagnostic tool for verifying the physiological responses of plant species to environmental changes and offer the opportunity to gain insight into the mechanisms behind the functioning of biodiversity–ecosystem relationships (Scherling *et al.* 2010; Eckert *et al.* 2012).

Natural variation in the metabolome is primarily associated with changes in secondary metabolism (Routaboul *et al.* 2012) as these metabolites are usually responsive to environmental stresses such as extremes of temperature or salinity, water availability or high light intensity, and as such define plant adaptive response, together with growth regulators. However, despite their tremendous importance for plants, most secondary metabolites remain poorly characterized. The exceptions to this are the flavonoid and terpenoid compounds, whose activity as adaptation signals has been intensively studied in model species and which have been found to perform a key role at the interface between plant and environment.

Flavonoids occur widely in plants and are a biologically and chemically diverse group that can be divided into various subgroups: anthocyanidins, flavonols, flavones, flavonones, chalcones, dihydrochalcones, isoflavonoids, pterocarpanes and dihydroflanonols. Their multifunctionality and diversity is evidence of the importance of flavonoids in plants: frost hardiness and drought resistance have been attributed to flavonoids or other phenolic compounds which are involved in the functioning of the cell wall and membranes (Tattini *et al.* 2004), along with photoprotection, which has been shown to be afforded by high quercetin levels in *Ligustrum vulgare* (Tattini *et al.* 2004; Agati *et al.* 2011). To a large extent, the compounds of this group contribute to the adaptation process of plants to environment changes such as cold, high aridity or irradiation, by serving an antioxidant function (Tattini *et al.* 2004; Treutter 2006; Doerfler *et al.* 2013, 2014).

The terpenoids are produced by a variety of plants and in particular conifers. Plants emit these volatile compounds that play a major role in the interaction between plants and their environment through their key role as insect attractants or repellents and/or biotic stress protectants in pathogen resistance. The best-known and most studied groups of terpenoids are the plant hormones abscisic acid and strigolactone. Both flavonoids and terpenoids are important in plant development and defence against biotic and abiotic stress such as the regulation of drought tolerance (Alder *et al.* 2012), as well as in plant adaptation to variability in environment conditions (Brewer *et al.* 2009; Ha *et al.* 2014).

In this study, we combined a common garden experiment with two complementary metabolomic techniques, gas chromatography coupled to mass spectrometry (GC-MS) and ultra-performance liquid chromatography coupled to high-resolution Fourier transformation mass spectrometry (UPLC-Orbitrap-MS), in order to achieve

a higher coverage of the metabolome than that provided by classic metabolome profiling with GC-MS alone (Doerfler *et al.* 2013). For our metabolomic study, we worked with ten different natural provenances of *P. pinaster* from France, the northwest and southeast of Spain, and Morocco (covering all six genetic pools of the Mediterranean basin, as described by Bucci *et al.* 2007) grown from seeds in a common garden design in order to analyse the sole contribution of genetic effects on the metabolome related to adaptation processes.

Material and methods

Plant material and growth conditions

Plant material was sampled from ten autochthonous populations of *P. pinaster* across the Mediterranean (Fig. 1a) basin, corresponding to different climatic environments in relation to water availability, temperature and irradiation (Table 1). The populations selected were the following: seven Spanish populations representing various climatic conditions from mountain ranges in the northwest (CDVO, ARMY, PTOV and SCRI) to the sandy areas of central Spain (ASPE, COCA), and the coastal region of southeast Spain (ORIA), two Northern French populations (MIMI and PLEU) and one Moroccan provenance (TAMR) (Fig. 1a, Table 1). For each provenance, seeds were collected from five mother trees separated from each other by at least 50 m to minimize inbreeding. Thus, fifty open-pollinated families were available for the experiments. Seedlings of the 10 provenances × 5 open family × 25 replicates of each family were grown in a randomized block design in the field in the northwest Spain (Experimental Station 'La Mata', Grado-Asturias: 43°32′N 7°00′W, 65 m altitude range, 1100 mm annual rainfall, 12.5 °C annual temperature, 1.1 aridity index, 3032.654 Wh/m^2/day direct normal irradiation) (the genetic structures of the provenances studied and

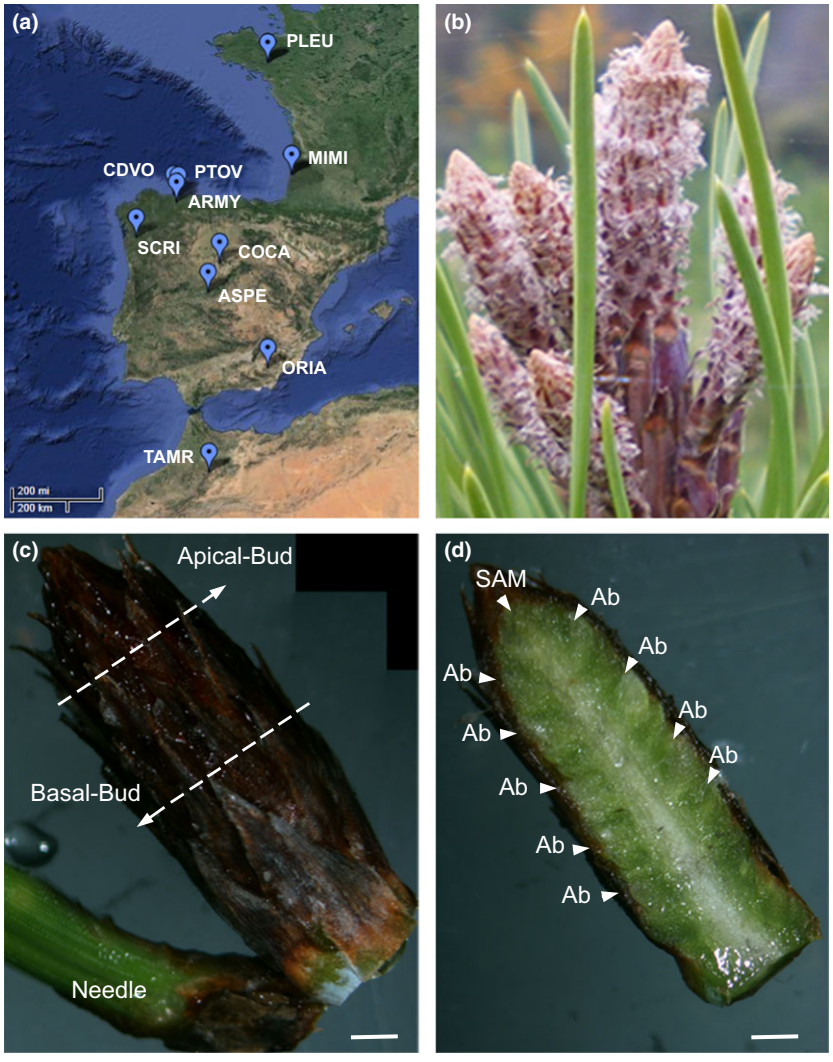


Fig. 1 (a) Localization of *Pinus pinaster* provenances. (b) Stage 1 of apical bud, budburst. (c) Apical bud under stereoscopy. (d) Transversal section of apical bud. Abbreviations: SAM, shoot apical meristem; Ab, axillary bud. Bars = 1 mm.

details of common garden design are described in detail in Gaspar *et al.* 2013). Plants were established in the field for 5 years and annual height (total growth) and number of whorls (polycyclism frequency) were measured at the end of the summer corresponding to the end of the fifth year of growth (September of 2013). From 2010 to 2012, phenological bud development was analysed quarterly (data not shown) to determine the best period for future sampling in terms of homogeneity in plant development. In 2013, three trees in the same phenological phase (at the beginning of active growth, budburst of spring) per provenance were randomly selected from the common garden trying to sample the maximal number of families. Shoot apical buds in the budburst developmental phase (stage 1, Fig. 1b), and mature needles (previous year growth) from the topmost whorl of each tree were sampled on the same day (22 April 2013) between 11.00 and 13.00 hours. Samples were immediately frozen in liquid nitrogen and kept at −80° C until metabolite extraction.

Metabolite extraction

Buds (Fig. 1c) were dissected, separating apical and basal sections according to Fig. 1d. Additionally, needles from last whorl were collected. About 30 mg (fresh weight) of each type of tissue under study (needles and apical and basal sections of bud) was ground in liquid nitrogen. Metabolite extraction was performed according to Valledor *et al.* (2014). Briefly, 600 μL of cold (4 °C) metabolite extraction solution [methanol:chloroform:H_2O (2.5:1:0.5)] was added to each tube and strongly vortexed. Then, the tubes were centrifuged at 20 000 ***g*** for 4 min at 4 °C. The precipitate was discarded, and the supernatant containing metabolites from each tube was transferred to a new tube, and 800 μL of chloroform:water (1:1) was added to allow phase separation. The tubes were vortexed and then centrifuged at 20 000 ***g*** for 4 min at 4 °C. After centrifugation, two layers formed: an upper, aqueous layer (methanol:water), containing the polar metabolites, and a lower organic layer (chloroform), containing the nonpolar. Each fraction was then transferred to separate microcentrifuge tubes and dried using a speed vac at a 25 °C.

LC-Orbitrap-MS analysis

The polar fraction of each sample was analysed twice on LC-Orbitrap-MS, first using the positive ion mode and then the negative. Recordings from both the DAD detector and the high-resolution mass spectrometer were monitored and saved to check system functioning and evaluate later results.

