New Forests
https://doi.org/10.1007/s11056-022-09933-x

Phylogeographic origin authentication of *Araucaria araucana* (Mol.) K Koch seedlings through the application of spectroscopy techniques in different infrared ranges and chemometric methods

Macarena Rojas-Rioseco[1,2,5] · Rosario del P. Castillo[2,5] · Jorge González-Campos[3] · Roberto Ipinza[4] · M. I. Sanhueza[2,5] · Rodrigo Hasbún[1]

Received: 1 October 2021 / Accepted: 6 July 2022

Abstract
The origin of seed and seedlings is an important factor for the success of restoration programs; an inadequate origin can have negative impacts on genetic and adaptive processes. A technique that allows authenticating the origin is infrared spectroscopy, a fast, accurate and low-cost tool. In Patagonia one species that required restoration programs, consequently, propagules traceability is *Araucaria araucana*. Phylogeographic studies showed significant differences between Chilean Andean and Coastal populations. The goal of this study was to discriminate the phylogeographic origin of *A. araucana* seedlings using spectroscopic and chemometric methods. Seedlings of both phylogeographic origins were cultivated in common garden and spectral information in four spectral ranges was recorded. Principal component analysis and soft independent modeling of class analogy (SIMCA) were applied. All the spectral ranges analyzed were able to discriminate phylogeographic origin, whose predictive models achieved a classification accuracy of 88–91%. The best models were SIMCA VIS–NIR and SIMCA FTIR. Wavelengths responsible for discrimination were associated with photosynthetic pigments, proteins and plant fibers. Andean seedlings have a higher content of Chlb, xanthophylls and plant fibers and the most important bands for the Coastal provenance are related to Chla and protein contents. It is shown that differences reported at the genetic level between both origins are expressed at the chemical level. In conclusion, infrared spectra obtained from *Araucaria araucana*, treated with chemometric methods, allow capturing the phylogeographic signal that separates Coastal and Andean origins. In the future, the resulting models could be used in restoration programs for this species.

Keywords Traceability · Conifers · Population differentiation · Plant materials selection · NIR · FTIR

✉ Macarena Rojas-Rioseco
macarrojas@udec.cl

✉ Rodrigo Hasbún
rodrigohasbun@udec.cl

Extended author information available on the last page of the article

Published online: 23 July 2022

Introduction

In the last decades, forest ecosystems have been highly threatened, fragmented and degraded principally by human activity. Nowadays, there is an increasing interest in restoration practices, and a critical decision is the correct choice of propagation material, where the most traditional practice is to use sources of material geographically close to the site to be restored (Broadhurst et al. 2008; Atkinson et al. 2021). The use of local genotypes gives a higher probability of survival, establishment and growth of propagules, and therefore, a higher probability of success of the restoration program. Non-local materials are more likely to suffer from maladaptation, leading to poor establishment, additionally, there is a risk of contamination or loss of local genetic pools, which could have negative destabilizing effects through genetic flow (Broadhurst et al. 2008). In some countries, among them Chile, the propagules used in restoration programs are obtained from forest nurseries, which in most cases provide material whose origin is unknown or untraceable (Atkinson et al. 2021; León-Lobos et al. 2020). It is therefore necessary to have a traceability tool that is able to authenticate or certify the origin of material, to ensure the use of propagules appropriate to the planting site on a scientific, reliable and accurate basis (Gutiérrez-Caro et al. 2012).

Until now, visual inspection (microscopy and macroscopy), molecular markers (DNA and protein-based), Information and Communication Technologies (ICT) and instrumental analysis techniques (spectrometry and chromatography) have been used for identification, discrimination and authentication of plant material (Rohman et al. 2019). However, these techniques fail in the desired characteristics of a traceability tool, such as being fast, easy to apply and relatively inexpensive (Alía et al. 2005; Godbout et al. 2018). An appropriate tool that satisfies these requirements is Infrared (IR) Spectroscopy techniques in combination with chemometric methods, highlighting for authentication, the Near Infrared (NIR) and Mid-Infrared (MIR) Spectroscopy (Rohman et al. 2014, 2019). IR spectroscopy has been widely used in different fields of applied science and proved to be a powerful origin authentication tool in agricultural and forestry products (Blanco and Villarroya 2002; Loewe et al. 2016). This technique allows quantitative and qualitative analysis of the chemical and physical properties of different biological materials, based on the measurement of the vibrational changes of the molecules after the interaction between the sample and IR radiation, which can be in the NIR (14,000–4000 cm^{-1}) or MIR (4000–400 cm^{-1}) regions (Lohumi et al. 2015). A unique absorption spectrum is acquired for each sample, creating a fingerprint. Chemometrics can be used to process the spectra, removing unwanted information such as spectral noise, scattering, among others, as well as to apply unsupervised (e.g. Principal Component Analysis (PCA) and supervised Pattern Recognition Methods (e.g. PLS-DA, SIMCA (Soft Independent Modeling of Class Analogy), KNN, SVM), to explore samples grouping or develop predictive classification models, respectively (Bystrzanowska and Tobiszewski 2020; Marini 2010).

Application of IR spectroscopy has been successful for the capture of the phylogenetic signal from different biological systems, distinguishing between different species, genera, varieties and subvarieties (e. g *Echinacea* species (Neal-Kababick and Flora 2010), *Hibiscus mutabilis* L. and *Berberidis radix* (Fu et al. 2015), *Eucalyptus nitens* and *E. globulus* (Castillo et al. 2008), among others) (Lohumi et al. 2015; Luykx and Ruth 2008; Rohman et al. 2014). Some research has reported the successful discrimination of some forest species according to their geographic origin from wood, seeds and leaves (Farhadi et al. 2017; Loewe et al. 2016; Sandak et al. 2011; Tigabu et al. 2005). Nevertheless, there are only a

few studies where these tools have been applied to discriminate woody species according to their phylogeographic origin. Jianguo et al. (2012) studied *Ulmus elongata* (Ulmaceae) leaves from different regions using FTIR spectroscopy and found variations at the band absorption maxima associated with polysaccharides, fatty acids, lignin and proteins. In this way, the application of IR spectroscopy is a very promising tool for forest authentication and traceability purposes.

In Chile, *Araucaria araucana* (conifer of a high social, cultural and economic value) is a species that requires conservation efforts through restoration programs and, therefore, traceability tools (Ipinza and Müller-Using 2021; Rodríguez et al. 1983). In this country this species is distributed in the Andes Mountains between 37° 30′ and 39° 30′ S latitude (Drake et al. 2009) and in the Coastal Mountains in two small and isolated populations between 37° 20′ and 38° 40′ S latitude (Martín et al. 2014). According to the Red List of the International Union for Conservation of Nature, the populations of both mountain ranges have been categorized as "Endangered" (IUCN 2022). For restoration project planning, it is critical to consider that the structure and distribution patterns of *A. araucana* have been influenced by historical events such as the Last Glacial Maximum (LGM) (Ruiz et al. 2007; Villagrán, 2018; Villagrán and Armesto 2005). Studies based on molecular markers and palynological data revealed significant and historical differences between the Andean and Coastal populations, which would correspond to two Evolutionarily Significant Units (ESU) (Bekessy et al. 2002; Martín et al. 2014; Ruiz et al. 2007). Bekessy et al. (2004) suggest that when restoration activities are carried out on this species, it is very important to avoid mixing the two ESUs, and to use only local germplasm.

IR spectroscopy is a promising tool, which may be able to capture the phylogeographic signal that differentiates both ESUs of *A. araucana*. Therefore, the main aim of this research is to discriminate the phylogeographic origin of *A. araucana* seedlings through the application of spectroscopy techniques in different IR ranges and chemometric methods. The specific objectives were to (1) evaluate the selectivity and discriminatory power of different ranges of the IR region, (2) establish and validate a predictive model for classification of phylogeographic origin, and (3) identify the spectral bands that contribute most to classification to extract information on the main chemical differences between seedlings of different origins. The models resulting from this research could be used in future restoration programs of *A. araucana* to trace the origin of the germplasm, reducing the risk of maladaptation and/or genetic contamination.

Methodology

Plant material (seed and seedling sampling and processing)

The seedlings of *A. araucana* were provided by the Forestry Institute INFOR Chile in the framework of the Integrated Monitoring System of Native Forest Ecosystems (SIMEF) project, in which seeds were sampled from 12 locations of the specie natural distribution (Fig. 1 and Table 1), considering the genetic structure reported for this conifer by Martín et al. (2014). Once the seeds were collected, they were kept in cold storage (4 °C) in the dark. The seeds were planted at the Carlos Douglas nursery located in Yumbel, Biobío Region, owned by the Chilean forestry and paper holding company CMPC. As a pretreatment to stimulate germination, the seed was cut and left to soak for 48 h in distilled water, in a cold chamber at 4 °C. For the nursery process, plastic trays

