ABSTRACT

**Background. Different parts of the tree *Calophyllum inophyllum* L. (nuts, leaves, roots, bark, fruits, nut oil and resin) are used ~~in the field of~~ as traditional medicines and cosmetics in most of the Pacific Islands~~: its nuts, leaves, roots, bark, fruits, nut oil and resin~~. The efficacy of the oil as a natural remedy and in ~~the practice of~~ traditional cosmetics has been described by many throughout the South Pacific, which ~~drew our attention to finding its diversity both chemically and biologically~~led us to investigate *C. inophyllum*'s chemical and genetic diversity. A correlative study of the nut resin and leaf DNA from three distinct archipelagos in the South Pacific ~~could possible show existing diversity patterns of the C. inophyllum plant in the South Pacific~~was carried out in order to identify diversity patterns in *C inophyllum* across the South Pacific. Methods. ~~This study was applied here on the oil resin~~*Callophylum inophyllum* plants were sampled from ~~three Pacific countries:~~ French Polynesia, New Caledonia and Fiji ~~in order to reveal chemical and biological diversity~~. We extracted~~pay special attention to the "t~~Tamanu" oil (nut oil) resin for chemo-diversity studies and sampled leaf tissues for ~~phylogenetic~~genetic studies. ~~A~~ We ~~highlight is the application of a~~applied a method of analysis designed for ~~kernels in~~ small quantities (at a microscale level), and used HPLC, with co-elution of in house standards, to establish the chemo-diversity of tamanu oil resin. ~~The various oil resin components of *C. inophyllum* were defined by comparing the samples chemical profiles from HPLC analyses against eleven in-house standards. This objective was accomplished through the HPLC co-chromatography of plant extracts and by~~Genetic diversity was assessed using chloroplast barcoding markers (the Acetyl-CoA carboxylase (accD) gene and the psaA-ycf3 intergenicspacer region) ~~universal gene markers in accD and PsaA-Ycf3 for the genetic identification of C. inophyllum samples from the coastal sites of the three geographically isolated island archipelagos~~. Results. ~~Through corroborative evidence from HPLC chemical analyses, the investigation into the effect of geographical isolation and genetic diversity of the plant were supported by multivariate analysis (PCA) of chemical data.~~ Our HPLC analyses revealed 11 previously known constituents of tamanu oil, with variability among plant samples. We also ~~revealed here the isolation of~~isolated and characterized 2 new neoflavonoids from tamanu oil resin namely, tamanolide E1 and E2 which are diastereoisomers. Although genetic analyses revealed low genetic variation, our

multivariate (PCA) analysis of the tamanu oil resin chemical profiles revealed differentiation among geographic regions. For the first time, the use of the chloroplast gene loci accD for the diversity studies of Calophyllum inophyllum was also presented. Albeit genetic analyses revealed low genetic variations, co-elution experiments discovered tamanu resin compounds to be clustered spatially along a geographic plane specifically more prominent at some sites. **Conclusion. We showed here that chromatographic analyses utilizing formalized in-house standards of oil resin compounds for co-elution studies against oil resin samples can, in the case of *C. inophyllum*,** identify patterns of variation among samples and be highly **discriminat**eory and veryinformative of underlying diversity of *C. inophyllum* samples **from different geographical origins.**

INTRODUCTION

The studyChemical and medicinal properties of the Alexandrian laurel (*Calophyllum inophyllum*), otherwise known commonly as beach mahogany, hasve been extensively studied throughout the world but more so in the Asia-Pacific region (Leguillier, 2015; Pawar *et al*., 2011; Patil *et al*., 1993). Endowed withThe plant has a myriad of medicinal uses, most of which are forinvolve topical applications, most studies have concentrated more on the medicinal aspects of the plant. Recent studies such as Leguiller *et al*., (2015) and Ansel *et al*., (2016) have managed to focused more attention on the cosmetic aspects of the plant in the Pacific region. which shows its importance in the regionOther common names for the plant in some Pacific island countries are dDilo (Fiji), Ffetau (Samoa), tamanou de bord de mer (New Caledonia) and Ttamanu in the Cook Islands and French Polynesia. (Friday and Okano, 2006). Historically, before the conversion of Polynesians to Christianity, the tTamanu trees were considered as sacred. They were planted inside the royal Marae (sacred areas) (Dweck and Meadow, 2002).

The Dilo or Tamanu plant *Calophyllum inophyllum* belongs to Kingdom Plantae, Phylum Spermatophyta, Class Dicotyledonae, Order Malpighiales, Familythe flowering plant family Calophyllaceae, and Genus *Calophyllum* (APG, 2009; Prabakaran and Brito, 2012). Nand is native to the Indo-Pacific region (Africa, India, South East Asia, Australia and the Pacific islands)., iIt grows to a height of 8 – 15m and has a large canopy with a network of branches. The wood is widely used for making cabinet and other furniture, for carving, and for boat and canoe building. Furthermore, different parts of the plant have been used in traditional medicine and as

excellent raw material for cosmetics (Dweck and Meadows, 2002). For example, the oil from ~~the~~its ~~almonds~~nuts ha~~ve~~s been used for medicine against skin infections, as a scar remover as well as for other cosmetic uses (Friday and Okano, 2006). In Fiji, the oil is used to cure arthritis and joint pain, as an eye wash for conjunctivitis (Cambie and Ash, 1994) and also to prevent infantile rash. The resin mixed with strips of bark and leaves is used as a ~~household application for the~~ treatment of sore eyes. The green fruit is used against tuberculosis (Cambie and Ash, 1994). In some islands of Polynesia, the oil has been used as an alternative for candle nut oil in lamps and also massaged into hair and used as a common topical application for skin diseases and burns (Prabakaran and Brito, 2012). A number of studies have revealed ~~the~~ interesting biological activity of the oil, which ~~provides~~has antibacterial (Yimdjo *et al*, 2004), antifungal (Saravan *et al*., 2011), anti-inflammatory activity against skin infections (Bhalla *et al*., 1980).Most recently it has proven useful for wound healing (Leguiller *et al*., 2015). Such is the uniqueness of its properties that ~~the~~ "t~~T~~amanu oil" has been recognised as an active cosmetic ingredient and recorded as *C. inophyllum* seed oil by the International Nomenclature of Cosmetic Ingredients (INCI) (Assouvie, 2013; Ansel *et al*., 2015).