For the high-performance liquid chromatography component, a Dionex Ultimate 3000 (ThermoFisher Scientific, USA) was used. The LC-Orbitrap-MS system (controlled by Xcalibur version 2.2, Thermo Fisher Corporation) was run in gradient mode using a 150 × 2.1 mm 3 μm Hypersil Gold reverse-phase column (ThermoFisher Scientific), set at 30 °C. Solvent A was acetonitrile and solvent B was water containing 0.1% acetic acid. Both mobile phases A and B were filtrated and degassed for 10 min in an ultrasonic bath prior to use. Gradient elution chromatography was performed starting with 10% A and 90% B and held 5 min. Within time interval (5–20 min), A% composition was increased to 90%. This composition was then maintained for 5 min. after which the system was 5 min equilibrated to initial conditions (10% A and 90% B). A flow rate of 0.3 mL/min was used. MS and MS/MS were performed using an LTQ Orbitrap XL high-resolution mass spectrometer (ThermoFisher Scientific) equipped with a HESI II (heated electrospray ionization) source. The high-resolution mass spectrometer was operated in full-scan mode with a resolution of 60 000. Full-scan spectra were acquired for mass range m/z 50–1000 in the positive mode and 65–1000 in the negative mode. The resolution and sensitivity of the Orbitrap were controlled by the injection of a mixed standard after the analysis of each batch (30 samples), and resolution was also checked with the aid of lock masses (phthalates). Blanks were also analysed during the sequence. Additionally, samples were reanalysed (MS/MS) under the same conditions, but selecting the Top three parent ions of each scan. Parent ions with a minimum intensity of 500 were fragmented by CID (normalized collision energy of 35, activation Q 0.25 and activation time 90 ms) and analysed in ITMS.

Chromatogram alignment and metabolite quantification

The raw data of LC-Orbitrap-MS were processed and compared using MZmine software (Pluskal *et al.* 2010). MS1 spectra were filtered establishing noise threshold at $2E^{04}$ and minimum peak height at 2.5 E^{05}. Peaks were smoothed and deconvoluted using a local minimum search algorithm (95% chromatographic threshold, minimum retention range 0.2 min, minimum relative height of 5% and minimum ratio top/edge of 0.5). Chromatograms were aligned using the RANSAC algorithm with a tolerance of 5 ppm of m/z and 0.2 min of retention time. Normalized peak areas were used for quantification, and their values were log-transformed before statistical analyses (Data S1a, Supporting information). MS/MS were only used for metabolite identification purposes.

Table 1 Origin, climatic and growth details of the provenances studied. Climatic data are taken from Sánchez-Gómez *et al.* (2010), irradiation data from European Climate Assessment & Dataset, and data for annual growth were obtained during 2012 for the samples analysed (see Material and Methods)

Provenance	Origin	Latitude and longitude	Altitude range (m)	Annual rainfall (mm)	Annual Temperature (°C)	Aridity index	Direct Normal Irradiance (DNI) (Wh/ m^2/day)	Polycyclism (number of whorls)	Annual Growth (cm)
Armayan (ARMY)	Spain	43°18′N 6°29′W	532	1160	11.8	0	3084.615	4.668	44.558
Mimizan (MIMI)	France	44°12′N 1° 13′W	20	935	13.2	0	4316.923	4.586	45.654
Pleucaeuc (PLEU)	France	47°47′N 2° 20°W	80	855	12.0	0	3386.153	5.106	48.650
Cadavedo (CDVC)	Spain	43°32′ N 6° 25′W	180	1316	13.2	00.8	3366.923	5.278	48.014
Pto. de Vega (PTOV)	Spain	43°33′N 6°38′W	98	1283	13.4	04.9	3373.846	4.806	48.774
San Cipriano (SCRI)	Spain	42°07′N 8°22′W	727	1600	12.3	15.8	4260.000	4.640	47.690
Coca (COCA)	Spain	41°15′N 4°30′W	780	454	12.3	51.4	5143.076	4.310	37.396
A. de San Pedro (ASPE)	Spain	40°12′N 5°3′W	728	1318	14.2	69.4	5417.692	4.728	40.910
Tamrabta (TAMR)	Morroco	33°36′N 5°01′W	1760	763	10.8	67.9	5565.384	3.980	36.424
Oria (ORIA)	Spain	37°30′N 2°20′W	1150	357	13.1	110.0	5935.384	4.450	44.594

Metabolite identification

The individual peaks were identified following different approaches; the first step was performed against our in-house library (>100 compounds) and manual annotation considering m/z and retention times (Fig. S1, Supporting information). In a second step, masses were assigned using the KEGG (Kanehisa & Goto 2000) and ChemSpider (Williams 2008) databases with a built-in M/Z mine utility with a 5-ppm threshold. Our final approach was to use the metlin (Tautenhahn *et al.* 2012) and MassBank (Horai *et al.* 2010) databases, searching for each MS and MS/MS, if available, individually. We considered as undoubtedly 'identified' those metabolites that were defined after the comparison with our standard compound library or by a matching of MS/MS to the small number of plant compounds for which their MS/MS is available in public databases, and 'tentatively assigned' metabolites to those molecular ions with exact masses corresponding to identified metabolites in databases. Metabolite identification against our library was confirmed by RT, mass, isotopic pattern and ring double-bound parameters. Data S2 (Supporting information) includes the detailed interpretation of experimental MS/MS spectra which support our tentative identifications of those masses referred across the manuscript that were not undoubtedly identified in first term. Data S3a (Supporting information) shows detailed peak identification information.

GC-MS analysis

Nonpolar metabolites were methylesterified with 295 μL tert-methyl-butyl-ether (MTBE) and 5 μL of trimethylsulfoniumhydroxide (TMSH) for 30 min at room temperature. Tubes were centrifuged for 3 min at 20 000 ***g*** to remove insoluble particles before transferring supernatant to GC-microvials.

GC-MS measurements were taken following a previously developed procedure (Furuhashi *et al.* 2012) on a triple quad instrument (TSQ Quantum GC; Thermo). In brief, one 1 μL of sample was injected, and GC separation was performed on a HP-5MS capillary column (30 m × 0.25 mm × 0.25 mm) (Agilent Technologies). Oven temperature was increased from 80 °C to 200 °C at 3 °C per min and then reduced to 25 °C at 10 °C per min and maintained at 25 °C for 3 min. Post run conditions were maintained at 30 °C for 4 min. The mass spectrometer was operated in electron impact (EI) mode at 70 eV in a scan range of *m*/*z* 40–600. Metabolites were identified based on their mass spectral characteristics and GC retention times through comparison with the retention times of reference compounds in an in-house reference library and current version of Gölm Metabolome Database (Hummel *et al.* 2007) using LC-Quant software (Data S1b, S3b, Supporting information).

Statistical and bioinformatics analysis

The quantification of the metabolites was performed after filtering out those peaks that were not consistent in terms of representativeness (mass/RT present in at least three samples of any provenance; each sample was split and randomly queued into two LC-MS run batches. Average peak area of the two technical replicates was considered). This step was followed by a log-

transformation to decrease the CV% by reducing the mean:SD dependence. Also, peak areas corresponding to each metabolite were normalized according to the total peak areas in the sample. Neutral masses obtained in positive and negative mode were evaluated to avoid duplicities (same retention time and neutral mass in the different modes), retaining the most intense peaks.

Normalized areas were scaled and subjected to principal component analysis (PCA), sparse partial least squares (sPLS) and heatmap-clustering analyses following the recommendations given by Valledor & Jorrín (2011) for processing omics data sets. Additionally, a regression tree analysis (Random Forest) was performed. One-way ANOVAs for comparing the accumulation of each metabolite in provenances with differential aridity were performed considering the mean metabolite levels for each provenance as experimental units. The statistical procedures described above were performed using the software R v2.12 Core functions, plus SeqKNN, doBy, pheatmap, randomForest and mixomics packages. Network topology was defined after applying sPLS regression using the environmental and morphological variables shown in Table 1 as responses. sPLS algorithm was employed to see correlations between predictor (metabolites matrix) and response variables (environment and growth). Plotting individual scores allowed the definition of provenances matrices. All variables were retained during sPLS analyses, but only those correlations that were statistically significant were used to plot networks. Furthermore, only those edges showing correlation values > |0.6| were then visualized and nodes with no edges above the threshold to other nodes in the network were removed from the analysis. Networks were plotted using the built-in function of the mixOmics package. In addition, a regression tree-based approach was conducted using the Random Forest algorithm. We simulated 5000 trees and applied both regression (quantitative variables) and classification (qualitative variables), employing metabolites as the predictive matrix and each of the environmental growth variables (Table 1) as response vectors.

Results

Needles and apical and basal sections of bud show differential metabolite profiles

To evaluate which tissue is the most appropriate for the analysis of the metabolomics of regulatory networks involved in the adaptation process of *P. pinaster*, three different tissues were tested: needle, bud basal and bud apical, sampling three different individuals from each of the 10 provenances under study, all randomly chosen from the common garden. The applied workflow combining methanol:chloroform:H_2O extraction coupled to LC-Orbitrap-MS and GC-MS, allowed, after peak filtering, the determination of 2403 masses (2386 polar and 17 nonpolar) (Data S1a, b, Supporting information). The combination of a customized algorithm-based library and public databases resulted in the confident identification of 107 metabolites and the tentative assignment of 841 molecular ion masses (Data S3a, b, Supporting information). This approach, combining ionization modes both for LC-Orbitrap-MS and GC-MS, gave a broad characterization of the metabolomes of the pine buds and needles that covered both primary and secondary metabolism pathways. The quality of the data sets produced enabled univariate and multivariate analyses to distinguish between organs and the different provenances.