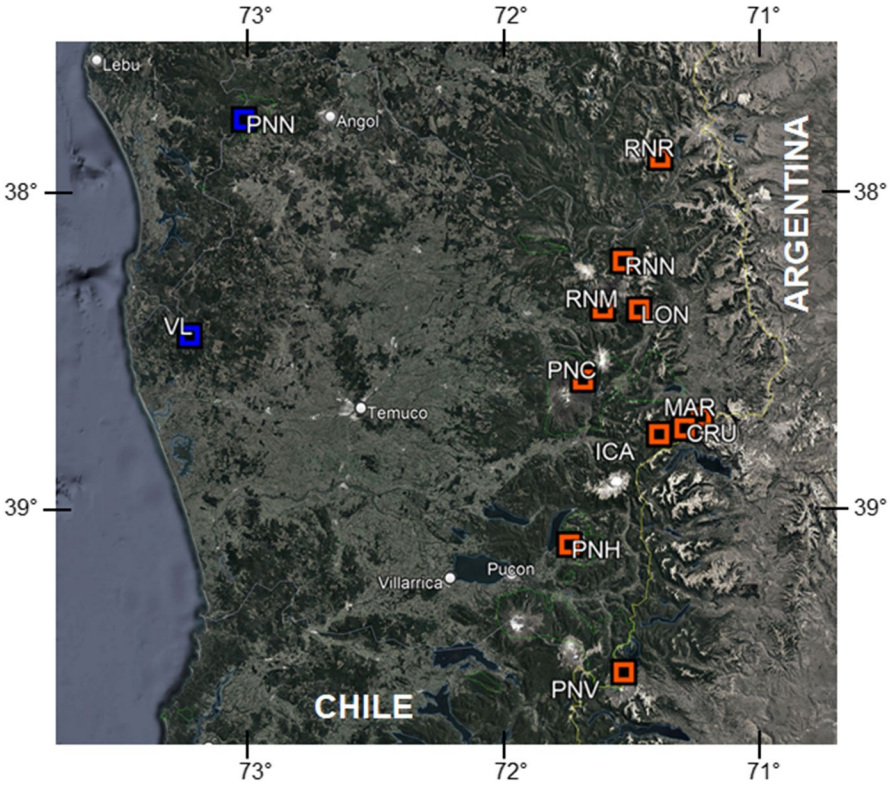

Fig. 1 Seed localities sampled in both ESU of *A. araucana* in seed period 2018. Blue boxes indicate localities in the Cordillera de la costa (Nahuelbuta National Park and VL: Villa las Araucarias). Orange boxes indicate localities in the Andes Mountains (RNR: Ralco National Reserve; RNN: Las Nalcas National Reserve; LON: Lonquimay; RNM: Malalcahuello National Reserve; PNC: Conguillio National Park; ICA: Icalma; MAR: Marimenuco; CRU: Cruzaco; PNH: Huerquehue National Park; PNV: Villarrica National Park). Image obtained from Google Earth

with 125 cc cavities and composted pine bark were used as substrate. Subsequent care was carried out in accordance with the company's quality standards, which consisted of irrigation, temperature curtains at 26 °C, moss removal and fertilization (aminoterra ®, seed mixture fertilizers and horizon ® application). Seedling sampling was carried out 11 months after planting. Nine seedlings were selected per provenance for a total of 108 seedlings. The selected seedlings were placed in paper bags and properly identified with their phylogeographic origin and were transported directly to the laboratory for processing. Seedlings were immediately placed in an oven at 40 °C for seven days until constant weight. Then the material was trimmed with scissors, selecting the upper 2/3 of the seedlings. The pulverization process was carried out with a Retsch MM400 ball mill mixer at a frequency of 30.0 1/s for 3 min. The pulverized material was stored in black flasks in darkness until the acquisition of spectra, in which the pulverized material was used directly (without being diluted or mixed with any solvent), except for the MIR analysis where the potassium bromide (KBr) pelletizing method was used (Sherman 1997).

Table 1 Information about seed sampling points of *Araucaria araucana* within its natural distribution in Chile

Population	Code	Region	Latitude	Longitude	Alt	MAP	MAT	TAR
PN Nahuelbuta	PNN	Coastal	−37.805609	−73.017985	1269	1604	6.1	19.9
Villa las Araucarias	VL	Coastal	−38.495328	−73.254247	664	1501	8.9	19.4
RN Ralco	RNR	Andean	−37.939491	−71.334323	1239	1696	8.8	25.4
RN Las Nalcas	RNN	Andean	−38.269358	−71.489768	976	2219	9.6	24.7
RN Malalcahuello	RNM	Andean	−38.425844	−71.56517	1382	1765	7.7	24.3
Lonquimay	LON	Andean	−38.426427	−71.421637	1376	1632	8.1	24.7
PN Conguillio	PNC	Andean	−38.647372	−71.698783	1236	1860	7.9	23.6
PN Huerquehue	PNH	Andean	−39.172097	−71.707628	1378	1464	6.7	23
Cruzaco	CRU	Andean	−38.800124	−71.235559	1424	1138	7.9	24.7
Icalma	ICA	Andean	−38.819732	−71.332654	1195	1355	8.7	24.4
Marimenuco	MAR	Andean	−38.762456	−71.184824	1401	1110	8.2	24.7
PN Villarica	PNV	Andean	−39.569	−71.514951	1187	1145	7.5	23

NP, National Park; NR, National Reserve; Alt, altitude (m); MAP, mean annual precipitation (mm); MAT, mean annual temperature (°C); TAR, temperature annual range (maximum temperature–minimum temperature; °C). Climate variable means obtained from WorldClim. Table facilitated by McIntosh et al. (Unpublished)

Acquisition of spectra

The processed samples were analyzed in different regions of the electromagnetic spectrum: visible, near IR and mid-IR. For VIS–NIR the USB4000 portable equipment from Oceans Optics (400–1037 nm) with a resolution of 0.2 nm and 5 scans was used, in reflectance mode. For NIR, two instruments were used, Bruker FT-NIR MPA II (12,489–3996 cm^{-1}) with a resolution of 16 cm^{-1} and 32 scans, and the portable Thermo Fisher MicroPHAZIR (1595–2396 nm) with a resolution of 8 nm and 5 scans, both in diffuse reflectance mode. To obtain the absorption spectra, the function A = log (1/R) was used (where A is the absorbance and R the reflectance), except for the FT-NIR spectra, which were derived with the Kubelka–Munk function (Pasquini 2003). Spectral acquisition in mid-IR was performed with the Agilent Cary 630 FTIR (3999–650 cm^{-1}) with a resolution of 4 cm^{-1} and 32 scans, in transmittance mode, for which it was necessary to make KBr pellets for each sample. The samples were homogeneously mixed between 5 and 10 w/w% in KBr in an agate mortar and then pressed at high pressure using a hand-operated press, generating a translucent pellet (Sherman 1997). More spectra were taken for plants in the Coastal populations to compensate for the gap with the number of Andean populations.

Spectral pre-processing

Spectral pre-processing allows improvement of the posterior development of mathematical models and data analysis through transformations and pre-processing (Lerma-García et al. 2018). It is known that Sample characteristics such as irregular particle size can have effects on the properties of the spectra (noise, scattering, baseline shift, among others) that make the subsequent analysis more difficult. Based on the above, mathematical transformations are

Table 2 Preprocessing and transformations applied to the four spectral matrices for PCA and SIMCA

Spectral matrix	PCA		SIMCA	
	Transformation	Pre-processing	Transformation	Pre-processing
VIS–NIR	1st derivative (25)	MC	1st derivative (15)	MC
FT-NIR	MSC	MC	2nd derivative (9)	MC
NIR	MSC	MC	2nd derivative (5)	MC
FTIR	Baseline correct	MC	Normalization	MC
	Normalization		2nd derivative (25)	

The number of the dot window used is indicated between parentheses

MSC, Multiplicative scatter correction; MC, Mean Center

applied to reduce these effects, the most applied being spectral derivatives, baseline correction, smoothing and those whose purpose is to compensate for dispersion such as MSC (Multiplicative Scatter Correction) and SNV (Standard Normal Variate). Pre-processing is also applied to reveal small differences between similar spectra (accentuates differences). The most common are mean-centering (MC) and autoscaling (AU). In this case the four spectral bases were treated with various combinations of transformations and pre-processing using Piruette 4.5 (Infometrix Inc) software. For the purposes of this research, only the results of the best treatments for the respective analyses are presented, corresponding to the treatments that presented the greatest number of correctly classified samples, which are shown in Table 2.

Chemometric analysis

Spectra matrices were analyzed by PCA and SIMCA. The analyses were performed with Piruette 4.5 (Infometrix Inc) software. PCA is an exploratory method that allows dimensionality reduction through the transformation of uncorrelated variables into a linear combination (principal component: PCs) (McGarigal et al. 2000). PCA allowed the identification of atypical samples, as well as to observe the behavior of the samples (clusters). The outliers were eliminated only when at least two different statistical tests (e.g. Mahalanobis distance, Q test, residuals plot, among others) indicate that these samples are outliers. SIMCA is a probabilistic, parametric and supervised method based on the generation of a PCA for each class, independently. A "space for each class" is constructed, whose limits determine whether a sample belongs to that class, with a certain level of confidence (Pomerantsev and Rodionova 2020). For the generation classification models, 70% of the samples corresponding to the training set of the models were used and two classes were stipulated, considering each mountain range or macrozone as a class.