The bioactive components (belonging to neoflavonoid, xanthone and triterpene secondary metabolite groups) in this plant are highly rich ~~in all~~and recognised as having medicinal properties as shown in Table 1. The composition of the main neoflavonoid compounds in the oil are as follows: calophyllolide, inophyllums (C, D, E and P), tamanolides D and P, calanolide Gut 270, and finally the calanolides A, B and D (Laure, 2005; Leu et.al., 2009; Assouvie, 2013). ~~Of~~Also of ~~high~~ great interest is the anti-HIV efficacy of ~~the~~ pyranocoumarins from the Calophyllum genus ~~in anti against -HIV~~ (Wang et al., 2006), with recent ~~in advances now~~studies also examining their use in ~~promoting~~ cicatrization, and as an anti-aging agent (Leguillier et al., 2015; Ansel et al., 2016). ~~Among this renewed interest into the phytochemistry of this plant is includes investigation of t~~The compound calanolide A, a minor constituent in of C. *inophyllum* resin extract (Ansel et.al., 2016) has also attracted recent interest because, ~~which is was~~ as ~~stated~~ noted by Wang et al., (2006), ~~as~~ it is the only natural product that has progressed into human clinical trials with positive results ~~for~~against HIV-1.

The analysis of the chemical composition of leaves of *C. inophyllum* from various sites in French Polynesia and that of fruits originating from various sites in India has revealed ~~a terroir effect~~regional differentiation (Laure *et al*., 2005, Pawar et al. 2010) between~~for~~ C. *inophyllum*

plants found in land ~~to~~and those found closer to the coast. In order to evaluate the chemical qualities of the different ~~T~~tamanu oils of the South Pacific, phytochemical analy~~z~~ses are therefore necessary. Given that secondary metabolites play a role in the adaptation of plants to biotic and abiotic factors, it would be interesting to know if, as revealed by Pawar *et al*., 2011, abiotic factors such as the geography of harvesting sites have an influence on the chemical profile of ~~the~~tamanu oil.

In addition to ~~these problems~~the general question of regional influence on the chemical profiles of tamanu oil, ~~it is questionable~~we also were interested in whether dipyranocoumrins and more particularly inophyllums can be used, as described by Pawar *et al*., 2011, as chemotaxonomic markers for the identification of *C. inophyllum* from various geographical areas of the South Pacific. ~~South~~. ~~For this~~To address these questions, ~~an approach to~~we characterized genetic diversity ~~is conducted~~ in parallel with the phytochemical diversity, following a polyphasic approach, as ~~observed in polyphasic~~ has been ~~work~~ applied to some microbial research. Polyphasic taxonomy is a consensus building method of taxonomy developed to incorporate phenotypic, genotypic and phylogenetic data for micro-organisms (Vandamme *et al*., 1996). Research into ~~encompassing~~applying a dialled down version of this method for plant research, utilising only a two pronged approach of genotype and phenotype variation for elucidation of diversity has been achieved by Pawar *et al*.~~,~~ (2011)~~,~~; Lynch *et al*. (~~,~~ 2016), and ~~;~~ Hu *et al*., (2007). ~~Similarly~~As in ~~to~~ these previous ~~work~~studies, the present study aims to gain more insight into the diversity of the plant *C. inophyllum* in the localities of Fiji, French Polynesia and New Caledonia by utilizing genetic data from barcoding universal gene markers in accD and psaA-ycf3. ~~U~~We expect that by using these as tools to study ~~the~~ genetic variation~~s~~ and integrating the observed patterns ~~to~~with corresponding ~~existent~~ HPLC chemical profiles ~~from~~ (phenotypic information) for each sample, we can gain new insights that could~~. An integration of these two pieces of information would~~ help ~~to~~ explain current diversity patterns ~~of~~in this plant. This approach is however different from Pawar *et al*.~~,~~ (2011) as it does not utilise simple sequence repeat markers. Instead, we used sequence data from two chloroplast regions that have been employed for barcoding: the accD gene and the psaA-ycf3 intergenic spacer region . ~~The 3 major locations for collections are shown on Oceania's map in Fig. 1.~~ The plastid accD gene, which encodes for the ß-carboxyl transferase subunit of acetyl co-enzyme A carboxylase, is present in the plastids of most flowering plants, including non-photosynthetic parasitic plants and is

involved in fatty acid biosynthesis. Associated with a fast evolving genome region in some evolutionary lineages, it has been used in barcoding experiment analyses for Magnoliophyta (Lahaye et al., 2007), Mesangiospermae (Lam *et al*., 2016) and Fabids (Xi *et al*, 2012). Studies with ~~C.~~ *Chlamydomonas reinhardtii* and higher plants have shown that ycf3 is required for the stable accumulation of the Photosystem 1 complex (Boudreau et al., 1997; Ruf et al., 1997) and it appeared to link with psaA