Principal component analysis multivariate analysis of polar (Fig. 2a, Data S4a, Supporting information) and nonpolar (Fig. S2, Data S4b, Supporting information) metabolites made it possible to distinguish and group samples belonging to the three types of organs. However, PCA of polar metabolites (Fig. 2a) allowed a clearer clustering of the different organs, with most of the metabolites with higher loading in components 1 and 2 of the PCA corresponding to primary metabolism (Table 2). Needle and bud apical tissues presented the highest number of characteristic polar metabolites (565 and 526, respectively) compared to the 147 found specifically in bud basal tissues (Fig. 2b).

Previous studies of bud development in *P. pinaster* have shown that different reserves are accumulated within different tissues of the bud depending on what organogenetic activity is occurring there (Jordy 2004). These observations are in line with the data obtained from the metabolome and the qualitative differences in metabolites found in the tissues analysed in this work. For example, one of the characteristic metabolites of apical bud tissue was tentatively assigned as succinic acid (N1665), which is an intermediate in the citrate cycle, whereas needle tissue contains compounds from other metabolic pathways, as suggested by the characteristic molecular ions corresponding to the mass of peonidin 3-O-glucoside (P1986), involved in pigment biosynthesis, or 4-methylthio-2-oxobutanoic acid (P1195), related to 2-oxocarboxylic acid metabolism. In contrast, tissue from the basal section of the bud was high in metabolites related to amino acid metabolism, such as L-ornithine (P1621), L-Methionine (P268) and vanillic acid (P290) and, as in apical tissue, hormone signalling, such as zeatin (P1335). The majority of these compounds are found among the metabolites with the highest loading in the two principal components (PC1

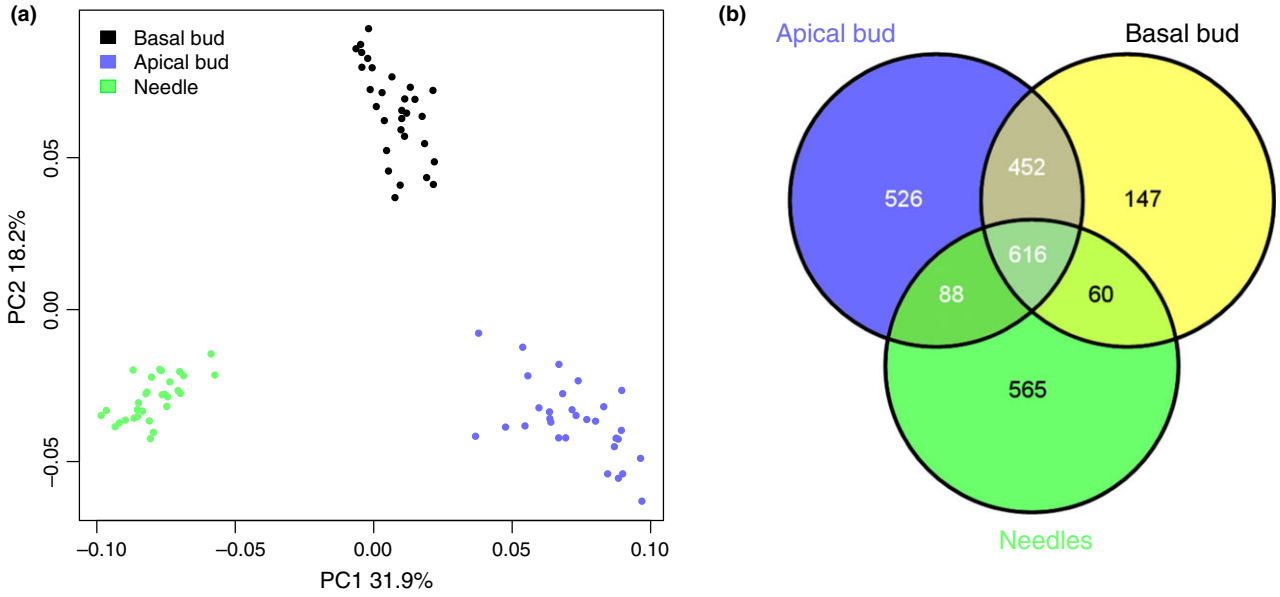


Fig. 2 Untargeted analysis of polar metabolites in needles, and apical and basal sections of the bud in all provenances studied (2386 polar metabolites identified: 90 unequivocally identified, 841 tentatively assigned and 1455 unknown). (a) Classification of the different samples according to principal components analysis (PCA). PC1 and PC2 allowed clustering of the three tissues analysed. (b) Venn diagram showing the total number of polar metabolites detected and the distribution across the three organs analysed.

Table 2 The 20 polar metabolites (methanol soluble) with higher loadings over first two components of PCA. Metabolites in bold were unequivocally identified after the comparison with our compound library or by comparison of the MS/MS to online databases, while the rest were assigned after comparing it very accurate mass to reference compound databases

ID	Assigned to/Identified	Nature of the compound/Pathway related
P299	p-Hydroxyphenylacetothiohydroximate	Glucosinolate biosynthesis (secondary metabolism)
N160	Musca-aurin-VII	Amino acid-related compounds (primary metabolism)
N1665	Succinic acid	Citrate cycle (primary metabolism)
P61	Unknown	Unknown
P202	2-(isopropylamino)ethanol	Ethanolamide
P54	3-Buten-1-1-amine	Amide
N271	2-C-Methyl-D-erythritol 4-phosphate	Terpenoid biosynthesis (secondary metabolism)
P290	**Vanillic acid**	**Amino acid-related compounds (primary metabolism)**
P413	Magnoshinin	Lignan (primary metabolism)
N164	Unknown	Unknown
P268	L-Methionine	Amino acid-related compounds (primary metabolism)
P166	Unknown	Unknown
P319	L-Tryptophan	Amino acid-related compounds (primary metabolism)
P488	L-Proline	Amino acid-related compounds (primary metabolism)
P1195	4-Methylthio-2-oxobutanoic acid	2-Oxocarboxylic acid metabolism (primary metabolism)
P249	**L-Aspartate**	**Amino acid-related compounds (primary metabolism)**
N197	Gambiriin A1	Flavonoids (secondary metabolism)
P605	Cyclohexanecarboxamide, N-[2-[4-(2-methoxyphenyl)-1-piperazinyl]ethyl]-N-2-pyridinyl-	Amide
P1335	**Zeatin**	**Plant hormone (secondary metabolism)**
P499	Unknown	Unknown

and PC2) that can be used to define a molecular signature allowing the distinction of the three analysed organs (Table 2).

This work represents a breakthrough in the description of the metabolome in different *P. pinaster* provenances and organs providing high-resolution data and

quantification. Complete data sets are available as supplementary information with the aim being the base of future reanalysis or experiments for getting new valuable information.

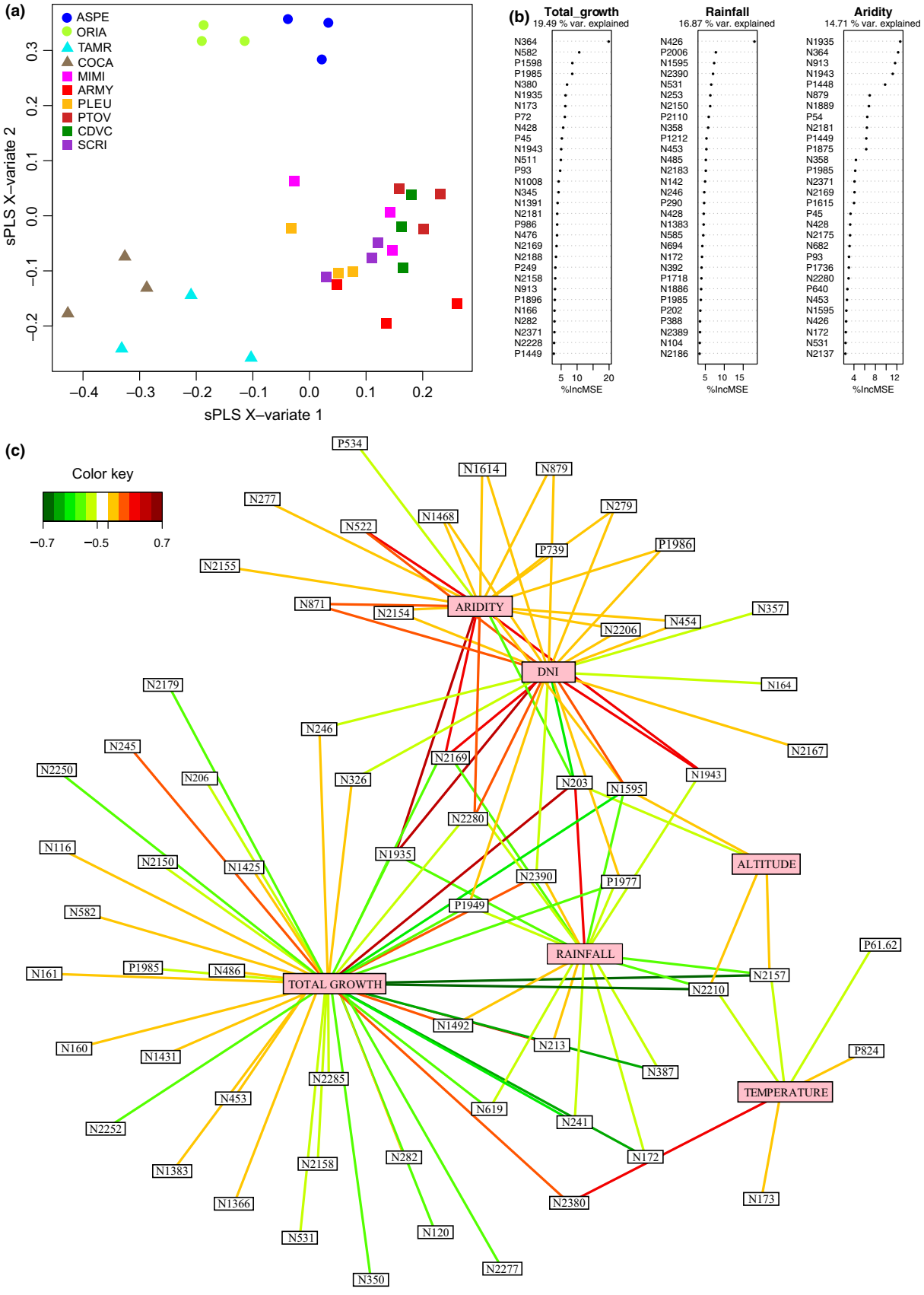
Natural variation through the study of the metabolome profile of P. pinaster