Validation of classification models (SIMCA)

Model validation made it possible to evaluate the performance of the models, estimating the uncertainty of future predictions with an external set. For the validation 30% of the total samples, which were not used in the training set, were used. It was validated by the percentage of correctly classified samples with 95% confidence (Massart et al. 1998). For the evaluation of model performance, sensitivity (Sn), specificity (Sp), classification error rate (ER) and

classification accuracy (CA) were evaluated. Sn describes the ability of the model to correctly recognize the membership of samples in each class. Sp describes the ability of the model to reject samples not belonging to that class. The parameters will be calculated with the equations described by Ballabio and Consonni (2013), which are shown below, where TP corresponds to true positive, TN to true negative, FP to false positive and FN to false negative. In addition, the robustness of the models was evaluated by applying the different calibration and validation sets 3 times.

$$Sn = \frac{TP}{TP + FN}$$

$$Sp = \frac{TN}{FP + \mathrm{T}N}$$

$$ER = 1 - \frac{(Sn + Sp)}{2}$$

$$CA = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

Data interpretation

We identified the bands that contribute most to discrimination by phylogeographic origins of *A. araucana* seedlings. For this purpose, the component loadings in PCA and the power of the discrimination plot in SIMCA were analyzed. Subsequently, a bibliographic search was carried out to identify possible compounds associated with phylogeographic origin. In addition, for FTIR spectra acquired, the mean of the spectra of each class and their respective standard deviation were calculated, and then a second derivative (Savitzky-Golay algorithm with 35-point window and second-order) was applied. The minimum of each spectrum was analyzed, identifying the bands that presented differences between origins and that maintained the same behavior with and without the standard deviations. Afterwards, chemical assignment was carried out through a bibliographic search.

Chlorophyll content measurements

To estimate the chlorophyll content of *A. araucana* seedlings, 3.5 year old plants grown in pots in a common garden located in the Coastal zone (INFOR Concepción Headquarters) (from seeds coming from the same sampling in 2018), of 30 cm height were selected. A total of 12 plants from the Coastal Range and 28 plants from the Andes were analyzed. The SPAD index (Soil Plant Analysis Development) was measured with the portable chlorophyllometer SPAD-502 (Minolta, Spectrun Technologies Inc., Illinois, USA). This index quantitatively evaluates the intensity of leaf green (650 to 940 nm), being proportional to the amount of chlorophyll present in the leaf. To acquire the information, the height of the plants was divided into 3 fractions of 10 cm, collecting readings in the upper, middle and lower areas. All readings were taken in triplicate in each area and then an average per plant was taken. Measurements were taken during the morning 8:00–12:00 am.

SPAD data analysis

Descriptive statistical analysis of the SPAD values obtained for each phylogeographic origin was performed. The assumptions of normality were tested using the Shapiro-Wilks test and homoscedasticity using Bartlett's test, both with ($\alpha = 0.05$). Given that the above assumptions were not fulfilled, the non-parametric Mann–Whitney–Wilcoxon test was applied ($\alpha = 0.01$).

Results

VIS–NIR, FT-NIR, NIR and FTIR spectra

In the different spectral ranges, the absorbance spectra (Fig. 2) of *A. araucana* seedlings from the two phylogeographic origins showed a similar profile with the same absorption maxima. In the VIS–NIR range, absorption maxima were observed at wavelengths 538, 616 and 672 nm, corresponding to green, orange and red spectral regions, respectively (Fig. 2A, B). In the FT-NIR spectra eight bands were observed specifically at wavelengths 6904, 6326, 5793, 5701, 5176, 4690, 4335 and 4258 cm^{-1} (Fig. 2C). The spectra in portable NIR acquired show four bands at wavelengths 1927, 2114, 2277 and 2307 nm (Fig. 2D). Finally, the mid-IR spectra acquired showed that between the zone 2600–2000 cm^{-1} doesn't present bands providing chemical information, and this region was excluded from the analysis. Multiple bands are observed. Bands at wavelengths 3360 and 2800 cm^{-1} can be distinguished, as well as numerous bands in the fingerprint region (1800–500 cm^{-1}) (Fig. 2E).

Discrimination of phylogeographic origin of *A. araucana* seedlings

PCA models were developed to observe the behavior of the samples of *A. araucana* and to identify if it is possible to observe separation of groups according to their phylogeographic origin. The results obtained are shown in Table 3. The percentage of variance explained by the first three components varied according to the spectral range analyzed, obtaining values from 81.64 to 91.52%. In no case was a clear differentiation between classes observed (Fig. 3A, C, E and G), although there were tendencies of separation according to the Andean and Coastal origin of the samples. This trend was observed in PC1 (58.1%), PC2 (25.9%), PC2 (12.52%) and PC2 (12.69%) in the PCAs developed with VIS–NIR, FT-NIR, NIR and FTIR spectral information, respectively. According to the loading plots presented in Fig. 3B, D, F, H and I, in the VIS–NIR range for the Coastal class (CS1) the most important wavelengths are at 515, 553 and 705 nm and for the Andean class (CS2) at 538, 603, 655 and 682 nm (Fig. 3B). For the model developed in the FT-NIR range, the most important spectral range for CS1 is between 11,517–7236 cm^{-1} and for CS2 7236–4490 cm^{-1}, where the most influential bands are 5138 and 4667 cm^{-1} (Fig. 3D). In the case of portable NIR PCA, the most influential bands for CS1 are at wavelengths 1902, 2270, 2292 and 2337 nm, and for CS2 at 1968 and 2145 nm (Fig. 3F). The information collected in mid-IR, enabled to observe that multiple bands between 4000 and 3500 cm^{-1} are mainly important for CS2, and the 1800–1300 cm^{-1} range for CS1 (Fig. 3H, I). The same information was corroborated by the second derivative of the FTIR spectra (Fig. 4), where some differences in the absorption maxima of certain bands according to the phylogeographic origin

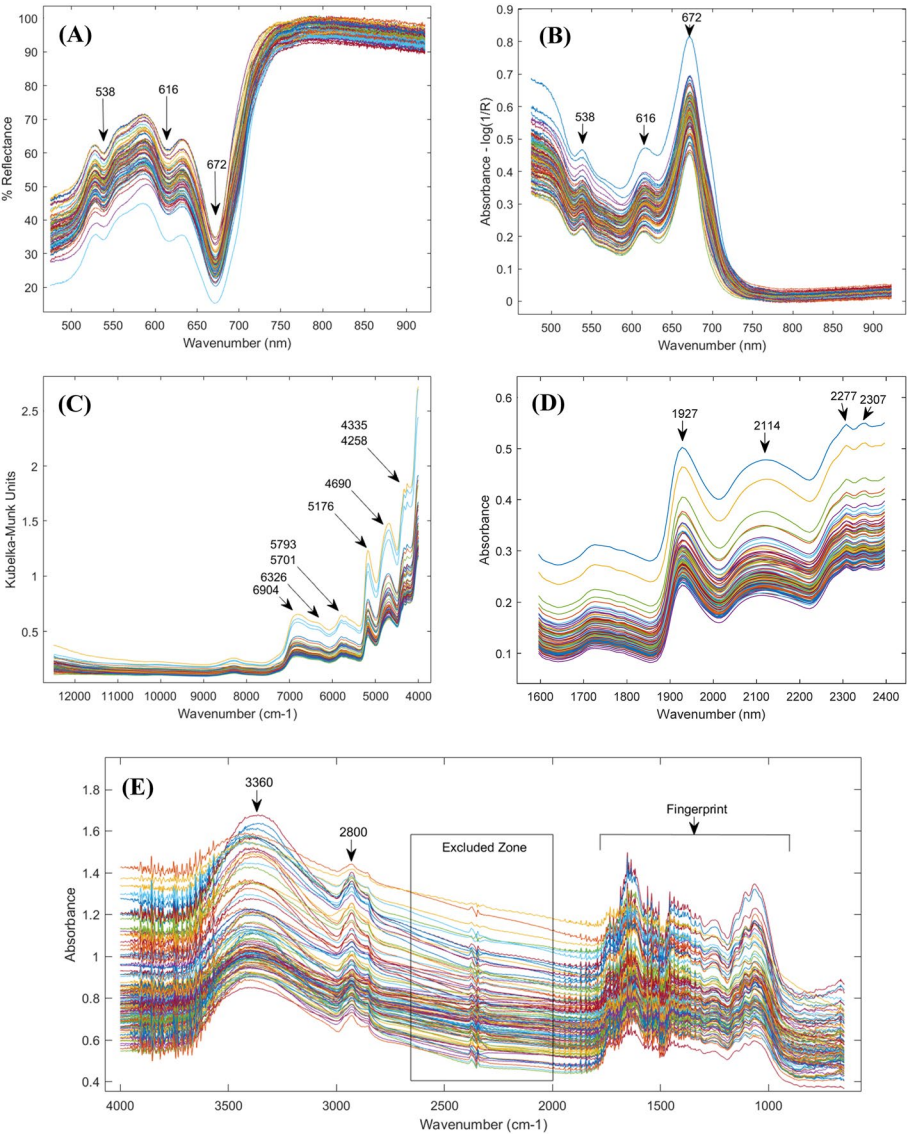

Fig. 2 Acquired raw spectra in the 4 spectral ranges of each powdered *Araucaria araucana* seedlings. **A** VIS–NIR reflectance spectra; **B** VIS–NIR pseudo-absorbance spectra; **C** FT-NIR absorbance spectra in Kubelka–Munk units; **D** absorbance spectra in NIR range with portable equipment; **E** FTIR absorbance spectra, dotted box indicates excluded spectral region (2600–2000 cm^{-1})

of the *A. araucana* seedlings were observed. The range 4000–3400 cm^{-1} shows higher absorbance in CS2 samples and bands in the wavelengths 1555, 1550, 1540 and 1026 cm^{-1} showed higher absorbance for the CS1.