MATERIALS AND METHODS

Sample collection

~~The~~ Tamanu ~~almonds~~nuts were collected during the fruit flushing season of June to August and even collected late into the month of December. In New Caledonia, ~~T~~tamanu ~~almonds~~nuts were collected under the scientific authorization of the South Province N°2050-2014. In Fiji, all samples were collected along ~~the~~ road~~-~~sides at each collection site. At each location, a replicate sampling plan was applied as observed in the geodata (Table 2). A total of two ~~L~~leaf samples were collected at each tree together with 3-5 ~~almonds~~nuts. From the leaf composite, one leaf was placed in a zip lock bag as a voucher ~~(as for the study as in the Fiji aspect)~~ and two leaf samples were stored at -80°C or immediately placed into RNA later solution for DNA extraction. The nuts~~almonds~~ were placed in polyethylene trays and aired to remove moisture for two weeks before they were placed into a solar drier and dried for ~~a period of~~ six to eight weeks. In total, there were 22 sampled sites and a total of 85 trees sampled from the 3 major collections in Fiji, French Polynesia and New Caledonia. Note that due to the nature of collections being conducted in isolated locations from both French Polynesia and New Caledonia, limited amount of materials were collected and thus vouchers specimens could not be deposited for these two research groups. Only the Fiji samples were deposited into a voucher collection in Fiji, kept at the South Pacific Regional Herbarium (Table S1).

Microscale Extraction, Purification and HPLC Analyses

From the collected nuts sampled from each of the 85 tree's, a set of 3-5 ~~almond~~nuts weighing 7-10g were picked from each set and subjected to a small scale cold extraction consisting of maceration and grinding of the ~~almonds~~nut in a cheese cloth and crushing in a mortar and pestle

to allow oil absorption. Samples were immersed in EtOAc and sonicated for 5 mins before this was partitioned with EtOH at a 1:1 ratio v/v. The EtOH layer was removed and dried in vacuo leaving the resin behind. Any remaining oil was de-fatted twice with 40mL hexane in total. Resinous extracts were dissolved in EtOAc: cyclohexane at a 1:1 ratio v/v to give a concentration of 10mg/mL for final injection. Chromatographic analyses were performed using an Agilent 1100 series gradient HPLC fitted with a UV/DAD system. The HPLC column used was a Interchrome Modula-Cart QS Uptisphere 5µm Si column, and the data were viewed on Agilent Chemstation software. The elution conditions were as follows: flow rate 1mL/min, pressure 2500psi max, injection volume 4 µL and the step gradient sequence.

The HPLC analyses of the samples from all three locations (Fiji: RawDataS1, New Caledonia: RawDataS2 and French Polynesia: RawDataS3) were subjected under numerous trials to acquire the maximum amount of compounds at the shortest runtime. On completion of the analyses, eighty resin samples were chosen from the total of eighty five samples and five were discarded due to poor alignment and high signal to noise. A further two samples were discarded due to poor alignment therefore a total of seventy eight samples were analysed in HPLC. In order to concatenate the chemical data and compare sample chromatograms observed at 280nm (which is where most picks were observed in the shortest time) against known compounds retention times, the removal of noise, redefining and alignments of all data was necessary in order to identify and assign each sample peak against the known compounds isolated in previous works by Leu *et al*., 2009; Leu, 2009; Laure, 2005 and Laure *et al*., 2008. Background fitting and identification of major peaks of the raw HPLC data were performed using the R package align DE v2.0.1. Chromatograms were aligned using a procedure in Scilab version 5.5.1 (Scilab Enterprises, 2012) derived from chromaligner (Wang *et al*., 2010).

Purification of known compounds from commercial ~~T~~tamanu oil resin and Isolation of new compounds

A batch of oil resin extract (157g) from commercial ~~T~~tamanu oil (provided by "Laboratoire de Cosmétologie du Pacifique Sud" manufacture) from French Polynesia was first partitioned with EtOH and an aqueous alkaline solution of Na_2CO_3 (10%, v/v). Its organic fraction was washed with distilled water and then dried with $MgSO_4$ to give a neutral fraction (53g) after solvent evaporation. This fraction was submitted to flash liquid chromatography on an open column with

a silica gel (240-300 mesh) using a stepwise gradient from cyclohexane to EtOAc yielding 12 fractions (Table S2). Fractions having similar R_f values on silica gel TLC (cyclohexane-acetone, 60:40, v/v) were combined. Fractions 7, 9 and 11 were submitted to repeated preparative HPLC using a Varian Dynamax Si column (250 × 21.4 mm id, 5µm with cyclohexane -EtOAc (10:90) in isocratic eluent conditions. This chromatographic purification network led to the isolation of new compounds tamanolides E1 and E2 as a mixture (2 mg) besides standard known compounds namely, calophyllolide, inophyllums (C, D, E, P), calanolides (Gut 70 and A, 12-oxo-calanolide) and tamanolides (D, P).

New Compounds Analysis

New compounds ~~from~~ were identified by NMR and were further analyzed with a Bruker Avance DRX500 spectrometer (^{1}H – 500.13 MHz) equipped with a 5 mm triple resonance inverse Cryoprobe TXI (^{1}H/^{13}C/^{15}N) in $CDCl_3$-99.8%, (δ_{1H}=7.26pm, δ_{13C}= 77.16ppm). The HR-ESI-MS data were collected using a QStar Elite mass spectrometer (Applied Biosystems SCIEX, Concord, ON, Canada) equipped with an ESI source operated in positive ion mode. Optical rotations were measured with a Perkin-Elmer 241 polarimeter equipped with a sodium (589 nm) lamp and a 1 dm cell. The FTIR spectra were established with a Thermo-Nicolet IR 200 spectrometer on a KBr cell and 4cm^{-1} resolution.