To reduce the dimensionality of the results and integrate metabolome data with that relating to morphological and environmental factors, sPLS analysis was used (Data S4c–e, Supporting information). The different organs were studied individually to reduce noise, given that each metabolome has proved to be different to the other two. The sPLS models showed that the ten provenances under study could be clustered into two large groups, defined by the scores of each population to sPLS components 1 and 2: one containing the provenances from highly arid areas (aridity > 50) (TAMR, ASPE, ORIA, COCA) (Mediterranean provenances) and another comprising the provenances from areas with low aridity (aridity <50) (ARMY, CDVO, MIMI, PLEU, PTOV, SCRI) (Atlantic provenances) (aridity data, based on Sánchez-Gómez *et al.* (2010), are in Table 1). In addition, although all three organs showed a similar clustering pattern in relation to aridity in the geographic origin of the provenances (Fig. 3a, Fig. S3a and S4a, Supporting information), it was in fact data from needles (Fig. 3c) and the basal section of the bud (Fig. S3c, Supporting information) which presented the most identified metabolite nodes. Indeed, needle data (Fig. 3a) even permitted the high aridity group to be subdivided into moderate aridity (COCA, TAMR, with 50 < aridity<69) and extreme aridity (ORIA, ASPE with aridity >69).

In this context, possible genetic effects on the metabolome related to specific local adaptations of the different provenances were then studied using the sPLS test, which allowed us to associate individual compounds with the specific environmental conditions in the site of origin of each provenance. Plotting the sPLS model correlations of metabolites and environmental and physiological data (Fig. S5, Supporting information) showed that polycyclism and temperature appeared in the same sector of the plot for all developmental stages, indicating a similar correlation pattern. Conversely, total growth and rainfall were always located in the opposite sector to aridity, altitude and irradiation (DNI), highlighting their opposing influences on the metabolome. Additionally, a regression tree analysis (Random Forest algorithm) was performed in order to see the extent of variance related to the geographic/climatic origin of the provenance or phenotypic expression which could be explained by differences in metabolites. This analysis showed that total growth, rainfall and aridity were the three variables that were best explained by the changes observed in the needle metabolome (Fig. 3b). Furthermore, in the other two organs studied, geographic and phenotypic parameters showed differential importance in explaining the variance on metabolome (Fig S3b and S4b, Supporting information), which may be closely related to the function of the organ in the plant, as discussed below.

The interaction networks built for needle data (Fig. 3c) indicated a correlation between key metabolites and the different growth-related traits of the provenances and environmental conditions in each provenance's geographic origin. For instance, molecular ions with masses corresponding to sorgolactone (N1935), 2,3-dihydroxy-p-cumate (P1949), tarennoside (N1595), acrylic acid (N203) and malvidin 3-(6″-p-coumaroyl glucoside) (N2280) were all correlated not only with growth rate, but also with rainfall, aridity and irradiation (DNI), suggesting different adaptive strategies depending on environment of the plant's origin (arid or nonarid) as it had already been observed by Krasensky & Jonak (2012). Furthermore, all these metabolites showed different accumulation levels (P-value < 0.05) between arid and nonarid provenances, all increasing in arid provenances, except acrylic acid, which decreased (Fig. 4 a–f). Most of these compounds are associated with the degradation pathway of aromatic compounds or are flavonoids and/or terpenoids. The most important terpenoid found was sorgolactone, a plant growth regulator belonging to the group known as strigolactones (SLs), which are related to branching

Fig. 3 Multivariable analysis of polar and nonpolar metabolites of needles. **(**a) Sparse partial least squares (sPLS) analysis. Different colours show populations analysed while squares represent provenances from areas with low aridity (aridity <50), triangles moderate aridity (50 < aridity < 69) and circles extreme aridity (with aridity > 69). (b) Random Forest analysis. Only total growth, rainfall and aridity, the three variables that were best explained by this model, are shown. ID numbers represent ID of metabolites shown in data 2–3 and %IncMSE indicates how mean squared error increased when this variable was randomly permuted with other variable of the data set (larger %IncMSE indicates greater importance of the variable to the model). (c) Interaction networks between different environmental variables—aridity, altitude, temperature, rainfall and direct normal irradiation (DNI) in shaded boxes—and the metabolites analysed (ID numbers represent ID of metabolites shown in data sets 1–3) in white boxes. Edge colour represents the correlation between metabolites and environmental growth variables. This plot was constructed after sPLS analysis using differential metabolites between provenances.



and apical dominance (Brewer *et al.* 2009; Ward *et al.* 2013). In regression tree analysis, sorgolactone (N1935) was also detected as one of the most significant metabolites in explaining variance related to aridity and total growth parameters (Fig 3b).

Furthermore, in the tissues from the basal section of the bud, several nodes of metabolites that correlated with irradiation, aridity, altitude and rainfall (Fig. S3c, Supporting information) were tentatively assigned to brusatol (N652) and isobrucein A (N654), and unequivocally identified as taxifolin (dihydroquercetin) (N150), and these also demonstrated significant differences between arid and nonarid provenances (Fig. 4 g–i). As with the needle tissue, most of the compounds are either terpenoid or flavonoid in nature, with taxifolin, the central element in the flavonoid biosynthesis pathway (Winkel-Shirley 2001), being that which presented the highest difference in levels between arid and nonarid provenances. In the regression tree analysis of the basal bud tissue, taxifolin (N150) was found to be one

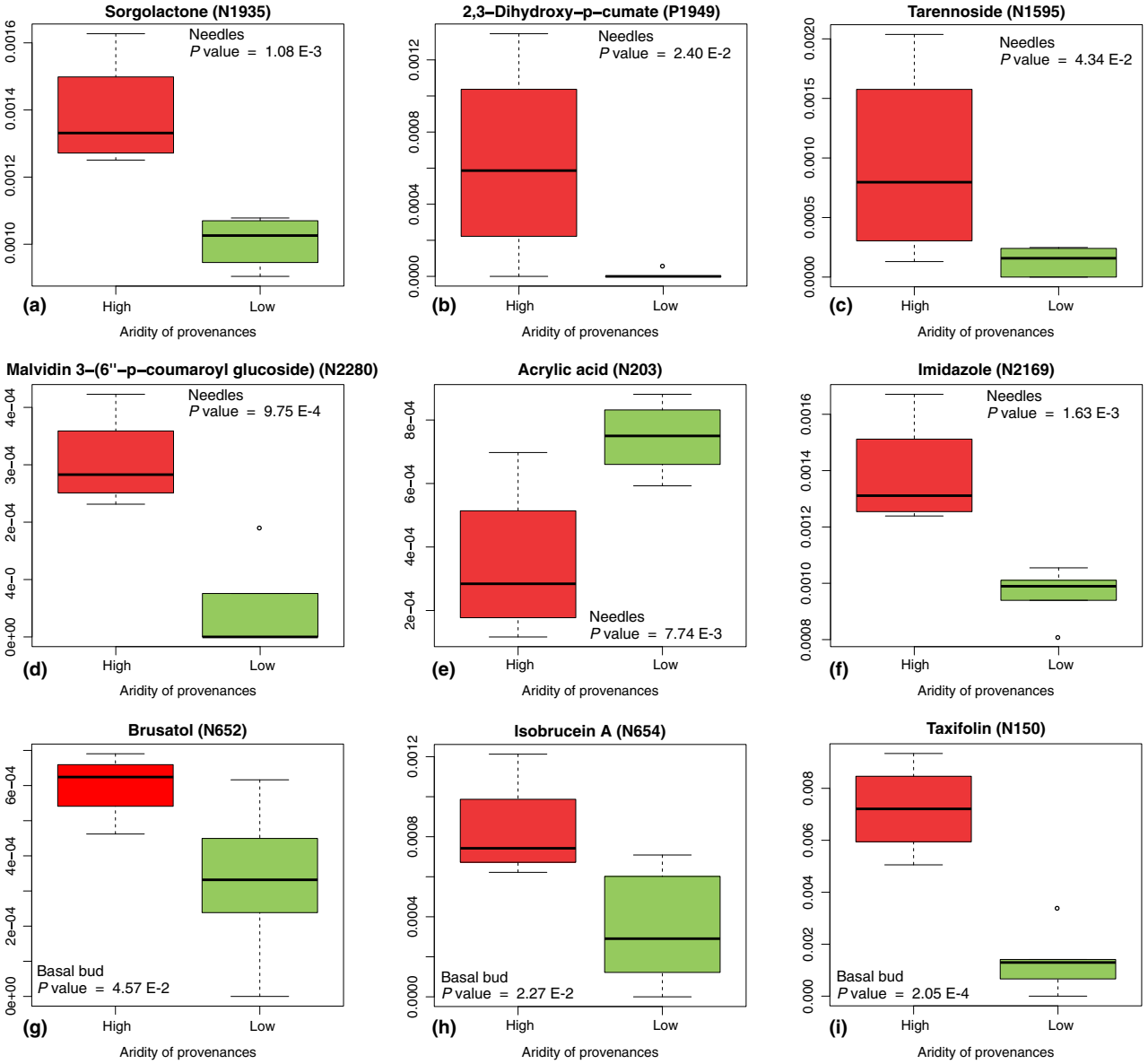


Fig. 4 Levels of selected metabolites showing nodes in sPLS networks from needles (a–f) and basal bud data (g-i). Box plot representation and one-way ANOVA employing 2 levels of aridity as a factor aridity <50 and aridity >50. Boxes at the right side (green in color on-line version) represent provenances characterized by a low aridity (ARMY, CDVC, MIMI, PLEU, PTOV, SCRI), while boxes at left (red in color on-line version) represent provenances of arid environments (ASPE, COCA, ORIA, TMBR). Experimental units used in the ANOVAS were the mean metabolite levels for each provenance.

of the most important metabolites in terms of explaining the variance of growth between provenances (Fig S3b, Supporting information) — arid provenances showed less annual growth and lower frequency of polycyclism than nonarid provenances (growth data shown in Table 1).