The SIMCA analysis showed a clear differentiation between *A. araucana* seedlings of Andean and Coastal origin, allowing to obtain %CA between 88.24 and 90.57% in

Table 3 Principal component analysis for the 4 spectral ranges analyzed

Spectral matrix	PCs	Explained variance (%)	Accumulative variance (%)
VIS–NIR	PC1	58.10	58.10
	PC2	24.08	82.18
	PC3	6.49	88.67
FT-NIR	PC1	56.54	56.54
	PC2	25.88	82.42
	PC3	7.82	90.24
NIR	PC1	72.28	72.28
	PC2	12.52	84.79
	PC3	6.72	91.52
FT-IR	PC1	59.69	59.69
	PC2	12.69	72.39
	PC3	9.25	81.64

externally validated models (Table 4). The VIS–NIR and FTIR SIMCA models presented the lowest number of misclassified samples, obtaining the highest classification accuracy percentages (> 90%) and, therefore, the lowest classification errors (9.5 and 9.4%, respectively). At the same time, these two models proved to be the most robust with the smallest %CA variation when varying the training and validation sets. Regarding the evaluation parameters of these two best models the Coastal class presents lower Sn values in relation to the Andean class, and therefore the latter presents lower Sp values. This means that several Coastal samples are not recognized in their own class but are recognized in the Andean class, affecting the specificity of the latter class. According to SIMCA the variable with the highest discriminating power in the VIS–NIR range was the wavelength 700 nm, followed by wavelengths 641, 657 nm and others (Fig. 5A). In addition, the wavelength 535 nm is also one of the most important variables to discriminate seedlings of *A. araucana* according to phylogeographic origin. In the case of the SIMCA FTIR the most discriminant bands are 3034, 2870, 1703, 1585, 1511, 1453, 1414, 1347 and 1293 cm^{-1} (Fig. 5B, C).

Chlorophyll index SPAD

It was observed that SPAD values obtained from *A. araucana* leaves had a lower mean in Coastal plants than in Andean plants, 61.24 ± 14.09 and 69.11 ± 12.46 respectively (Table 5 and Fig. 6). The same behavior was observed in the medians of each class. On the other hand, the Andean class presented a greater range of SPAD values (min = 35.00 and max = 103.43). The data didn't fit a normal distribution ($p = 0.03$) and didn't comply with the assumption of homogeneity of variance ($p = 0.71$). After Wilks' nonparametric test, it was observed that there were significant differences between the median SPAD values of both phylogeographic origins with 99% confidence ($W = 0.001$). As the SPAD value is an index of chlorophyll content, the results obtained indicate that Andean seedlings have a significantly higher chlorophyll content than the Coastal seedlings.

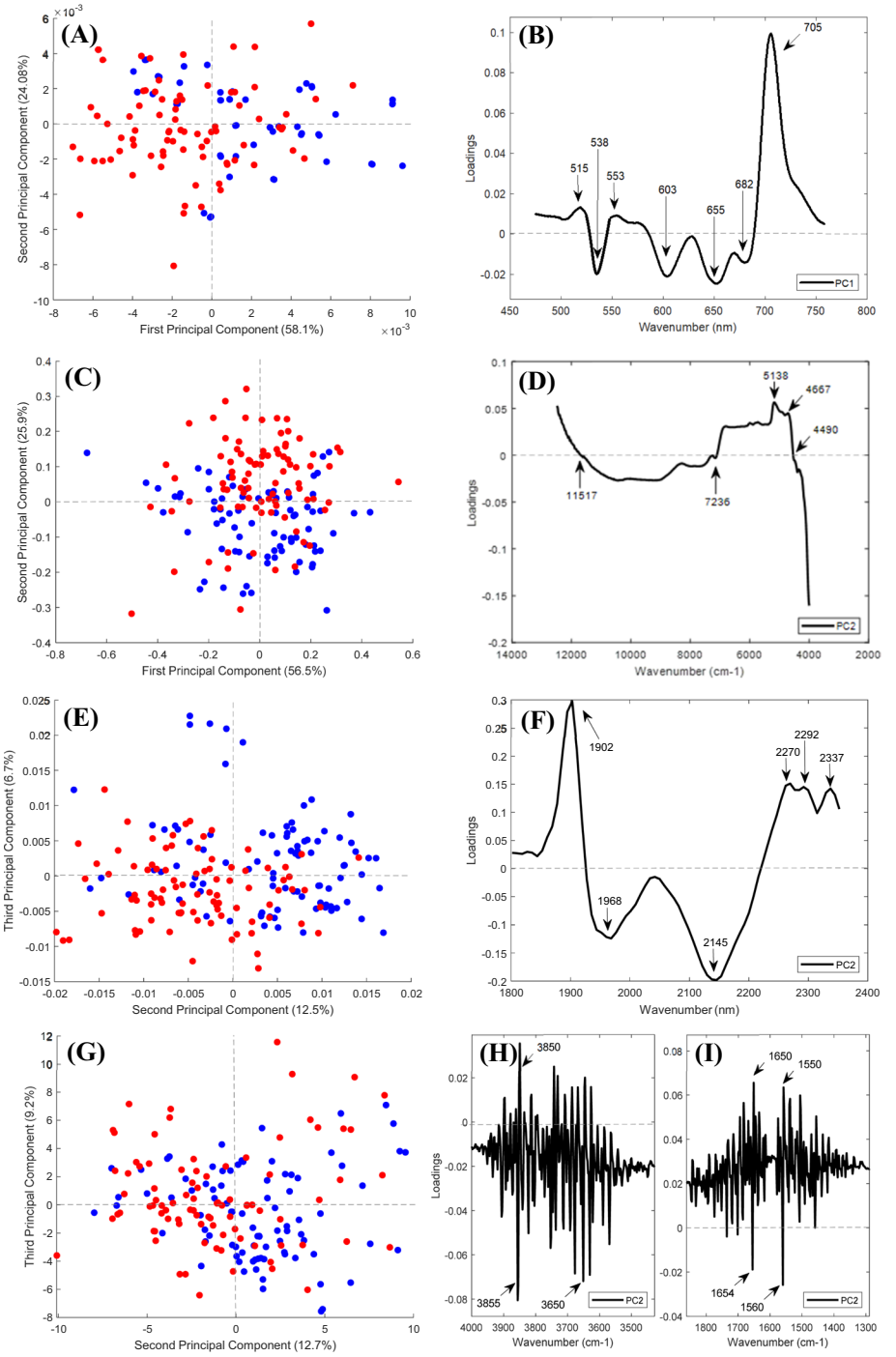

Fig. 3 Scores and loadings plot of PCAs that better explain the tendency of separation by phylogeographic origin of *A. araucana* seedlings, for: **A** and **B** VIS–NIR; **C** and **D** FT-NIR; **E** and **F** portable NIR; **G**, **H** and **I** FTIR. In scores, blue and red boxes correspond to Coastal and Andean samples, respectively

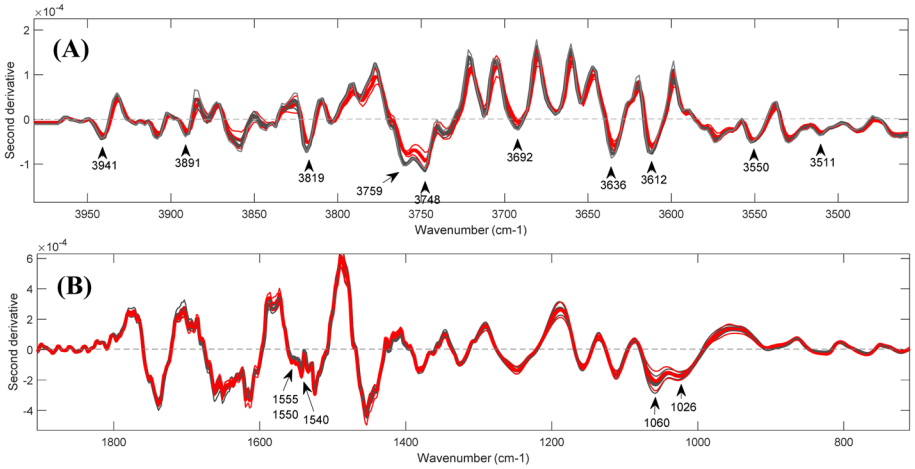

Fig. 4 Second derivative of the average FTIR spectra of each class. **A** Range 4000–3400 cm^{-1} and **B** Range 2000–600 cm^{-1}. Red and gray spectra correspond to CS1 and CS2, respectively; thick lines correspond to the average and thin lines to the respective average ± SD

Discussion

Spectral signatures

The VIS–NIR spectral pattern (Fig. 2B) obtained is like those reported for *Pinus sylvestris* L. (Danusevičius et al. 2014), *P. strobus* L. and *Quercus rubra* L. (Hallett et al. 1997). In the 400–530 nm range (green zone) carotenoid pigments absorb, specifically at 535 nm the xanthophyll photoprotective pigments absorb (Carranco et al. 2011). Between the 600–700 nm (red zone) the higher absorbance band can be observed, precisely this is the mainly absorption range of chlorophyll. According to literature band assignments (Table 6), the bands observed in FT-NIR (Fig. 2C) and NIR portable (Fig. 2D) are mainly related to carbohydrate (e. g C–H, C–O, O–H and CH_2) and protein (e. g amide I, amide II and N–H) functional groups (Schwanninger et al. 2011; Türker-Kaya and Huck 2017). Finally, in the mid-IR spectral pattern (Fig. 2E) the band identification was complex due to the number of bands present, but bands related to O–H and N–H functional groups can be seen near 3300 cm^{-1}, as well as CH_3 and CH_2 groups near to 2800 cm^{-1} and multiple bands in the fingerprint region (Beekes et al. 2007; Malek et al. 2014; Pandey and Pitman 2003).