DNA extraction and Amplification

Plant DNA was extracted from young leaf tissue using a (Macheral) Plant Nucleospin II kit according to the manufacturer's instruction. The leaf samples were first lyophilized in liquid nitrogen and crushed before being subjected to the manufacturer's protocols. Two plant universal chloroplast regions were targeted for amplification namely; the Acetyl-CoA carboxylase (accD) gene and the psaA-ycf3 intergenic spacer region. The amplifications were performed in a Thermo Scientific Arktik Thermo Cycler. The PCR was performed using a Qiagen taq polymerase kit and primers were diluted to a concentration of 60ρmol before use in the PCR mix. All sample amplifications were eluted in distilled milliQ water at 50 µL volumes and tested for yield in 1% agarose gel electrophoresis. The PCR program consisted of 92°C of incubation during 2 mins for 1 cycle and 8 cycles of PCR (denaturation at 92°C for 30sec followed by specific annealing temperatures for accD (44°C) and psaA-ycf3 (43°C) all at 30sec, and 72°C for 1min), followed

by 40 cycles of (92°C for 30sec, extension at primer specific temperature's accD (46°C) and psaA-ycf3 at 30 sec, 72°C for 1 min). A final 5 min extension was then carried out at 72°C and for only 1 cycle.

Sequencing Alignment and Data Analysis

The samples were sequenced by GATC Biotech sequencing services, Germany. Data received in AB1 format were viewed under the Applied Biosystem's sequencing scanner and corrected for contiguous read lengths and miscalled sequence data were removed especially in the beginning and at the ends of the raw data files. Files were then aligned using MEGA6 (Tamura *et al*., 2013) before they were matched to nucleotide sequences using the BLASTn tool available at https://blast.ncbi.nlm.nih.gov/Blast.cgi. Matches Blasted in NCBI using the nucleotide search tool. Only outgroup hits with 96-100% similarity were included into the construction of a multiple sequence alignment. Sequences were aligned using Clustal W (Thompson *et al*., 1994) and trees were generated using the MEGA6 phylogeny tools. New haplotypes were deposited into the NCBI Genbank using the Bankit tool and accession numbers ~~generated in the process~~ ~~(~~are given in Table S3~~)~~.

Phylogenetic construction

All newly determined sequences were checked using the sequencer software for peak intensity and also contiguous length. All sequences with contiguous lengths greater than 200bp were considered for tree construction using MEGA6. Maximum likelihood (ML) trees were constructed for all accD sequences which satisfied our contiguous length criterian. Initial tree(s) for a heuristic search were first obtained by applying Neighbor-Joining (Saitou and Nei, 1987) and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach implemented in MEGA, and tree searches were conducted assuming a Tamura-Nei model and a discrete Gamma distribution to model evolutionary rate differences among sites (5 categories (+G, alpha parameter = 0.1523)). The analysis involved 18 nucleotide sequences and all positions containing gaps and missing data were eliminated. There was a total of 234 positions in the final dataset. The optimal tree obtained has been drawn to scale, with branch lengths indicating the number of substitutions per site ~~has been shown~~. A maximum likelihood tree was also constructed for the psaA-ycf3 Fiji sequence~~d~~ data ~~s~~ only as the Tahiti and

New Caledonia sequenced data s for this chloroplast region could not be amplified, possibly due to DNA deterioration. A Tamura-Nei model (Tamura and Nei, 1993) was assumed. The optimal tree has been shown, drawn to scale, with branch lengths measured indetermined by the number of substitutions per site. The percentage of trees in which the associated taxa clustered together was shown next to the branches. The analysis of this second data set involved 11 nucleotide sequences. Positions containing gaps and missing data were eliminated. There was a total of 55 positions in this dataset.

Statistical Analysis

Multivariate statistical analysis was applied to all the chemical data in order to reveal the extent to which the 12 major compounds making up the chemical compositions of the 47 C. *inophyllum* oil resin samples were geographically distributed. Data were normalized by log transformation prior to Principle Component Analysis (PCA). PCA was performed with the package Ade4 ofusing R version 3.1 software (R Development Core Team, 2018) and a PCA biplot was drawn using \the eMicrosoft Excel software. The diversity of chemical compounds that contributed to the highest discrimination at a geospatial level were visualized and a scatter plot was generated from Microsoft eExcel.

RESULTS

Chemical composition of extracts

The fractionation, purification and identification of resin extracts (by spectroscopic methods) yielded 11 compounds from commercial tTamanu oil (Fig. 2 A-K), which were identified as constituents that had been previously reported (Laure et al., 2008; Leu et al., 2009). In addition to calophyllolide, inophyllums (C, D, E, P), calanolides (Gut 70 and A, 12-oxo-calanolide) and tamanolides (D, P), these analyses also led to the isolation of two new compounds X1 and X2 as an epimeric mixture. The presence of these 2 compounds X1 and X2 from the same peak in two fractions was clearly shown in ^{1}H theoretical (Fig. 3A) and experimental spectra's in (Fig. 3B). As the chemical shifts shown in Table 3 of these compounds were quite similar, their absolute presence proven on the ^{1}H NMR revealing more signals then expected from one compound and which actually corresponded to two sets of signals for two very close compounds that we

proposed here as X1: tamanolide E1 and X2: tamanolide E2 (Fig. 2L). The chemical characteristics of tamanolide E1 and E2 are: amorphous yellowish powder; $[\alpha]_{25}^{D} = -19.6$ (c 0.003, $CHCl_3$). FTIR (CCl_4): 3075, 2979, 2926, 2908, 2877, 2832, 1740, 1695, 1643, 1606, 1575, 1461, 1382, 1209, 1152, 1121 cm^{-1}. ^{1}H NMR ($CDCl_3$, 500MHz) and ^{13}C ($CDCl_3$, 125 MHz) (see Table 3); HR-ESI-MS m/z 383.1850 $[M+H]^+$ (calcd. for $C_{23}H_{27}O_5$, 383.1853.).

Chemical Diversity of C. *inophyllum* within the 3 regions

HPLC analysis of ***C. inophyllum*** was performed to investigate the clustering patterns of phytochemical components based on geographical distribution (archipelago). Eleven isolated compounds (Fig.4: peaks 1-11), along with two new compounds (Fig.4: peak 12), were used as external standards to qualitatively analyze~~d~~ four sample of *C. inophyllum* from three different locations, Fiji West (FW1), New Caledonia Noumea (N1 and N3) and Rotuma Fiji (R4). The phytochemical components of each sample were identified by comparing the retention time of the external standards (Fig. 4).