In the final series of analyses, when metabolome data from the apical section of the bud were integrated with the morphological and environmental data (Fig. S4, Supporting information), a very high number of metabolites were positively correlated with the altitude of each provenance's geographic origin. Regression tree analysis also showed altitude to be one of the best variables to explain variance in apical bud metabolomes (Fig S4b, Supporting information). Whilst most of the metabolites characterized were unknown compounds, some were identified, for example buddledin-A (P820), a terpenoid, and columbamine (N918), an isoquinoline alkaloid element, and were correlated positively with altitude and negatively with total growth. These results support, once again, the important role of secondary metabolism in the growth and resistance response of *P. pinaster* to light and drought. Through the Random Forests approach (Fig S3b, Supporting information), buddledin-A (P820) and compounds similar to columbamine, such as dehydrocorydaline (N927), another isoquinoline alkaloid element, and isoflavone like N992, were found to be key metabolites with respect to variance related to altitude.

The metabolomic pathways with most relevance in the differentiation between arid and nonarid provenances

The complexity of the metabolome data from the three developmental stages and 10 provenances analysed was reduced by studying specific pathways in relation to the relative abundance of the metabolites identified (Fig. S6, Supporting information). In this way, it was possible to show how, in each organ, specific metabolic pathways are overexpressed and inhibited. For example, the biosynthesis of amino acids in the apical section of bud and hormone metabolism in the basal section are enriched while in needles both pathways are inhibited. In contrast, in needles flavonoid biosynthesis is more highly expressed than in basal or apical sections of the bud.

Moreover, comparing the data on the basis of provenance, in needles, only the flavonoid and anthocyanin biosynthesis and carbon metabolism pathways showed differences between arid (Mediterranean provenances) and nonarid origins (Atlantic provenances) (Fig. 5). In contrast, in the basal section of the bud, differences between arid and nonarid provenances were observed in relation to the terpenoid, plant hormone and phenylpropanoids biosynthesis pathways, as well as for flavonoid and carotenoid biosynthesis. However, as mentioned in a previous section, when single metabolites were analysed rather than complete pathways, the clustering of the populations demonstrated differences between Mediterranean and Atlantic provenances with far higher clarity (Fig. 4).

On analysing the flavonoid biosynthesis pathway in needles, the organ with highest number of identified metabolites, and the relative abundance of these compounds in arid and nonarid provenances (Fig. 6), it was found that, in general, flavonoids (especially flavones, flavonols and anthocyanins) were overaccumulated in arid provenances. However, some flavonoid precursors such as 4-coumaroylquinic acid, and specific flavonoids such as hesperetin and the previously mentioned dihydroquercetin, along with stilbenoids (pinosylvin) connected with phenylpropanoid biosynthesis, were also found to be increased in nonarid provenances.

Discussion

Changes in plant metabolism are the key to plant developmental processes, underpinning many of the ways in which plants respond to the environment. An understanding of how much variation in the metabolome is influenced by inherent genetic factors and/or by external environmental factors is a crucial area of study in relation to the application of metabolomics research. In fact, the emergence of environmental metabolomics as one of the latest omics technologies has been one of the most important recent developments in plant physiology (Scherling *et al.* 2010). Its applications span the entire landscape of plant ecology, facilitating the understanding of various processes, from plant plasticity and adaptation through to community composition and even genetic modifications in crops (Scherling *et al.* 2010; Brunetti *et al.* 2013). However, to date hardly any studies of this type have been performed in nonmodel species, and still less in conifers.

In this work, we aimed to ascertain whether the different *P. pinaster* provenances can be characterized from their metabolomes, and if so, which metabolites are responsible for this differentiation, and whether certain metabolites are correlated with different growth capabilities and geoclimatic conditions of origin. Seeking to attain maximum coverage of the metabolome, we used a combination of LC-Orbitrap-MS and GC-MS, thereby putatively identifying 948 (107 identified, 841 tentatively assigned) and 1455 unknown metabolites of *P. pinaster*. As the metabolome is dynamic, we sampled three plants from each of ten populations all grown under the same common garden conditions from seed, where samples were all taken on the same day to minimize

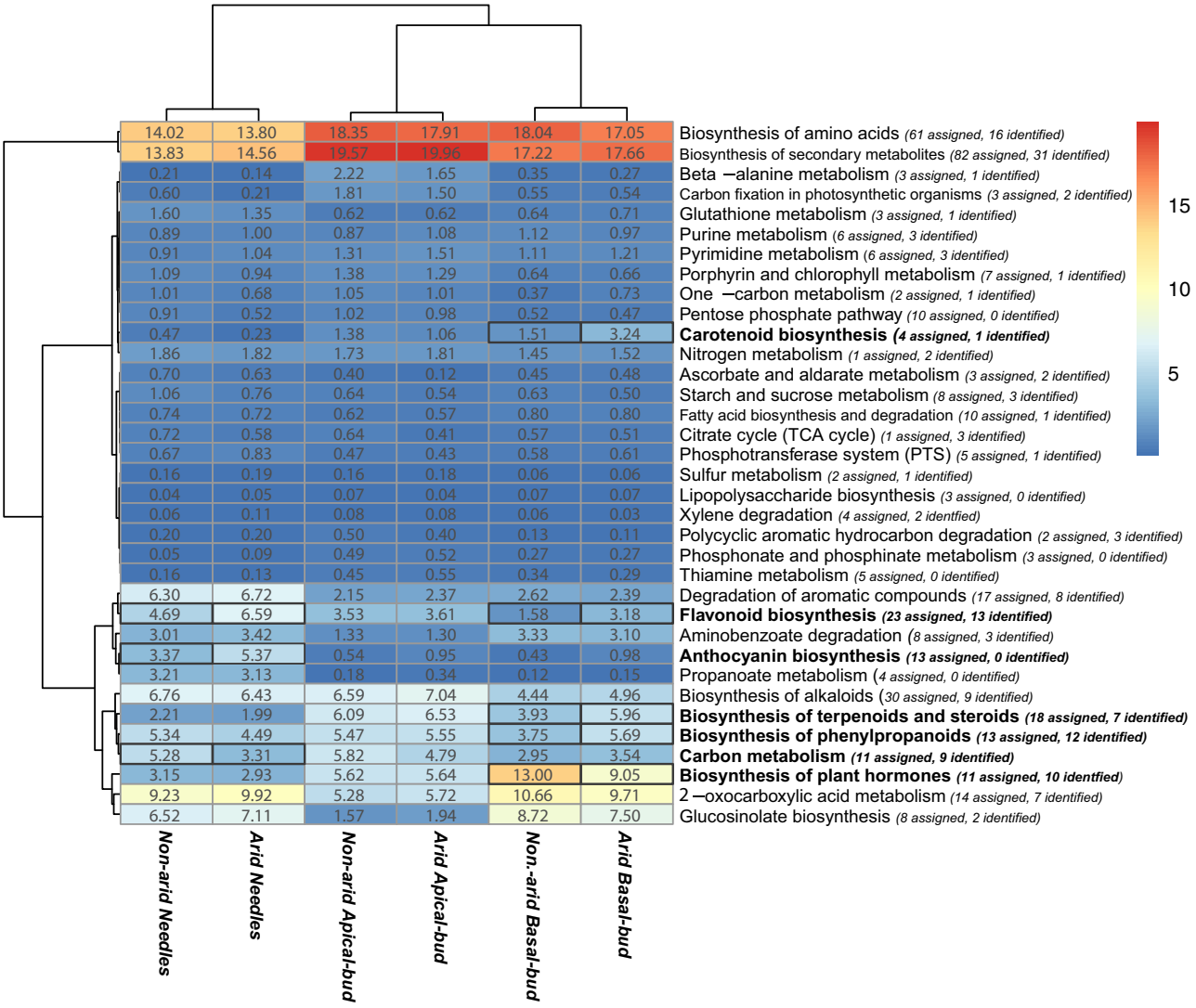


Fig. 5 Heatmap-clustering analyses of pathways identified through the metabolome of apical and basal section of bud and needles. Arid represents data from the following populations: ORIA. TAMR, ASPE, COCA; nonarid data are from: CADV, ARMY, MIMI, PLEU, PTVO, SCRI. Bold letters and rectangles superimposed on the plot indicate the changes mentioned in the text. Numbers inside of the boxes indicate normalized abundance of each pathway calculated as the sum of all identified/assigned metabolites within each pathway according to KEGG pathway.

environmental influences on the metabolome. In addition, we assayed three different organs: needles and apical and basal sections of the apical bud, to test metabolome stability and the suitability of our approach. The combination of LC-Orbitrap-MS and GC-MS, together with the bioinformatics tools used for metabolite identification and data mining, proved to be highly efficient, enabling us to show that needles and apical and basal sections of the bud present different metabolome profiles, even in the closely related and proximal tissues of the bud base and apex and that the tissue-specific metabolomes are fairly constant between individuals from the same provenance. While the analysis of the nonpolar metabolites made it possible to cluster the metabolites on the basis of whether they derived from needles and apical or basal sections of the bud, it was the analysis of polar metabolites which showed the best clustering performance, probably explained by the fact that polar fractions contain primary and secondary metabolites including primary metabolites related to the biosynthesis of amino acids (metabolites with higher loading in components 1 and 2 of the PCA, summarized in Table 2), while nonpolar samples are almost exclusively composed of fatty acids.