Discrimination of phylogeographic origin

The results in the different spectral ranges indicated the feasibility of detecting chemical differences allowing to distinguish both phylogeographic origins of *A. araucana.* Although no clear separations between groups were observed in the PCAs, there was a tendency of separation between seedlings of Coastal and Andean origin, mainly attributable to small differences in their chemical composition. According to the results obtained, VIS–NIR and FTIR were the two best spectral ranges for the research objective. In the case of VIS–NIR,

Table 4 Summary of SIMCA model in the 4 spectral ranges

Spectral matrix	N° Samples	N° Factors			Training						Validation								Robustness Evaluation (Mean ± SD) %
					Classification						Classification								
		CS1	CS2	I.D		Pred CS1	Pred CS2	No match	%CA	%ER	N° Samples		Pred CS1	Pred CS2	No match	Sn	Sp	%CA	%ER	
VIS–NIR	99	4	4	0.24	CS1	37	1	0	98.99	1.0	42	CS1	13	3	0	0.81	0.96	90.48	9.5	87.3 ± 2.3
					CS2	0	61	0	98.99	1.0		CS2	1	25	0	0.96	0.81	90.48	9.5	
					Mean				98.99	1.0		Mean						90.48	9.5	
FT-NIR	116	3	3	0.28	CS1	50	4	0	93.97	6.0	51	CS1	20	4	0	0.83	0.93	88.24	11.8	86.5 ± 3.0
					CS2	3	59	0	93.97	6.0		CS2	2	25	0	0.93	0.83	88.24	11.8	
					Mean				93.97	6.0		Mean						88.24	11.8	
NIR	126	6	4	0.53	CS1	62	1	0	88.89	11.1	54	CS1	25	1	0	0.96	0.85	88.89	11.1	86.11 ± 3.9
					CS2	13	50	0	88.89	11.1		CS2	4	23	1	0.85	0.96	88.89	11.1	
					Mean				88.89	11.1		Mean						88.89	11.1	
FTIR	123	5	5	0.39	CS1	58	3	0	93.50	6.5	53	CS1	21	5	0	0.81	1.00	90.57	9.4	89.9 ± 1.1
					CS2	5	57	0	93.50	6.5		CS2	0	27	0	1.00	0.81	90.57	9.4	
					Mean				93.50	6.5		Mean						90.57	9.4	

CS1, Coastal class; CS2, Andean class; I.D, interclass distance. Pred CS1, samples predicted as class 1; Pred CS2, samples predicted as class 2; %CA, classification accuracy percentage; %ER, error rate percentage. Sn, Sensibility; Sp, specificity

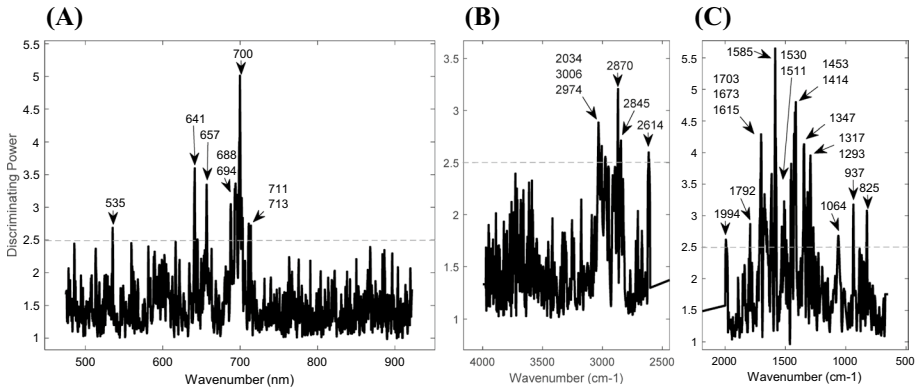

Fig. 5 Variable discrimination power (DP) plot of SIMCA models developed to discriminate between phylogeographic origins of *A. araucana* seedlings. Panels **A** and **B**, **C** correspond to the PD plots of the SIMCA models in the VIS–NIR and FTIR spectral range, respectively

Table 5 Descriptive statistics of the values of the chlorophyll meter measurements (SPAD) for each level (Coast and Andes)

Macrozone	n	Median	Mean	SD	Min	Max
Costa	12	62.78	61.24	14.09	37.47	79.87
Andes	28	69.33	69.11	12.46	35.00	103.43
Total	40	68.20	67.93	12.97	35.00	103.43

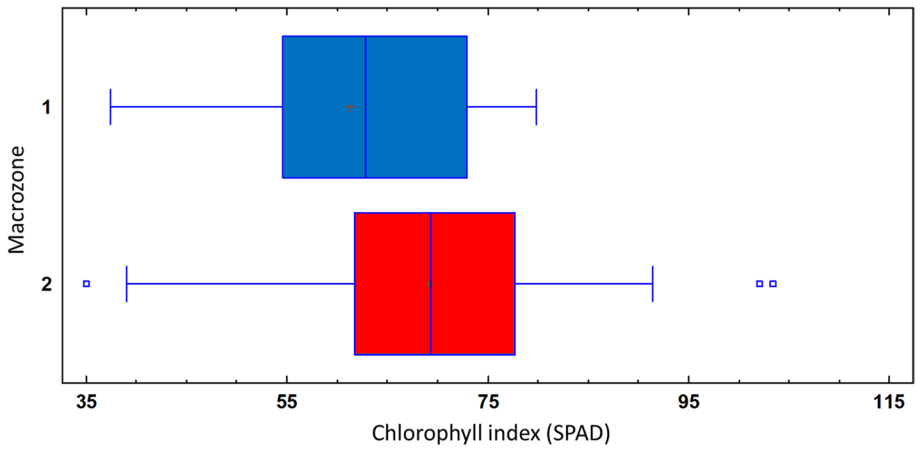

Fig. 6 Box-and-whisker plot of medians SPAD values for each class. Macrozones 1 and 2 are Coastal and Andean, respectively

the PC1 (51.6%) explained this separation trend with a higher % of variance than reported by molecular markers (16%; $p=0.004$) (Martín et al. 2014). Related to the SIMCA model the %CA obtained was lower than the %CA reported for the identification of seed sources of *P. sylvestris* through SIMCA and PLS-DA modeling (%CA of 100%) (Tigabu et al. 2005). According to PCA and SIMCA results the most discriminant variable was close

Table 6 Assignment of bands that present differences found in the spectroscopic characterization and multivariate analysis of the samples

λ (cm^{-1})	Assignment	Functional group	References
FT-NIR			
6904, 6326	1st + 2nd OT of N–H str	Proteins	Türker-Kaya and Huck (2017)
5793, 5701	1st OT of CH_2		Schwanninger et al (2011)
5176	N–H + O–H str	Proteins	Beć and Huck (2019)
4690	Amide I + Amide II	Proteins	
4335, 4258	C–H str + C–H def or 2nd OT of CH_2 ben	Polysaccharides	Schwanninger et al (2011)
λ (nm)	Assignment	Functional group	References
NIR			
1902	2nd OT C=O str	Carbohydrates	Türker-Kaya and Huck (2017)
1927	2nd OT C–O str	Carbohydrates	Türker-Kaya and Huck (2017)
1968	N–H asym str + Amide II or 2nd OT of C=O str	Proteins	Türker-Kaya and Huck (2017)
2114	O–H ben + O–H str	Cellulose	Türker-Kaya and Huck (2017)
2145	Amide I + Amide II	Proteins	Türker-Kaya and Huck (2017)
2270	O–H + C–O str	Carbohydrates	Türker-Kaya and Huck (2017)
2277	Aliphatic C–H str		Stenberg et al (2010)
2292	N–H str + C–O str	Proteins	Stenberg et al (2010)
2307	N–H str + C–O str	Proteins	Stenberg et al (2010)
2337	C–H str + CH_2 def	Hemicellulose, cellulose and lignins	Schwanninger et al (2011)
λ (cm^{-1})	Assignment	Functional group	References
FTIR			
4000–3500	O–H and N–H sym str	Plant fibers and proteins	Türker-Kaya and Huck (2017)
3360	O–H str		Zhuang et al (2020)
3034	C–H str		Zhuang et al (2020)
2870	CH_3 sym str	Mainly lipids and proteins with small contribution from carbohydrates	Malek et al (2014)
2800	C–H str		Beć and Huck (2019), Liu (2021)
1703	C=O str		Zhuang et al (2020)
1585	C–C ben		Zhuang et al (2020)
1555, 1550, 1540	C=N str + N–H str	Proteins	Beratto et al (2017)
1511	C=C aromatic stretch		Türker-Kaya and Huck (2017)
1453	C–H sym ben		Zhuang et al (2020)
1414	O–H bend		Türker-Kaya and Huck (2017)
1347	C–H ben of CH_2		Brandenburg and Seydel (2001)
1293	Amide III	Proteins	Beekes et al (2007)
1026	C–O str		Zhuang et al (2020)

sym, symmetrical; asym, asymmetricall; str, stretchingl; ben, bending; def, deformation; OT, overtone