The alignment of all samples through R and Scilab provided a matrix containing the integrated peak area of the identified features. Data were normalized by log transformation prior to PCA analysis. The samples can be discriminated into three main regions (Fig. 5) with the French Polynesian population on the left of the graph, the New Caledonia population on the bottom right and the Fiji population on the upper right. Within French Polynesia and Fiji, no discrimination between the sites or archipelagos was observed with these first two components.

The different proportions of the compounds leading to biogeographic discrimination of the three regions ~~has been~~is shown in Fig. 6. Compounds calophyllolide and tamanolide E1/E2 were found to be negatively loaded on axis 1 which covers 7 to 59% of the total information, the proportion of tamanolide E1/E2, Gut 70 calanolide, tamanolide and inophyllum E were the main and significant compounds prominent in the composition of the oil resin from French Polynesia, while tamanolide P, D and inophyllum C were found to be positively loaded on axis 2 covering 15-83% of the total information and were more prominently represented ~~more~~ in the Fiji samples, while New Caledonia appears to be profoundly represented by other constituents such as compound P69, which although unknown and unidentified, appeared to be unique from the PCA

analysis. Outgroups were observed in the chemical diversity of the Tahiti samples namely from the Tuamotu's and Australes as seen in Fig. 5.

Genetic discrimination

Mega 6 identified the Tamura-Nei model as the best fitting model in ML analyses. Phylogenetic relationships assuming this model of substitution for C. *inophyllum* arehave been shown in Fig. 7A and Fig. 7B for the accD gene and for the psaA-ycf3 spacer region, respectively.. Interestingly, only Raiatea from French Polynesia showed a unique genotype, where it produced two2 new haplotypes (T22 and T23) in the accD tree caused by substitutions at positions 134bp and 146bp.

DISCUSSION

CNew chemical isolates and chemical variability section

Compounds X1 and X2 had a molecular formula of C23H27O5 as determined by high-resolution mass spectrum (HR-ESI-MS) (m/z 383.1850 [M+H]+, calcd 383.1853) and NMR data implying 11 degrees of unsaturation (Table 3, Fig. S1-S10). The 13C NMR spectrum gave a total of 23 separated resonances and the 135-DEPTQ sequence showed the presence of six methyl, one methylene, six methine and ten quaternary carbons including a ketone carbonyl at δ 191.7 ppm for compounds X1 and X2. The 1H resonances of both compounds (Fig. S1 and S9) were typical of a pyranocoumarin [8] as characterized by the downfield shifted proton at δ 6.16 (1H, s, H-3) for an α,β-unsaturated lactone (ring B), and a set of doublet signals [δ 6.66 ppm (1H, d, J = 10.0Hz, H-8), 5.60 ppm (1H, d, J = 10.0Hz, H-7)], and two methyl singlets at δ 1.54 ppm (3H, s, H-19) and δ 1.54 (3H, s, H-20) for the first one (X1) and at δ 1.53 ppm (3H, s, H-19) and δ 1.53 ppm (3H, s, H-20) for the second one (X2), for a pyrane ring (ring C).

Furthermore, the cycle ring D showed a characteristic cis-configuration with a coupling constant of 3.4Hz between H-10 δ 4.69 ppm (1H, qd, J = 6.5, 3.4 Hz) and H-11 δ 2.68 ppm (1H, qd, J = 7.2, 3.4 Hz) for X1 and between H-10 δ 4.70 ppm (1H, qd, J = 6.5, 3.4 Hz) and H-11 δ 2.69 ppm (1H, qd, J = 7.2, 3.4 Hz) for X2 and by NOE crosspeaks correlation between the two methyl groups H-21 δ 1,.42 ppm (3H, d, J = 6.6Hz), H-22 δ 1,.15 ppm (3H, d, J = 7.2Hz) for X1, and H-

21 δ 1,.41 ppm (3H, d, J = 6.6Hz), H-22 δ 1,.15 ppm (3H, d, J = 7.2Hz) for X2, like inophyllum E.

With the aid of the COSY experiment, an isobutyl unit was identified by further analysis of the remaining 1H resonances [δ 3.79 ppm (1H, brsxt, J = 7.0Hz, H-13), 1.75 ppm (1H, m, H-14a), 1.45 ppm (1H, m, H-14b), 0.95 ppm (3H, t, J = 7.4Hz, H-15) and 1.22 ppm (3H, d, J = 6.7Hz, H-16)] for X1 and δ 3.79 ppm (1H, brsxt, J = 7.0Hz, H-13), 1.75 ppm (1H, m, H-14a), 1.45 ppm (1H, m, H-14b), 0.96 (3H, t, J = 7.4Hz, H-15) and 1.22 ppm (3H, d, J = 6.7Hz, H-16)] for X2. This isobutyl moiety was assigned to be at the C-4 position, based on the HMBC crosspeaks between H-3 (δ 6.16 ppm) and C-13 (δ 37.5 ppm), between H-16 (δ 1.22 ppm) and C-4 (δ 104.7 ppm), between H-14 (δ 1.45, 1.75 ppm) and C-4 (δ 104.7 ppm), between H-13 (δ 3.79 ppm) and the carbons C-3 (δ 109.3 ppm) and C-4a (δ 104.7 ppm). The rest of the HMBC correlations used to obtain a complete assignment of the 1H and 13C NMR chemical shift, have been summarized and presented in Table 3. The 1H and 13C NMR of both compounds were almost identical to those of inophyllum E with the exception of an isobutyl group at C-4 and tamanolide D with the exception of the ring D.