The number and type of metabolites present in different tissues are clearly linked to the specific organogenic

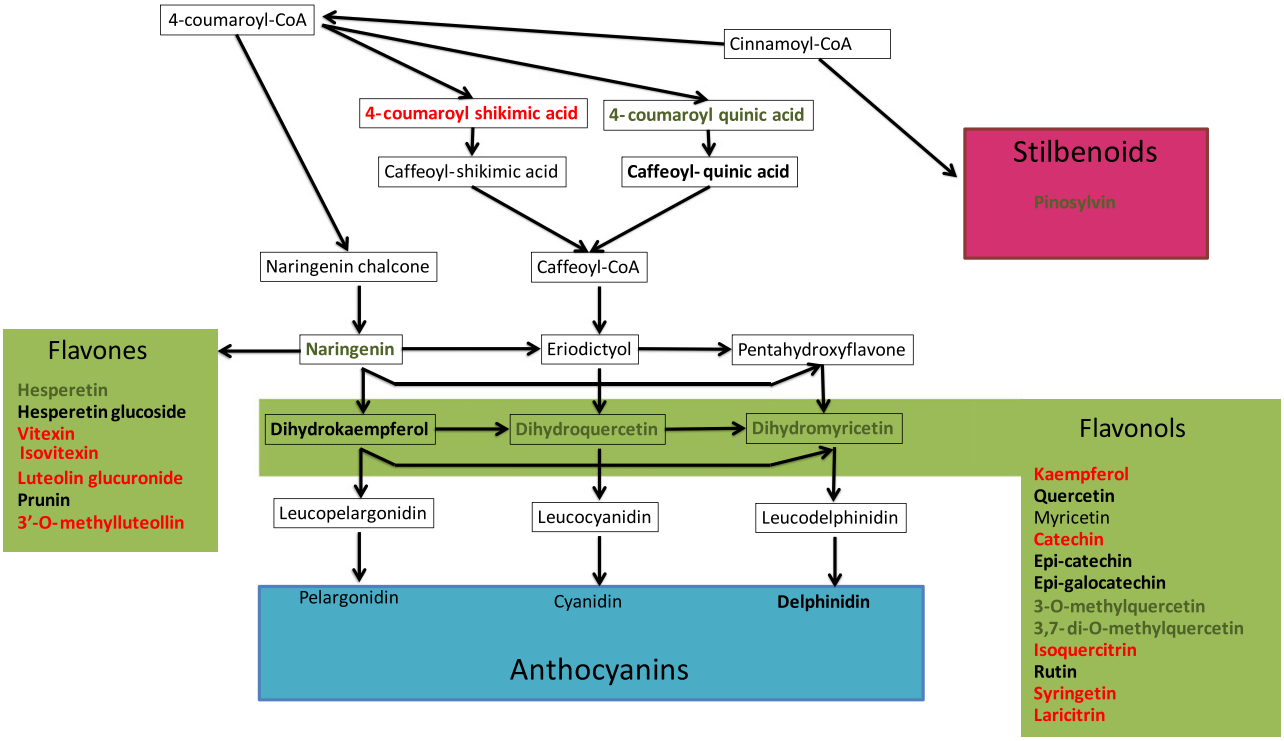


Fig. 6 Summary of flavonoid pathways. Words in bold indicate unique metabolites identified in needles: red represents high levels in populations from arid areas (ORIA, TAMR, ASPE, COCA), and green represents high levels in populations from nonarid areas (CADV, ARMY, MIMI, PLEU, PTVO, SCRI).

activity occurring in each situation: that is, needles and the basal section of the bud (where axillary meristems are located) represent organs involved in cell production, while the apical section of the bud focuses on the accumulation of starch grains (Jordy 2004) and maintenance of the stem cells in the rib meristem. For correct growth, the plant must maintain a constant flow of cells through the meristem, where the input of dividing pluripotent stem cells offsets the output of differentiating cells. This flow depends both on extracellular signalling within the shoot apical meristem, governed by a spatial regulatory feedback loop that maintains a reservoir of stem cells, and on factors that prevent meristem cells from differentiating prematurely (Carles & Fletcher 2003). This is reflected in the data obtained for polar metabolism, where the apical section presented the highest number of metabolites (1682), supporting the idea of the special complexity of the regulatory mechanisms in shoot apical meristems. Needles, a specialized organ for photosynthesis, have the second highest total number of metabolites of the three tissues studied (1329) and the highest number of differential metabolites compared with the other developmental stages (565) (Fig. 2b). Needles are involved in multiple cellular processes; however, their main function is to provide energy to the plant through the absorption of light at different wavelengths (through photosynthetic pigments such as chlorophylls) to initiate photosynthesis. The basal section of the bud is the tissue with the least differential metabolites (147) and this accords with its function as an intermediate state between needles and the apical section of bud, it being where primordial needles start to emerge from axillary buds (Jordy 2004). This tissue thus presents, on the one hand, meristematic cells like those of the apical section of the bud, and on the other, cells that are beginning to differentiate in order to carry out photosynthetic activity, resulting in very high metabolomic and hormone activity (Jordy 2004), comparable to that of needles.

Authors such as Pien *et al.* (2001) have shown that during early leaf development, carbohydrate metabolism is spatially regulated within the shoot apical bud and that this asymmetry is related to hormone biosynthesis in the apex. However, it is not only primary metabolism that changes during apical bud development, as secondary metabolism is also altered. For example, changes in levels of compounds like flavonols and stilbenes have been observed during needle development in *Picea abies* that differed between provenances (Slimestad 1998).

Overall, these analyses did not show a substantial number of outliers or random behaviour of the metabo-

lome within each organ (Fig. 2a), proving the suitability of the method used for the comparison of provenances, reflected in the low coefficient of variation of metabolites and the low noise of the data within each organ (in Data S4a, Supporting information, standard deviation of components 1 and 2 of polar PCA, respectively, 0.063 and 0.048 is shown). Ultimately, it is clear that using the metabolome profiles to analyse natural variation could provide different results depending on which organ is analysed. However, while the analysis of the different organs is not of great value in itself, in combination with the data related to the conditions in the area of provenance, it provides a comprehensive picture of the natural variation in this species.

The high variability of the natural environment presents great challenges for plants in terms of their ability to grow, compete with neighbouring plants, and respond appropriately to different abiotic and biotic pressures (Brunetti *et al.* 2013). Phenotypic plasticity enables plants to withstand such environmental dynamism, in both the short and long term, and is governed by genes which determine not only the characteristics of an organism but also the degree of responsiveness of each characteristic to environmental stimuli. However, it is the metabolome that gives each organism its specific ecophysiological properties enabling it to interact with the physical environment and with other species within its ecosystem. Numerous population genetics studies have been performed to analyse quantitative traits and natural variation in *P. pinaster* and other plant species (e.g. González-Martínez *et al.* 2002; Atwell *et al.* 2010), but few have been conducted using metabolite profiling. Our results in this study demonstrate that metabolite profiling could be a powerful tool for the assessment of the metabolic phenotype of a population in response to tree diversity and ecophysiological adjustment.

The use of the sPLS algorithm (Lê Cao & Boitard S Besse 2011), a step beyond simply reducing the dimensionality of the data, allowed us to build strong interaction networks between physiological and environmental data and the metabolome. By analysing the needle metabolome, we were able to cluster three types of *P. pinaster* populations with respect to the aridity of their area of origin (Fig. 3 a–b): one group with the populations from northwest Spain and France (ARMY, CDVO, MMI, PLEU, PTOV), areas with little or no aridity; a second group with the populations from central Spain and northern Morocco (COCA, TAMR) with medium aridity; and a third group with those from the extremely arid areas of central and southeastern Spain (ARMY and ORIA). Although it was not possible to distinguish these three groups considering the metabolomes of the apical or basal section of the bud, using these data sets did allow the differential clustering of the populations from the northwest Spain and France (Atlantic, aridity <50) as against those from the other regions (Mediterranean, aridity >50). These data are in agreement with the observations of Gaspar *et al.* (2013) who showed significant differences ($\alpha<0.0001$) in the survival rates between the provenances from Atlantic and Mediterranean regions in PEG-induced osmotic stress experiments. Additional genetic studies by chloroplast microsatellite markers (Fig. S7, Supporting information) have clearly identified three main gene pools of *P. pinaster* in Europe: an 'Atlantic gene pool', which includes all the northern areas of the Iberian Peninsula and France, a 'central gene pool' covering southeastern Spain, and a 'southeastern gene pool', including Italy, Corsica, Sardinia and Northern Africa (Bucci *et al.* 2007).