to 700 nm, that according to Merzlyak et al. (2003) has a high correlation with chlorophyll content ($r > 0.90$). In addition, it was observed that the most influential bands for the Coastal class are related to chlorophyll a (Chla) and the most important bands for the Andean class are related to Chlb and xanthophylls (Azcón-Bieto et al. 2008; Ustin et al. 2009). Based on the above, Coastal and Andean *A. araucana* seedlings present chemical differences related to photosynthetic pigments that allow their discrimination, which in turn was indirectly estimated/ratified using the SPAD analysis. These results agree with the reports of Gong et al. (1997) who successfully discriminated between different conifer species due to variations in pigment content. This information also matches the significant differences obtained in the SPAD values between seedlings of both macrozones. According to Wu et al. (2008) the photosynthetic pigment content in leaves is a reliable indicator of photosynthetic activity, mutations, nutritional status and degree of stress, and it can be affected by environmental factors such as temperature, salinity or drought (Tanaka et al. 2008). It is well documented that the environmental conditions between these two mountain ranges are quite different, including soil type, temperature ranges, water availability, altitude, among others, factors that affect the development and functioning of the plants (Bekessy et al. 2004). It is currently known that there is a strong relationship between chlorophyll concentration and nitrogen (N) content, where soil type and quality are of great importance (Sánchez et al. 2018). In the literature, several studies have shown that higher contents of N in soil tend to increase the concentration of total chlorophyll in leaves (positive correlation) (Latsague et al. 2014; Sánchez et al. 2018; Torres-Netto et al. 2005). On the other hand, chlorophyll content has a high correlation with leaf morphology, in this sense, McIntosh et al. (unpublished results) showed that leaf morphological traits of this species, such as needle width, area, succulence, thickness, length, among others, differed significantly between Andean and Coastal seedlings grown in a common garden. The present study identified that Andean araucarias would have a higher content of Chlb and xanthophylls, both of which play important roles in plants by acting as antennae pigments protecting against photooxidative damage to Chl (González et al. 2014). These compounds act as protectors against light excesses by dissipating the energy in the form of heat (xanthophyll cycle), as well as contributing to a greater antioxidant capacity. In addition, climates characterized by high temperatures and high solar radiation are often accompanied by water restrictions, in Andes the water is often not available for a major part of the year (in the form of snow) requiring greater protection against reactive oxygen species (ROS). These results are consistent with the report of Aherne et al. (2009), who indicate that this pigment content is controlled by physiological, genetic and biochemical attributes and can even be influenced by geographic location.

Regarding FTIR spectra, following the separation trend of PC2, the second derivative and the discrimination power plot, it was identified by the assignment of functional groups that the differences between Andean and Coastal seedlings were mainly attributed to differences in proteins and vegetable fibers content. CS1 was mainly influenced by bands between 2000 and 650 cm^{-1}, where the most important bands are attributed to proteins, specifically Amide I and II (Beekes et al. 2007; Sablinskas et al. 2003). Numerous proteins have been reported to be involved in plant defense mechanisms against biotic factors (Blanco-Labra and Aguirre-Mancilla 2002). It is known that both ESUs of *A. araucana* differ in the composition of companion species, as well as in the fungal biota present, being both mountains exposed to different pathogens (Delmastro and Donoso 1980). Small Coastal populations have the worst conservation status which could be more susceptible requiring a higher content of defense compounds. Despite the high relevance of defense compounds associated with protective molecules, Moran et al. (2017) report these traits

have been overlooked in provenance studies. In the case of the Andean class, there are multiple bands associated with O–H vibration mainly attributable to plant fibers, as well as bands of C–H, CH_2 and CH_3 groups, which correspond to the main functional groups that constitute the xanthophylls structure (Maoka, 2019). A higher content of plant fibers in seedlings of Andean origin could be related to the need for greater hardening due to the more adverse development conditions encountered in the Andean zone (Bekessy et al. 2004). As shown in Table 1, the Andes Mountains have higher annual temperature ranges, which translates into more extreme temperatures than in the Coastal range.

About the evaluation parameters of the resulting models, the lower Sn in the Coastal class can be attributed to the large genetic difference evidenced between the PNN and VL populations (Bekessy et al. 2002; Martín et al. 2014). Bekessy et al. (2002) using RAPDs (Random Amplified Polymorphic DNA) determined that VL population was significantly different from the PNN population. They also indicate that precisely this population, so different in genetic terms, is also ecologically different, developing in granitic/metamorphic soils at lower altitudes than the other populations analyzed (Nahuelbuta and Andean populations). These results support the hypothesis of racial divergence reported by Delmastro and Donoso (1980) and suggest that this difference could be due to local adaptations. On the other hand, the low Sp of the Andean class is presumably related to the wide ecological range covered by this class, which is constituted by 10 localities or populations in this study. Thus, with samples originating from a wide ecological range, a broader range of genetic variations and hence chemical responses is to be expected. Other studies have revealed that PNN locations are more similar genetically to Andean populations than to VL (Ruiz et al. 2007). Through allozyme analysis, it was observed that the low elevation Andean populations (PNV, LON and PNC) were significantly closer to the Coastal populations, suggesting this phenomenon is the result of long-distance dispersal (Ruiz et al. 2007). Thus, the authors hypothesize that low-elevation populations in the Andes would have been colonized by Coastal refugia after the LGM, making sense that Coastal samples tend to be classified in the Andean class.

Importance of plant material origin

This study shows that there are significant chemical differences between both phylogeographic origins of this species, which are detectable by vibrational and chemometric techniques. In a few words, the spectra are nothing more than a phenotypic expression of signals related to the chemical and structural composition resulting from the environment and from the genetic variation derived from the evolutionary history of the species in question. As observed, differences in pigment, protein and plant fiber content were attributed to climatological and edaphological differences between the two mountain ranges, suggesting a process of local genetic adaptation in this species. These results are consistent with the adaptative genetic variation observed in the genome of *A. araucana* (Varas-Myrik et al. 2022). They report an adaptative divergence between Andean and Coastal individuals, this local adaptation was highly correlated with environmental variation, specifically to the Temperature Annual Range (TAR). The analysis of *A. araucana* seedling traits showed the same results of differentiation between this two macrozones (McIntosh et al. Unpublished results). Our research supports the hypothesis of divergence between Andean and Coastal *A. araucana* seedlings proposal by Delmastro and Donoso (1980). Ignoring the local adaptation of the populations of this species could cause the failure of restoration attempts due

to a poor adaptation of the individuals to the conditions of the site to be restored and even contamination or loss of the local gene pool (Broadhurst et al. 2008). The latter could have major consequences since the introduction of foreign genes can have destabilizing effects on the genetic integrity of native populations (Keller et al. 2000). According to McKay et al. (2005), the preservation of intra- and inter-population genetic variation is what guarantees the maintenance of the evolutionary potential of species. On the other hand, other studies suggest that the use of foreign material could even affect a series of biotic interactions (e.g. resistance to pathogens) (Linhart and Grant 1996). Due to all of the above, the results obtained in this research suggest that when managing this species *A. araucana*, its phylogeographic origin should be taken into consideration, avoiding the use of non-local material, given that the adaptive genetic variation of this species can be of great importance at the time of its management.

Conclusion

The application of IR spectroscopy and chemometric methods have shown excellent results for the discrimination of *Araucaria araucana* seedlings according to their phylogeographic origin. Exploratory analyses showed trends of separation between the two origins, with higher percentages of variance explained than those reported by molecular tools. In the different ranges, the SIMCA models presented high prediction percentages (88–90%), being feasible to classify according to phylogeographic origin in all the spectral ranges analyzed. The best results were obtained in the VIS–NIR and FTIR ranges with prediction percentages of 90% in models validated with external sets. Concerning the 10% of error, it is suggested to perform a deeper analysis, including analysis of the georeferencing of the samples and their environmental conditions, as well as to apply molecular tools due the possibility of hybrids. Wavelengths responsible for discrimination according to phylogeographic origin were associated with photosynthetic pigments, proteins and plant fibers. Andean seedlings have a higher content of chlorophyll (mainly attributed to Chlb content), xanthophylls and plant fibers. For seedlings of Coastal origin, the most important bands are related to Chla and protein contents. These chemical variations between ESUs are likely the result of environmental and genetic variations derived from the evolutionary history of *A. araucana*, demonstrating that differences reported at the genetic level are expressed at the chemical level.

Spectra in the VIS–NIR and FTIR region obtained from *Araucaria araucana* seedlings, treated by chemometric methods, captured the phylogeographic signal which separates the Coastal Mountain from the Andes Mountain. The application of these tools is suggested as alternatives to conventional methods for the discrimination of the phylogeographic origin of *A. araucana*. It should be considered that the models resulting from this research could be improved by increasing the number of samples, thus improving the robustness and sensitivity values of the models, which could be used in future *A. araucana* restoration programs to trace or authenticate the origin of the germplasm to be used, reducing the risks associated with working with foreign material.

Acknowledgements The authors are grateful to the Postgraduate Direction of the University of Concepción for granting a scholarship for the development of postgraduate studies, to ANID Subdirección de Capital Humano/Doctorado Nacional/2021-21210494, to the "Sistema de monitoreo de ecosistemas forestales" (SIMEF) project financed by the Food and Agriculture Organization of the United Nations (FAO)—Global Environment Facility (GEF), executed by the Forestry Institute (INFOR, Chile), the staff of the Carlos

Douglas Nursery of CMPC, especially to the general manager Juan Andrés Celhay, and the Plant Metabolomics Laboratory of Faculty of Pharmacy, UdeC.

Authors' contributions MR-R: Conceptualization, Methodology, Formal analysis, Investigation, Writing. RC: Formal analysis, Editing, Supervision; JG-C: Selection, Sampling, Samples acquisition, Writing, SPAD analysis; RI: Selection, Samples acquisition; MI Sanhueza: Formal analysis; RH: Writing, Formal analysis, Editing, Supervision. All authors have read and agreed to the published version of the manuscript.