Complete assignments of X1 and X2 were made based on 1D and 2D NMR experiments, of which compound X1 was suggested as tamanolide E1 while compound X2 was the H-13 epimer of X1 and suggested as tamanolide E2. Proton NMR spectrum (Fig. 3) of different HPLC fractions revealed X1 and X2 as a mixture from two compounds whose signals were close to each other with a varied ratio (70/30 to 40/60). The chemical shifts of both compounds are listed in Table 3. Different fractions of a collected HLPC peak show a variation of proportion 70/30 next 40/60 of two close signals in the 1H NMR spectrum (Fig. 3). ~~The a~~Assignment of the resonances was ~~performed~~ based on the ~~basis of~~ 1D and 2D NMR experiments (1H, 13C-DEPTQ135, COSY, HSQCed, HMBC and NOESY). One compound was identified as tamanolide E1 while the second compound was concluded to be the H-13 epimer form tamanolide E2.

Finally, the structures of 11 known compounds are shown (Fig. 2) including the two new compounds assigned as tamanolide E1 and E2 after elucidation from exhaustive NMR analysis (1H, 13C-DEPTQ135, COSY, HSQCed, HMBC, NOESY), including the comparison of experimental and ~~theorical~~theoretical results (Fig. 3A and B) and by comparison with NMR data of the phenylcoumarin back bone from inophyllum E and tamanolide D.

Consistent with earlier findings (Prabakaran and Britto, 2012; Laure *et al*., 2008), calophyllolide remains as the dominant major compound in tTamanu resin for all chemotypes isolated. Calanolide A which is the most potent of all C. *inophyllum* compounds with anti-HIV-1 activity (Wang *et al*., 2006) was found as a minor component in C. *inophyllum* oil. The pyranocoumarine's inophyllum B and P (also active against HIV-1-) were alsohave been isolated here. The former is not only the most active but it is also the only natural product undergoing clinical trials against HIV-1 (Liang, 2006). As seen in Fig. 4, the 280nm UV spectrum revealed the maximum number of peaks and therefore standards and samples were targeted for visualization at this wavelength.

All the standards isolated here as seen in Fig. 4 have been isolated in previous studies from French Polynesian C. *inophyllum* oil (Leu *et al*., 2009a; Leu *et al*., 2009b; Laure, 2005). A unique characteristic of C. *inophyllum* oil is the presence of a resin which has been shown as well to contain most neoflavonoid biologically active constituents (Ansel *et al*., 2016). Previous studies by McKee *et al*., 1998 and Li *et al*., 2007 have indicated the use of HPLC analyses to investigate differences in chemical profiles from different populations. In our case, we have looked at samples from three localities. The population from New Caledonia was characterised by peaks not found in earlier studied populations and these novel peaks suggest new and specific chemomarkers. Unique to this study as shown in the biplot on Figure 6 is that we were able to show the chemodiversity of the tTamanu resin oil across the three study sites if we only use thosebased on chemical compounds that contribute to the highest discrimination at a geographical level from the PCA analysis. This point is important and rather unique in the study as this reveals the chemical specificity of compounds and their proportions in Ttamanu resin by geographical locations across the South Pacific.

DNA variability

This is the first reported work on C. *inophyllum for the* usinguse of the universal chloroplast barcoding regions, the gene accD and the intergenic region psaA-ycf3 on C. *inophyllum*. Low levels of genetic variation were found with these markers, and this variation was not functionally linked to differences in chemical composition. The absence of genetic variation among Fiji psaA-ycf3 regions suggests that for C. *inophyllum* low genetic variation in C. *inophyllum. might also be the case for populations in other geographical localities.* However, analyses of additional

chloroplast gene regions or nuclear ISSR markers may yet reveal more discrimination within the Fiji archipelago and between populations from different archipelagos.

CONCLUSIONS

Our results are informative in revealing that chemical differences in tamanu resin can be a tool for the discrimination of samples and geographic regions. In our case, chromatographic data proved to be more informative and highly discriminative thaen DNA barcoding data, possibly owing to low genetic variation in the chloroplast regions utilized. Additional chloroplast barcoding regions or utilization of nuclear microsatellites may give a better perspective on ~~the current~~ patterns of genetic diversity. Of interest were haplotypes (T22 and T23) which were distinct from the more commonly found C. *inophyllum* genotype. Polyphasic taxonomy that considers both chemical diversity and genetic diversity presents the best approach to delineate biological variation across geographical boundaries. However, higher levels of genetic resolution are required to ~~explain~~characterize variation in the disjunct distribution of C. *inophyllum* across the South Pacific and to bring insight into the diversification processes that have occurred following geographic isolation. This diversification ~~has included~~ novel pyranocoumarins compounds characterized by this study: ~~in~~ tamanolide E1 and E2 (C-13 epimers as a mixture).

ACKNOWLEDGEMENTS

This paper is dedicated to the memory of our wonderful colleague, supervisor and friend Prof William Aalbersberg who has passed away. We thank and greatly appreciate him for his wisdom and inspiring research leadership. We are grateful to the people of French Polynesia, Fiji and New Caledonia for giving us access to sampling areas to collect our nuts and leaf samples. We thank Cloe Check, Juliette Prevost, Nicolas Martin, and Ranitea Ly for performing HPLC analysis on all the samples. We also thank Mr. John Bennett for supplying oil extracts and assisting in the processing of samples in Fiji as well as Mr. Olivier Touboul (LCPS) for ~~T~~tamanu sample collection in French Polynesia. Special thanks to Mr. Marika Tuiwawa and Mr. Alivereti Naikatini for their assistance in the sample collection and deposition of samples in Fiji.