These genetic mappings clearly overlap with the three groups of populations defined through metabolome profiling in the current work, supporting the notion that secondary metabolism, and in particular flavonoids and terpenoids, is connected to light and drought tolerance. This is exemplified by compounds such as malvidin 3,4-p-coumaroyl glucoside (anthocyanin) (N2280) and taxifolin (dihydroquercetin) (N150), central elements in flavonoid biosynthesis (Fig. 6) (Winkel-Shirley 2001), whose levels are extremely high in populations from arid zones (southern Spain and Morocco) and very low in the populations from northwest Spain and France (nonarid) (Fig. 4d,i). This increase could possibly be related to the antioxidant function of malvidin 3,4-p-coumaroyl glucoside and taxifolin and the need for additional protection from the light and drought in high aridity locations. Other flavonol compounds, like quercetin 3-O-glucoside or cyaniding 3-O-glucoside (crucial compounds in anthocyanin biosynthesis), have also been demonstrated to exhibit high antioxidant activity, thus enhancing plant stress tolerance (Nakabayashi *et al.* 2014). It is further assumed by several authors that anthocyanins have a crucial role in the photoprotection process, providing antioxidant functions in response to excess light (Treutter 2006; Valledor *et al.* 2012; Doerfler *et al.* 2013, 2014). Additionally, in Sitka spruce, Dauwe *et al.* (2012) have documented substantial variation in anthocyanin metabolism and flavonoid biosynthesis in seasonal temperature-related adaptation between populations along latitudinal gradients, suggesting that changes in metabolism during cold acclimation may reflect genetic adaptation to local climate. Regression tree analyses showed that, when organs are analysed independently, most of the metabolome variance between provenances was related to the geographic and climatic conditions of the origin of the provenance, confirming that the aridity

of the origin region is an essential parameter to explain the variance of the metabolome (Fig 3b). A detailed comparison of the data showed that the percentage of variance explained by each parameter was dependent on the studied organ (Fig S2b and S3b, Supporting information). This is not surprising as the metabolome of the different organs changes considerably to allow their very different functions in the plant. The great majority of the compounds used in the sPLS analysis to cluster the samples were confirmed by regression tree approach to be key in explaining the different aspects of variance that related to aridity and growth in needles and that corresponding to altitude in apical and basal sections of the bud. Furthermore, additional compounds of a flavonoid nature were detected in the regression tree analysis of the apical section of the bud that were also related to altitude of the region origin of the provenance.

Plants are estimated to contain more than 9000 flavonoids, representing a significant proportion of the total number of metabolites in plants. These compounds are involved in many physiological mechanisms such as flower or fruit colour, light protection and interactions between plants and microbes, animals or other plants (Routaboul *et al.* 2012). Our results suggest that flavonoid overaccumulation could be associated with drought tolerance in *P. pinaster* (Fig. 6) as was already suggested by Nakabayashi *et al.* (2014). However, those provenances from the more extreme arid areas do not always show higher levels of flavonoids such as malvidin 3,4-p-coumaroyl glucoside and taxifolin (dihydroquercetin), and it is in fact the provenances from medium aridity zones which present the highest values of these compounds. This could be attributed to the fact that flavonoids are also increased in response to light and cold stress (Treutter 2006; Routaboul *et al.* 2012). In general, these types of stress induce the accumulation of reactive oxygen species (ROSs). ROSs play an important role as signalling molecules that initiate stress responses in plants. The increased generation of ROS in the cellular components is attributed to a disturbance in cellular homeostasis, such as programmed cell death or death by cytotoxicity (Nakabayashi *et al.* 2014). Additionally, regression tree analysis of the basal section of the bud revealed taxifolin (dihydroquercetin) to be one of the most relevant metabolites as far as explaining the variance in growth between provenances was concerned (Fig S3b, Supporting information). Growth data (Table 1) showed that provenances from arid areas were characterized by lower annual growth and reduced frequency of polycyclism compared to nonarid provenances. In accordance with these data, it has been recently demonstrated that some flavonoids, besides leading to enhanced oxidative and drought tolerance in plants (e.g. Nakabayashi *et al.* 2014), can also act as an endogenous polar auxin transport inhibitor in the shoots of Arabidopsis (Yin *et al.* 2013), thereby playing a crucial role in plant growth. The variety and complexity of flavonoid functions suggest that plants use the each compound in response to particular forms of oxidative stress (Onkokesung *et al.* 2014), which also affects the development of the plant (Yin *et al.* 2013).

Terpenoids and carotenoids also feature strongly in this present work, in both metabolite network and pathway analyses. In the clustering of arid and nonarid provenances, sorgolactone (a plant hormone of terpenoid structure related to carotenoid biosynthesis and belonging to the SL group) is a key metabolite. This is probably connected with the differential growth models which have been found for arid and nonarid provenances, as sorgolactone is directly related to apical growth and inhibition of branching (Brewer *et al.* 2009). In *P. pinaster,* polycyclic growth (multiple shoot flushes in a single season) is common and shows high family and individual heritability (Zas & Fernández-López 2005), as well as being clearly connected with adaptation to water availability. Trees that show a high level of polycyclism usually present increased branching and lack of apical dominance, while nonpolycyclic trees present strong apical dominance and less branching (Rozas *et al.* 2011; Cline & Harrington 2007). These data point to the likelihood that the regulation of this trait is linked to the control of shoot apical meristem development and drought tolerance adaptation. Recently, studies in Arabidopsis, like that of Ha *et al.* (2014), have shown that exogenous SL treatment rescues the drought-sensitive phenotype of SL-deficient mutants, confirming the additional role of SL as a positive regulator of the drought stress response.

Conclusion

This work describes clear correlations between the metabolome and the geoclimatic environment of the origin of *P. pinaster* provenances. The common conditions under which the seeds of these provenances were propagated suggest that these differences are potentially linked to genetic factors related to local adaptation of the species. These data provide a valuable step forward in our understanding of the diversity and adaptive processes of *P. pinaster*. The robustness of metabolome data within developmental stages, achieved through the combination of novel mass spectrometry technologies and bioinformatics tools, allowed us to explore the variation between provenances. Clustering of the provenances proved possible on the basis of their metabolites in relation to the geoclimatic conditions of the origin of the population, aridity of the provenance being identi-

fied as the key factor. Our results also show the differential regulation of primary metabolism during organ development and also point to genome specialization, particularly in those pathways related to flavonoids and terpenoids metabolism, both of which play a role in improving stress resistance in arid provenances and also affect growth and branching capabilities. The application of multivariate regression tests allowed the definition of the specific metabolites involved in explaining the different growth and environmental variables considered. Despite further studies on *P. pinaster* being necessary to link genetic and environmental conditions, growth models and flavonoids and terpenoids variation, this work provides a first step in the emerging area of metabolomics research in relation to natural variation studies. However, we are still far from having a complete metabolic description of natural variation in plant compounds.

Acknowledgements

The authors wish to especially thank Wolfgang Busch, for critically reading the manuscript and to Ronnie Lendrum for the English language review. We express our gratitude to five anonymous reviewers for constructive and helpful comments and suggestions. This research was supported by funds from INIA and FEDER cofunding through the projects: RTA 2010-00120-C32-01, RTA 2013-00048-C03-02. MO was supported by the projects M200871201 (Academy of Sciences of the Czech Republic) and LO1415 (Ministry of Education, Youth and Sports of the Czech Republic). LV was supported by MEYS CR Re. CZ.1.07/2.3.00/20.0256 and the Spanish Government through the Juan de la Cierva Programme (JCI-2012-12444). MM was funded by the Juan de la Cierva Programme (JCI-2011-08958) (Spanish Ministry of Economy and Competitiveness).

References

Agati A, Biricolti S, Guidi L, Ferrini F, Fini A, Tattini M (2011) The biosynthesis of flavonoids is enhanced similarly by UV radiation and root zone salinity in *L vulgare* leaves. *Journal of Plant Physiology*, **168**, 204–212.

Alder A, Jamil M, Marzorati M *et al.* (2012) The path from β-Carotene to Carlactone, a strigolactone-like plant hormone. *Science*, **335**, 1348–1351.

Atwell S, Huang YS, Vilhjálmsson BJ *et al.* (2010) Genome-wide association study of 107 phenotypes in a common set of Arabidopsis thaliana inbred lines. *Nature*, **465**, 627–631.

Brewer PB, Dun EA, Ferguson BJ, Rameau C, Beveridge CA (2009) Strigolactone acts downstream of auxin to regulate bud outgrowth in Pea and Arabiodopsis. *Plant Physiology*, **150**, 482–493.

Brunetti C, George RM, Tattini M, Field K, Davely MP (2013) Metabolics in plant environmental physiology. *Journal Experimental Botany*, **64**, 4011–4020.

Bucci G, González-Martínez SC, Le Provost G *et al.* (2007) Range-wide phylogeography and gene zones in *Pinus pinaster* Ait. revealed by chloroplast microsatellite markers. *Molecular Ecology*, **16**, 2137–2153.

Cañas RA, Canales J, Muñoz-Hernández C *et al.* (2015) Understanding developmental and adaptive cues in pine through metabolite profiling and co-expression network analysis. *Journal of Experimental Botany*, **63**, 6431–6444.

Carles CC, Fletcher JC (2003) Shoot apical meristem maintenance: the art of a dynamic balance. *Trends in Plant Science*, **8**, 394–401.

Cline MG, Harrington CA (2007) Apical dominance and apical control in multiple flushing of temperature woody species. *Canadian Journal of Forest Research*, **37**, 74–83.

Dauwe R, Holliday JA, Aitken SN, Mansfield SD (2012) Metabolic dynamics during autumn cold acclimation within and among populations of Sitka spruce (*Picea sitchensis*). *New Phytologist*, **194**, 192–205.

Doerfler H, Lyon D, Nägele T *et al.* (2013) Granger causality in integrated GC–MS and LC–MS metabolomics data reveals the interface of primary and secondary metabolism. *Metabolomics*, **9**, 564–574.

Doerfler H, Sun X, Wang L, Engelmeier D, Lyon D, Weckwerth W (2014) mzGroup analyzer-predicting pathways and novel chemical structures from untargeted high-throughput metabolomics data. *PLoS ONE*, **9**, e96188.