Funding Research funded by the Articulation Scholarship of the Postgraduate Direction of the University of Concepción and Doctoral Scholarship ANID-2021.

Data availability Not applicable.

Code availability Not applicable.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Aherne A, Jiwan M, Daly T, O'Brien N (2009) Geographical location has greater impact on carotenoid content and bioaccessibility from tomatoes than variety. Plant Foods Hum Nutr (dordrecht, Netherlands) 64:250–256. https://doi.org/10.1007/s11130-009-0136-x

Alía R, Alba N, Agúndez D, Iglesias S (2005) Manual para la comercialización y producción de semillas y plantas forestales. Materiales de base y de reproducción (Organismo Autónomo Parques Nacionales. Ministerio de Medio Ambiente). Serie Forestal

Atkinson RJ, Thomas E, Roscioli F, Cornelius JP, Zamora-Cristales R, Franco-Chuaire M, Alcázar C, Mesén F, López H, Ipinza R, Donoso PJ, Gallo L, Nieto V, Ugarte J, Sáenz-Romero C, Fremout T, Jalonen R, Gaisberger H, Vinceti B, Kettle C (2021) Seeding resilient restoration: an indicator system for the analysis of tree seed systems. Diversity. https://doi.org/10.3390/d13080367

Azcón-Bieto J, Fleck I, Aranda X, Gomez-Casanovas N (2008) Fotosíntesis, factores ambientales y cambio climático. In: Fundamentos de Fisiología Vegetal

Ballabio D, Consonni V (2013) Classification tools in chemistry. Part 1: linear models. PLS-DA Anal Methods 5:3790–3798. https://doi.org/10.1039/c3ay40582f

Beć K, Huck C (2019) Advances in near-infrared spectroscopy and related computational methods. Molecules. https://doi.org/10.3390/molecules24234370

Beekes M, Lasch P, Naumann D (2007) Analytical applications of Fourier transform-infrared (FT-IR) spectroscopy in microbiology and prion research. Recent Progress Prion Res 123(4):305–319. https://doi.org/10.1016/j.vetmic.2007.04.010

Bekessy SA, Allnutt T, Premoli A, Lara A, Ennos R, Burgman M, Cortes M, Newton A (2002) Genetic variation in the vulnerable and endemic Monkey Puzzle tree, detected using RAPDs. Heredity 88:243–249. https://doi.org/10.1038/sj.hdy.6800033

Bekessy SA, Lara A, González M, Cortes M, Gallo L, Premoli AC, Newton AC (2004) Variación en *Araucaria araucana* (Molina) K. Koch (Araucaria o Pehuén). In: Variación Intraespecífica en las especies arbóreas de los bosques templados de Chile y Argentina (Claudio Donoso, Andrea Premoli, Leonardo Gallo y Roberto Ipinza, p 434). Editorial Universitaria

Beratto A, Agurto C, Freer J, Peña-Farfal C, Troncoso N, Agurto A, Castillo R (2017) Chemical characterization and determination of the anti-oxidant capacity of two brown algae with respect to sampling season and morphological structures using infrared spectroscopy and multivariate analyses. Appl Spectrosc 71(10):2263–2277. https://doi.org/10.1177/0003702817715654

Blanco M, Villarroya I (2002) NIR spectroscopy: a rapid-response analytical tool. TrAC Trends Anal Chem 21(4):240–250. https://doi.org/10.1016/S0165-9936(02)00404-1

Blanco-Labra A, Aguirre-Mancilla C (2002) Proteínas Involucradas en los Mecanismos de Defensa de Plantas. Acta Universitaria 12(3):3–28

Brandenburg K, Seydel U (2001) Handbook of vibrational spectroscopy. Wiley

Broadhurst LM, Lowe A, Coates DJ, Cunningham SA, McDonald M, Vesk PA, Yates C (2008) Seed supply for broadscale restoration: maximizing evolutionary potential. Evolut Appl 1(4):587–597. https://doi.org/10.1111/j.1752-4571.2008.00045.x

Bystrzanowska M, Tobiszewski M (2020) Chemometrics for selection, prediction, and classification of sustainable solutions for green chemistry—a review. Symmetry. https://doi.org/10.3390/sym12122055

Carranco M, Calvo M, Pérez-Gil F (2011) Carotenoids and their antioxidant function: a review. Arch Latinoam Nutr 61:233–241

Castillo R, Contreras D, Freer J, Ruiz J, Valenzuela S (2008) Supervised pattern recognition techniques for classification of *Eucalyptus* species from leaves NIR spectra. J Chil Chem Soc 53:1709–1713

Danusevičius D, Masaitis G, Mozgeris G (2014) Visible and near infrared hyperspectral imaging reveals significant differences in needle reflectance among Scots pine provenances. Silvae Genetica. https://doi.org/10.1515/sg-2014-0022

Delmastro R, Donoso C (1980) Review of distribution, variation and utilization of gene resources of *Araucaria araucana* (Mol.) Koch in Chile. Simposio. IUFRO en melhoramiento genético e productividade de espécies florestais de rápido crescimento, Brasil

Drake F, Martín A, Machuca M, Martínez J, Drake-Martin F, Martin L (2009) Networking sampling of *Araucaria araucana* (Mol.) K. Koch in Chile and the bordering zone of Argentina: implications for the genetic resources and the sustainable management. Iforest Biogeosci for. https://doi.org/10.3832/ifor0524-002

Farhadi M, Tigabu M, Pietrzykowski M, Danusevičius D, Odén PC (2017) Application of near infrared spectroscopy for authentication of *Picea abies* seed provenance. New for 48(5):629–642. https://doi.org/10.1007/s11056-017-9589-1

Fu H, Fan Y, Zhang X, Lan H, Yang T, Shao M, Li S (2015) Rapid discrimination for traditional complex herbal medicines from different parts, collection time, and origins using high-performance liquid chromatography and near-infrared spectral fingerprints with aid of pattern recognition methods. J Anal Methods Chem. https://doi.org/10.1155/2015/727589

Godbout J, Bomal C, Farr K, Williamson M, Isabel N (2018) Genomic tools for traceability: opportunities, challenges and perspectives for the Canadian forestry sector. For Chron 94(01):75–87. https://doi.org/10.5558/tfc2018-010

Gong P, Pu R, Yu B (1997) Conifer species recognition: an exploratory analysis of in situ hyperspectral data. Remote Sens Environ 62(2):189–200. https://doi.org/10.1016/S0034-4257(97)00094-1

González A, Briceño H, Chirinos J, Buonocore R, Villarreal Á (2014) Variation of the concentration of chlorophyll a, b, Total chlorophyll and photosynthetic rate in the mangrove Avicennia germinans Punta de Palmas, municipality Miranda, Zulia state, Venezuela. Revista Investigaciones Científicas UNERMB 5:67–82

Gutiérrez-Caro B, Quiroz I, Gracia E (2012) Bases para un reglamento de semillas y plantas de especies forestales utilizadas en Chile

Hallett R, Hornbeck J, Martin M (1997) Predicting elements in white pine and red oak foliage with near infrared reflectance spectropscopy. J near Infrared Spectrosc. https://doi.org/10.1255/jnirs.101

Ipinza R, Müller-Using S (2021) Migración asistida de *Araucaria araucana*. Santiago De Chile FAO y MINAGRI. https://doi.org/10.4060/cb2901es

IUCN International Union for Conservation of Nature (2022). Monkey puzzle. *Araucaria araucana*. The IUCN red list of threatened species. https://www.iucnredlist.org/es/species/31355/2805113

Jianguo G, Yuhuan W, Gendi X, Wenqiao L, Guohao Y, Jun M, Liu P (2012) Phylogeography of *Ulmus elongata* based on Fourier transform-infrared spectroscopy (FTIR), thermal gravimetric and differential thermal analyses. Biochem Syst Ecol. https://doi.org/10.1016/j.bse.2011.10.018

Keller M, Kollmann J, Edwards PJ (2000) Genetic introgression from distant provenances reduces fitness in local weed populations. J Appl Ecol 37(4):647–659. https://doi.org/10.1046/j.1365-2664.2000.00517.x

Latsague M, Sáez P, Mora M (2014) Efecto de la fertilizacización con nitrógeno, fósforo y potasio, sobre el contenido foliar de carbohidratos, proteínas y pigmentos fotosintéticos en plantas de *Berberidopsis corallina* Hook.f. Gayana. Botánica 71:37–42

León-Lobos P, Bustamante-Sánchez M, Nelson C, Alarcón D, Hasbún R, Way M, Pritchard H, Armesto JJ (2020) Lack of adequate seed supply is a major bottleneck for effective ecosystem restoration in Chile: friendly amendment to Bannister et al. (2018). Restor Ecol. https://doi.org/10.1111/rec.13113

Lerma-García M, Cortés V, Talens P, Barat J (2018) Variety discrimination of fruits, edible plants, and other foodstuffs and beverages by infrared spectroscopy. Comp Anal Chem 80:127–163. https://doi.org/10.1016/bs.coac.2018.03.004