REFERENCES

Angiosperm Phylogeny Group, 2009. "An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III". *Botanical Journal of the Linnean Society* 161: 105–121 DOI : 10.1111/j.1095-8339.2009.00996

Ansel J , Lupo E , Mijouin L , Guillot S , Butaud J , Ho R , Lecellier G , Raharivelomanana P, Pichon C. 2016. Biological activity of Polynesian *Calophyllum inophyllum* oil extract on human skin cells. *Planta Medica* 82 : 961-966.
DOI : 10.1055/s-0042-108205

Assouvie N. 2013. Le Tamanu (*Calophyllum inophyllum* L.) en Polynésie Française et autre espèces du genre *Calophyllum*: De l'usage en médecine traditionnelle à l'émergence d'un médicament anti-VIH. Thèse de doctorat en pharmacie, Université de Bordeaux Ségalen, p 122, n° 2013BOR2P019.

Bhalla TN, Saxena RC, Nigam SK, Misra G, Bhargava KP. Calophyllolide—A new non-steroidal anti-inflammatory agent. Indian J. Med. Res. 1980, 72, 762–765.

Boudreau E, Takahashi Y, Lemieux M., and Rochaix, J.-D. (1997). The chloroplast ycf3 and ycf4 open reading frames of Chlamydomonas reinhardtii are required for the accumulation of the photosystem I complex. EMBO J. 16, 6095–6104. DOI: 10.1093/emboj/16.20.6095

Cambie RC, and Ash J. 1994. Dicotyledons in Fijian Medicinal Plants. CSIRO Australia, 119 - 120.

Dweck AC and Meadows T. 2002. Tamanu (*Calophyllum inophyllum*) - the African, Asian, Polynesian and Pacific Panacea. *Int J Cosmet Sci*. 24(6):341-8. doi: 10.1046/j.1467-2494.2002.00160.x.

Friday JB, and Okano D. 2006. Calophylum inophyllum (Kamani). Species profiles for Pacific Island Agroforestry. Ver 2.1. Permanent Agriculture Resources (PAR), Hōlualoa, Hawai'i. http:// www.traditionaltree.org

Hu Y, Zhang Q, Xin H, Qin L-P, Lu B-R, Rehman K, Zheng H (2007). Association between chemical and genetic variation of Vitex roduntifolia population from different locations in China: its implication for quality control of medicinal plants. Biomedical Chromatography 21:967–975. DOI: 10.1002/bmc.841

Ishikawa, T. 2000. Anti HIV-1 Active Calophyllum Coumarins: Distribution, Chemistry and Activity. *Heterocycles* 53 (2) : 453-474. DOI: 10.3987/REV-99-526

Itoigawa M, Ito C , Hugh T , Tan W , Kuchide M , Tokuda H , Nishino H , Furukawa H. 2001. Cancer chemo preventive agents, 4-phenylcoumarins from *Calophyllum inophyllum*. *Cancer Letters* 169 : 15–19. DOI : 10.1016/S0304-3835(01)00521-3

Jin L , Tabe Y , Kimura S , Zhou Y , Kuroda J , Asou H , Inaba T , Konopleva M , Andreeff M , Miida T. 2011. Antiproliferative and proapoptotic activity of GUT-70 mediated through potent inhibition of Hsp90 in mantle cell lymphoma. *British Journal of Cancer* 104 : 91 – 100. DOI: 10.1038/sj.bjc.6606007

Kostova I, and Mojzis J. 2007. Biologically active coumarins as Inhibitors of HIV-1. *Future HIV Therapy* 1(3): 315-329. DOI: 10.2217/17469600.1.3.315

Lahaye R, Van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, Maurin O, Duthoit S, Barraclough T, and Savolainen V. 2007. DNA barcoding the floras of biodiversity hotspots. *PNAS* Vol 5 (8): 2923-2928. DOI: 10.1073 /pnas.0709936105

Lam V, Merckx V, and Graham S. 2016. A few-gene plastid phylogenetic framework for mycoheterotrophic monocots. *American Journal of Botany* 103(4): 1-17. DOI: 10.3732/ajb.1500412

Laure F. 2005. Etude de la composition chimique et de la biodiversité du Calophyllum inophyllum de Polynésie Française. Thèse de doctorat de chime, Université de la Polynésie Française, p 357, n° 2005POLF0001.

Laure F, Raharivelomanana P, Butaud J-F, Bianchini J-P, Gaydou EM. 2008. Screening for Anti-HIV-1 inophyllums by HPLC-DAD of Calophyllum inophyllum leaf extracts from French Polynesia Islands. *Analytica Chimica Acta* 624: 147-153.
DOI: 10.1016/j.aca.2008.06.046

Léguillier T, Lecsö-Bornet M, Lémus C, Rousseau-Ralliard D, Lebouvier N, Hnawia E, Nour M, Aalbersberg W, Ghazi K, Raharivelomanana P, Rat P. 2016. The Wound Healing and Antibacterial Activity of Five Ethnomedical Calophyllum inophyllum Oils: An Alternative Therapeutic Strategy to Treat Infected Wounds. *PLoS ONE* 10(9):1-20. DOI: 10.1371/

Leu T, Raharivelomanana P, Soulet S, Bianchini J-P, Herbette G and Faure R. 2009. New tricyclic and tetracyclic pyranocoumarins with an unprecedented C-4 substituent. Structure

elucidation of Tamanolide, Tamanolide D and Tamanolide P from *Calophyllum inophyllum* of French Polynesia. *Magnetic Resonance in Chemistry* 46: 989-993. DOI: 10.1002/mrc.2482

Leu T. 2009. Contribution à la connaissance de la flore polynésienne: évaluation de l'intérêt pharmacologique de quelques plantes médicinales et étude phytochimique du Tamanu (Calophyllum inophyllum L. – Clusiaceae) [dissertation]. Papeete, Tahit: Université de la Polynésie Française.