Eckert AJ, Wegrzyn JL, Cumbie WP *et al.* (2012) Association genetics of the loblolly pine (*Pinus taeda*, *Pinaceae*) metabolome. *New Phytologist*, **193**, 890–902.

Furuhashi T, Fragner L, Furuhashi K, Valledor L, Sun X, Weckwerth W (2012) Metabolite changes with induction of *Cuscuta haustorium* and translocation from host plants. *Journal of Plant Interactions*, **7**, 84–93.

Gaspar MJ, Velasco T, Feito I, Alía R, Majada J (2013) Genetic variation of drought tolerance in *Pinus pinaster* at three hierarchical levels: a comparison of induced osmotic stress and field testing. *PLoS ONE*, **8**, e79094.

González-Martínez SC, Alía R, Gil L (2002) Population genetic structure in a Mediterranean pine (*Pinuspinaster*Ait.): a comparison of allozyme markers and quantitative traits. *Heredity*, **89**, 199–206.

Ha CV, Leyva-González MA, Osakabe Y *et al.* (2014) Positive regulatory role of strigolactone in plant responses to drought and salt stress. *PNAS*, **111**, 851–856.

Horai H, Arita M, Kanaya S *et al.* (2010) MassBank: a public repository for sharing mass spectral data for life sciences. *Journal of Mass Spectrometry*, **45**, 703–714.

Huang X, Wei X, Sang T *et al.* (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics*, **42**, 961–967.

Hummel J, Selbig J, Walther D, Kopka J (2007) The Golm Metabolome Database: a database for GC-MS based metabolite profiling. In: *Metabolomics* (eds Nielsen J, Jewett MC), (pp. 75–95). Springer, Berlin, Heidelberg.

Jordy M-N (2004) Seasonal variation of organogenetic activity and reserves allocation in the shoot apex of *Pinus pinaster* Ait. *Annals of Botany*, **9**, 25–37.

Kanehisa M, Goto S (2000) KEGG: encyclopedia of genes and genomes. *Nucleic Acids*, **28**, 27–30.

Krasensky J, Jonak C (2012) Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. *Journal of Experimental Botany*, **63**, 1593–1608.

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

Meijón M, Satbhai SB, Tsuchimatsu T, Busch B (2014) Genome-wide association study using cellular traits identifies a new regulator of root development in Arabidopsis. *Nature Genetics*, **46**, 77–81.

Nakabayashi R, Yonekura-Sakakibara K, Urano K *et al.* (2014) Enhancement of oxidative and drought tolerance in Arabidopsis by overaccumulation of antioxidant flavonoids. *The Plant Journal*, **77**, 367–379.

Onkokesung N, Reichelt M, van Doorn A, Schuurink RC, van Loon JJA, Dicke M (2014) Modulation of flavonoid metabolites in *Arabidopsis thaliana* through overexpression of the MYB75 transcription factor: role of kaempferol-3,7-dirhamnoside in resistance to the specialist insect herbibore *Pieris brassicae*. *Journal Experimental Botany*, **65**, 2203–2217.

Pien S, Wyrzykowska J, Fleming AJ (2001) Novel marker genes for early leaf development indicate spatial regulation of carbohydrate metabolism within the apical meristem. *The Plant Journal*, **25**, 663–674.

Pluskal T, Castillo S, Villar-Briones A, Orešič M (2010) MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics*, **11**, 395.

Routaboul J-M, Dubos C, Beck G *et al.* (2012) Metabolite profiling and quantitative genetics of natural variation for flavonoids in Arabidopsis. *Journal Experimental Botany*, **63**, 3749–3764.

Rozas V, Ignacio García-González I, Zas R (2011) Climatic control of intra-annual wood density fluctuations of *Pinus pinaster* in NW Spain. *Tree*, **25**, 443–453.

Sánchez-Gómez D, Majada J, Alía R, Feito I, Aranda I (2010) Intraspecific variation in growth and allocation patterns in seedlings of *Pinus pinaster* Ait. submitted to contrasting watering regimes: can water availability explain regional variation? *Annals of Forest Science*, **67**, 505–512.

Savolainen O, Lascoux M, Merila J (2013) Ecological genomics of local adaptation. *Nature Reviews Genetics*, **14**, 807–820.

Scherling C, Roscher C, Giavalisco P, Schulze E-D, Weckwerth W (2010) Metabolomics unravel contrasting effects of biodiversity on the performance of individual plant species. *PLoS ONE*, **5**, e12569.

Slimestad R (1998) Amount of flavonols and stilbenes during needle development of *Picea abies*; variations between provenances. *Biochemical Systematics and Ecology*, **26**, 225–238.

Tattini M, Galardi C, Pinelli P, Massai R, Remorini D, Agati G (2004) Differential accumulation of flavonoids and hydroxycinnamates in leaves of *Ligustrum vulgare* under excess light and drought stress. *New Phytologist*, **163**, 547–561.

Tautenhahn R, Cho K, Uritboonthai W, Zhu Z, Patti G, Siuzdak G (2012) An accelerated workflow for untargeted metabolomics using the METLIN database. *Nature Biotechnology*, **30**, 826–828.

Treutter D (2006) Significance of flavonoids in plant resistance: a review. *Environmental Chemistry Letters*, **4**, 147–157.

Valledor L, Jorrín J (2011) Back to the basics: maximizing the information obtained by quantitative two dimensional gel electrophoresis analyses by an appropriate experimental design and statistical analyses. *Journal of Proteomics*, **74**, 1–18.

Valledor L, Cañal MJ, Pascual J, Rodríguez R, Meijón M (2012) Early induced protein 1 (*PrELIP1*) and other photosynthetic, stress and epigenetic regulation genes are involved in *Pinus radiata* D. don UV-B radiation response. *Physiologia Platarum*, **146**, 308–320.

Valledor L, Escandón M, Meijón M, Cañal MJ, Weckwerth W (2014) An universal protocol for the combined isolation of metabolites, DNA, long RNAs, small RNAs, and proteins from green tissues. *The Plant Journal*, **79**, 173–180.

Ward SP, Salmon J, Hanley SJ, Karp A, Leyser O (2013) Using Arabidopsis to study shoot branching in biomass willow (Salix spp.). *Plant Physiology*, **162**, 800–811.

West JM, Julius SH, Kareiva P *et al.* (2009) U.S. natural resources and climate change: concepts and approaches for management adaptation. *Environmental Management.*, **44**, 1001–1021.

Williams AJ (2008) Public chemical compound databases. *Current opinion in drug discovery & development*, **11**, 393–404.

Winkel-Shirley B (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiology*, **126**, 485–493.

Yin R, Han K, Heller W *et al.* (2013) Kaempferol 3-O-rhamnoside-7-O-rhamnoside is an endogenous flavonol inhibitor of polar auxin transport in Arabidopsis shoots. *New Phytologist*, **201**, 466–475.

Zas R, Fernández-López J (2005) Juvenile genetic parameters and genotypic stability of *Pinus pinaster* Ait. Open-pollinated families under different water and nutrient regimes. *Forest Science*, **51**, 165–174.

M.M. designed and performed the research, analysed the data and wrote the manuscript. I.F. performed research and contributed to the design of the common garden. M.O. contributed in-house LC-MS Library and performed research. C.D. performed research. W.W. contributed to metabolite identification. J.M. contributed to the design of the common garden and the population genetics analysis. L.V. designed and performed the research, analysed the data and wrote the manuscript. All authors read, corrected and approved the final manuscript. The data sets employed in this work are available in Supporting Information.

Data accessibility

Sampling localization and common garden design DOI: 10.1371/journal.pone.0067781. Chloroplast microsatellite markers data doi: 10.1111/j.1365-294X.2007.03275.x. Metabolome data and statistical parameters: enclosed in this submission.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Comparison of retention times of some of the identified analytes and its corresponding standards.

Fig. S2 Untargeted analysis of nonpolar metabolites of needles, apical and basal section of bud according to Principal Components Analysis (PCA) scores of nonpolar metabolomic data-set (17 unequivocally identified against pure standard or full MS/MS data).

Fig. S3 Multivariable analysis of polar and nonpolar metabolites of basal section of bud.

Fig. S4 Multivariable analysis of polar and nonpolar metabolites of apical section of bud.

Fig. S5 Graphical representation of environmental and morphological variables of two-dimensional Sparse Partial Least Squares (sPLS).

Fig. S6 Heatmap-Clustering analyses of pathways identified through the metabolome of apical and basal section of bud, and needles.

Fig. S7 Ward's minimum-variance dendrogram based on the Mahalanobis pairwise distance matrix among the gene zones of *P. pinaster* used in this work and based on microsatellite data from Bucci *et al*. (2007).

Data S1 Peaks obtained after UPLC-MS of polar (methanol soluble) samples (a) and GC-MS of non-polar samples (b).

Data S2 Interpretation of MS2 spectra corresponding to parent m/z 315.12, 637.26, 359.13, 521.20, 519.17, 117.01, 149.02, and 464.25 tentatively identified as Sorgolactone (N1935), Malvidin 3-(6''-p-coumarylglucoside) (N2280), Tarennoside (P1977), Isobrucein A (N654), Brusatol (N652), Succinic Acid (N1665), 4-Methylthio-2-oxobutanoic acid (P1195), and peonidin-3-o-glucoside (P1986), respectively.

Data S3 Identification of polar (methanol soluble) and nonpolar metabolites in the 2403 peaks that were analyzed.

Data S4 Results of multivariate statistical tests. Score and loadings matrices of PCA analyses of polar (a) and nonpolar (b) metabolites considering all tissues.