Linhart Y, Grant M (1996) Evolutionary significance of local genetic differentiation in plants. Annu Rev Ecol Syst 27:237–277. https://doi.org/10.1146/annurev.ecolsys.27.1.237
Liu X (2021) IR spectrum and characteristic absorption bands. In Organic chemistry I. Kwantlen Polytechnic University. https://kpu.pressbooks.pub/organicchemistry/chapter/6-3-ir-spectrum-and-characteristic-absorption-bands/
Loewe V, Navarro-Cerrillo R, García-Olmo J, Riccioli C, Sánchez-Cuesta R (2016) Discriminant analysis of Mediterranean pine nuts (*Pinus pinea* L.) from Chilean plantations by near infrared spectroscopy. Food Control. https://doi.org/10.1016/j.foodcont.2016.09.012
Lohumi S, Lee S, Lee H, Cho BK (2015) A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration. Trends Food Sci Technol 46(1):85–98. https://doi.org/10.1016/j.tifs.2015.08.003
Luykx D, Ruth S (2008) An overview of analytical methods for determining the geographical origin of food products. Food Chem 107:897–911. https://doi.org/10.1016/j.foodchem.2007.09.038
Malek K, Wood B, Bambery K (2014) FTIR imaging of tissues: techniques and methods of analysis. Opt Spectrosc Comput Methods Biol Med. https://doi.org/10.1007/978-94-007-7832-0_15
Maoka T (2019) Carotenoids as natural functional pigments. J Nat Med. https://doi.org/10.1007/s11418-019-01364-x
Marini F (2010) Classification methods in chemometrics. Curr Anal Chem 6:72–79
Martín MA, Mattioni C, Lusini I, Molina JR, Cherubini M, Drake F, Herrera MA, Villani F, Martín LM (2014) New insights into the genetic structure of *Araucaria araucana* forests based on molecular and historic evidences. Tree Genet Genom 10(4):839–851. https://doi.org/10.1007/s11295-014-0725-1
Massart D, Vandeginste BGM, Buydens LMC, De Jong S, Lewi PJ, Smeyers-Verbeke J (1998) Handbook of chemometrics and qualimetrics—part A. Elsevier
McGarigal K, Cushman S, Stafford S (2000) Multivariate statistics for wildlife ecology. Research. https://doi.org/10.1007/978-1-4612-1288-1
McIntosh M, González-Campos J, Demaree P, Toro O, Ipinza R, Bustamante-Sánchez MA, Hasbún R, Nelson CR (2022) Unpublished results. Trait variation between and within Andes and Coastal mountain ranges in the iconic South American tree *Araucaria araucana* in Chile
McKay JK, Christian CE, Harrison S, Rice KJ (2005) "How local is local?"—A review of practical and conceptual issues in the genetics of restoration. Restor Ecol 13(3):432–440. https://doi.org/10.1111/j.1526-100X.2005.00058.x
Merzlyak M, Solovchenko A, Gitelson A (2003) Reflectance spectral features and non-destructive estimation of chlorophyll, carotenoid and anhocyanin content in apple fruit. Postharvest Biol Technol 27:197–211. https://doi.org/10.1016/S0925-5214(02)00066-2
Moran E, Lauder J, Musser C, Stathos A, Shu M (2017) The genetics of drought tolerance in conifers. New Phytol. https://doi.org/10.1111/nph.14774
Neal-Kababick J, Flora (2010) Rapid authentication of botanicals for dietary supplement cGMP's compliance by automated FT-NIR using chemometric modeling. Perkin Elmer Appl 1–4
Pandey K, Pitman A (2003) FTIR studies of the changes in wood chemistry following decay by brown-rot and white-rot fungi. Int Biodeterior Biodegrad 52:151–160. https://doi.org/10.1016/S0964-8305(03)00052-0
Pasquini C (2003) Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. J Braz Chem Soc 14:198–219
Pomerantsev A, Rodionova O (2020) Popular decision rules in SIMCA: critical review. J Chemom. https://doi.org/10.1002/cem.3250
Rodríguez R, Matthei O, Quezada M (1983) Flora arbórea de Chile. Editorial de la Universidad de Concepción
Rohman A, Nugroho A, Lukitaningsih E, Sudjadi (2014) Application of vibrational spectroscopy in combination with chemometrics techniques for authentication of herbal medicine. Appl Spectrosc Rev 49(8):603–613. https://doi.org/10.1080/05704928.2014.882347
Rohman A, Windarsih A, Hossain MA, Johan MR, Ali ME, Fadzilah NA (2019) Application of near- and mid-infrared spectroscopy combined with chemometrics for discrimination and authentication of herbal products: a review. J Appl Pharmaceut Sci 9:137–147. https://doi.org/10.7324/JAPS.2019.90319
Ruiz E, González F, Torres C, Fuentes-Arce G, Mardones M, Stuessy T, Samuel R, Becerra J, Silva M (2007) Genetic diversity and differentiation within and among Chilean populations of *Araucaria araucana* (Araucariaceae) based on allozyme variability. Taxon. https://doi.org/10.2307/25065913
Sablinskas V, Steiner G, Hof M (2003) Applications. In: Handbook of spectroscopy. Wiley, pp 89–168. https://doi.org/10.1002/3527602305.ch6

Sánchez E, Ruiz JM, Romero L, Preciado-Rangel P, Flores-Córdova M, Márquez-Quiroz C (2018) ¿Son los pigmentos fotosintéticos buenos indicadores de la relación del nitrógeno, fósforo y potasio en frijol ejotero? Ecosistemas y Recursos Agropecuarios 5:387–398
Sandak A, Sandak J, Negri M (2011) Relationship between near-infrared (NIR) spectra and the geographical provenance of timber. Wood Sci Technol 45:35–48. https://doi.org/10.1007/s00226-010-0313-y
Schwanninger M, Rodrigues J, Fackler K (2011) A review of band assignments in near infrared spectra of wood and wood components. J near Infrared Spectrosc 19:287–308. https://doi.org/10.1255/jnirs.955
Sherman H (1997) Infrared spectroscopy. In: Handbook of instrumental techniques for analytical chemistry. Prentice Hall, Upper Saddle River, pp 247–283.
Stenberg B, Viscarra Rossel R, Mouazen A, Wetterlind J (2010) Chapter five-visible and near infrared spectroscopy in soil science. Adv Agron 107:163–215. https://doi.org/10.1016/S0065-2113(10)07005-7
Tanaka Y, Sasaki N, Ohmiya A (2008) Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. Plant J 54(4):733–749. https://doi.org/10.1111/j.1365-313X.2008.03447.x
Tigabu M, Oden P, Lindgren D (2005) Identification of seed sources and parents of *Pinus sylvestris* L. using visible-near infrared reflectance spectra and multivariate analysis. Trees 19:468–476. https://doi.org/10.1007/s00468-005-0408-5
Torres-Netto A, Campostrini E, Oliveira JG, Bressan-Smith RE (2005) Photosynthetic pigments, nitrogen, chlorophyll a fluorescence and SPAD-502 readings in coffee leaves. Sci Hortic 104(2):199–209. https://doi.org/10.1016/j.scienta.2004.08.013
Türker-Kaya S, Huck CW (2017) A review of mid-infrared and near-infrared imaging: principles, concepts and applications in plant tissue analysis. Molecules (basel, Switzerland) 22(1):168. https://doi.org/10.3390/molecules22010168
Ustin S, Gitelson A, Jacquemoud S, Schaepman ME, Asner G, Gamon J, Zarco-Tejada P (2009) Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. Remote Sens Environ. https://doi.org/10.5167/uzh-23317
Varas-Myrik A, Sepúlveda-Espinoza F, Fajardo A, Alarcón D, Toro O, Castro-Nallar E, Hasbún R (2022) Predicting climate change-related genetic offset for the endangered southern South American conifer *Araucaria araucana*. For Ecol Manag. https://doi.org/10.1016/j.foreco.2021.119856
Villagrán C (2018) Biogeografía de los bosques subtropical-templados del sur de sudamérica. Hipótesis Históricas Magallania (punta Arenas) 46:27–48
Villagrán C, Armesto J (2005) Fitogeografía histórica de la Cordillera de la Costa de Chile. In Historia, biodiversidad y ecología de los bosques costeros de Chile (Cecilia Smith-Ramírez, Juan J. Armesto y Claudio Valdovinos). Editorial Universitaria. http://repositorio.uchile.cl/handle/2250/120075
Wu C, Niu Z, Tang Q, Huang W (2008) Estimating chlorophyll content from hyperspectral vegetation indices: modeling and validation. Agric for Meteorol 148(8):1230–1241. https://doi.org/10.1016/j.agrformet.2008.03.005
Zhuang J, Li M, Pu Y, Ragauskas A, Yoo CG (2020) Observation of potential contaminants in processed biomass using Fourier transform infrared spectroscopy. Appl Sci. https://doi.org/10.3390/app10124345

Authors and Affiliations

Macarena Rojas-Rioseco[1,2,5] · Rosario del P. Castillo[2,5] · Jorge González-Campos[3] · Roberto Ipinza[4] · M. I. Sanhueza[2,5] · Rodrigo Hasbún[1]

[1] Laboratorio de Epigenética Vegetal, Departamento de Silvicultura, Facultad de Ciencias Forestales, Universidad de Concepción, 4070386 Barrio Universitario s/n, Concepción, Chile

[2] Departamento de Análisis Instrumental, Facultad de Farmacia, Universidad de Concepción, 4070386 Barrio Universitario s/n, Concepción, Chile

[3] Instituto Forestal, Línea de Conservación y mejoramiento genético, Camino a Coronel km 7.5, San Pedro de la Paz, Chile

[4] Instituto Forestal, Línea de Conservación y mejoramiento genético, Isla Teja S/N, Valdivia, Chile

[5] Laboratorio de Bioespectroscopia y Quimiometría, Centro de Biotecnología, Universidad de Concepción, 4070386 Barrio Universitario s/n, Concepción, Chile