Li HL, Zhang WD, Zhang C, Han T, Liu R-H, Hu J, Chen HS. 2007. Comparative analysis of chemical profile of wild and cultivated populations of Corydalis saxicola by high performance liquid chromatography. *Phytochemical Analysis* 18:393–400. DOI: 10.1002/pca.994

Liang W, Tao M, and Gang L. 2006. Recent Progress in Calophyllum Coumarins as Potent Anti-HIV Agents *In* Liang X, and Fang W. 2006. *Medicinal Chemistry of Medicinal Natural Products*. Wiley-Interscience. pp 325-350.

Lynch RC, Vergara D, Tittes S, White K, Schwartz CJ, Gibbs MJ, Ruthenburg TC, deCesare K, Land, DP & Kane, NC (2016). Genomic and Chemical Diversity in Cannabis. Critical Reviews in Plant Sciences 35: (6), 349-363. DOI: 10.1080/07352689.2016.1265363

Nei M, and Kumar S. 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York.

McKee TC, Covington CD, Fuller RW, Bokesch HR, Young S, Cardellina JH II, Kadushin MR, Soejarto DD, Stevens P, Cragg GM, Boyd MR. 1998. Pyranocoumarins from tropical species of genus Calophyllum: a taxonomic study of extracts in National Cancer Institute Collection. *Journal of Natural Products* 61:1252–1256. DOI: 10.1021/np980140a

Patil AD, Freyer AJ, Eggleston DS, Haltiwanger RC, Bean MF, Taylor PB, Caranfa MJ, Breen AL, Bartus HR, Johnson RK, Hertzberg RP, Westley JW. 1993. The Inophyllums, novel inhibitors of HIV-1 reverse transcriptase isolated from Malaysian tree, *Calophyllum inophyllum Linn*. *Journal of Medicinal Chemistry* 36: 4132–4138. DOI: 10.1021/jm00078a001

Prabakaran K, Britto SJ. 2012. Biology, Agroforestry and Medical Value of Calophyllum inophyllum L. (Clusiacea): A Review. *International Journal of Natural Products Research* 1 (2): 24-33. Pawar KD, Swati JP, Shubhada RT. 2011. Association between chemical and genetic variation in Calophyllum inophyllum, a medicinally important tree of the Western Ghats of India. Plant Systematic Evolution 292: 257–265. DOI 10.1007/s00606-010-0409-8

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Ruf S, Kossel H, and Bock, R. (1997). Targeted inactivation of a tobacco intron-containing open reading frame reveals a novel chloroplast-encoded photosystem I–related gene. J. Cell Biol. 139, 95–102. DOI: 9314531.

Rzhetsky A, and Nei, M. 1992. A simple method for estimating and testing minimum evolution trees. *Molecular Biology and Evolution* 9: 945-967. DOI: 10.1093/oxfordjournals.molbev.a040771

Saitou N, and Nei M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4: 406-425. DOI: 10.1093/oxfordjournals.molbev.a040454

Saravanan, R & Dhachinamoorthi, D & Senthilkumar, K & Thamizhvanan, K. (2011). Antimicrobial activity of various extracts from various parts of Calophyllum inophyllum L.. Journal of Applied Pharmaceutical Science. 1. 102-106.

Saxena RC, Nath R, Palit G, Nigam SK, Bhargava KP. 1982. Effect of Calophyllolide, a non-steroidal anti-inflammatory agent, on capillary permeability. *Journal of Medicinal Plant Research* 44(4): 246-248. DOI: 10.1055/s-2007-971459

Scilab Enterprises .2012. Scilab: Free and Open Source software for numerical computation (Windows 8, Version 5.5.1) [Scilab]. Available from: http://www.scilab.org

Tamura, K. 1992. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G + C-content biases. *Molecular Biology and Evolution* 9: 678-687. DOI: 10.1093/oxfordjournals.molbev.a040752

Tamura K, Stecher G, Peterson D, Filipski A, and Kumar S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* 30: 2725-2729.

Tamura K. and Nei M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* 10:512-526. DOI: 0131~4038/93/1003-0

Thompson JD, Higgins DG, and Gibson TJ. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*. 22 (22): 4673-4680. DOI: 10.1093/nar/22.22.4673

Quisumbing E. 1951. Medicinal Plants of the Philippines. In Dweck, A.C and Meadows, T. 2002. Tamanu (*Calophyllum inophyllum*) – the African, Asian, Polynesian and Pacific Panacea. *International Journal of Cosmetic Science* (24): 1-8. DOI: 10.1046/j.1467-2494.2002.00160

Vandamme P, Pot B, Gillis M, de Vos P, Kersters K, Swings J. Polyphasic taxonomy, a consensus approach to bacterial systematics. Microbiological Reviews. 1996;60(2):407-38. DOI: PMC239450

Wang L, Ma T, and Liu G. 2006. Recent Progress in *Calophyllum* Coumarins as Potent Anti-HIV Agents *In* Liang, X.; and Fang, W. 2006. *Medicinal Chemistry of Medicinal Natural Products*. Wiley-Interscience. pp 325-350.

Wang S, Ho T, Kuo C, Tseng C. 2010. Chromaligner: a web server for chromatogram alignment. *Bioinformatics* 26 (18): 2338-2339. DOI: 10.1093

Yimdjo MC, Azebaze GA, Nkengfack AE, Meyer AM, Bodo B, Fomum ZT. 2004. Antimicrobial and cytotoxic agents from *Calophyllum inophyllum*. *Phytochemistry* 65: 2789–2795. DOI: 10.1016/j.phytochem.2004.08.024

Xi, Z; Ruhfel, R. B; Schaefera, H; Amorimd, A. M; Sugumarane, M; Wurdack, K. J; Endress, P. K; Matthews, M. L; Stevens, P. F; Mathews, S; and Davis, C. 2012. Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. *PNAS* 109(43): 17519–17524. DOI: 10.1073/pnas.1205818109