UPLC-MS/MS-based molecular networking and NMR structural determination for the untargeted phytochemical characterization of the fruit of Crescentia cujete (Bignoniaceae)

UPLC-MS/MS-based Molecular Networking and NMR structural determination for the untargeted phytochemical characterization of the fruit of *Crescentia cujete* (Bignoniaceae)

Andrés Rivera-Mondragón,[a,*] *Emmy Tuenter,*[a] *Orlando Ortiz,*[b] *Maria E. Sakavitsi,*[c] *Theodora Nikou,*[c] *Maria Halabalaki,*[c] *Catherina Caballero-George,*[d] *Sandra Apers,*[a] *Luc Pieters*[a] *and Kenn Foubert*[a]

[a]Natural Products & Food Research and Analysis (NatuRA), Department of Pharmaceutical Sciences, University of Antwerp, Universiteitsplein 1, 2610, Antwerp, Belgium.

[b]Herbario PMA, Universidad de Panamá, Estafeta Universitaria, Panama City, Republic of Panama.

[c]Department of Pharmacognosy and Natural Products Chemistry, Faculty of Pharmacy, National and Kapodistrian University of Athens, Zografou, 15771 Athens, Greece.

[d]Centre of Innovation and Technology Transfer, Institute of Scientific Research and High Technology Services (INDICASAT-AIP), Building 208, City of Knowledge, Panama, Republic of Panama.

*Corresponding author at:

Natural Products & Food Research and Analysis (NatuRA), Department of Pharmaceutical Sciences, University of Antwerp, Universiteitsplein 1, 2610, Antwerp, Belgium.

E-mail address: arivera.qclabherbals@gmail.com (A. Rivera-Mondragón)

Note: Sandra Apers passed away on 5 February 2017.

ABSTRACT

The fruit pulp of *Crescentia cujete* is traditionally used in folk medicine for the treatment of a variety of respiratory conditions and gastrointestinal disorders. Due to the lack of a comprehensive phytochemical description of the fruit of this plant, its active compounds and rational quality control parameters have not yet been described. An untargeted metabolomics approach combining UPLC-MS/MS-based molecular networking with conventional isolation and NMR methods was carried out for the phytochemical profiling of the fruit pulp of *Crescentia cujete*. Sixty-six metabolites, including nine *n*-alkyl glycosides, twenty-three phenolic acid derivatives (such as cinnamoyl and benzoyl derivatives), fifteen flavonoids, four phenylethanoid derivatives and fifteen iridoid glycosides were identified at different levels of confirmation: eighteen confirmed structures (Level 1), six probable structures (Level 2) and forty two tentative candidates (Level 3). Among these, all four phenylethanoid (**25**-**28**) derivatives were described for the first time within this species. In addition, 8-*epi*-eranthemoside (**13**), crescentiol A (**21**) and crescentiol B (**24**) were reported as three undescribed iridoid glucosides. The use of molecular networking has resulted in a detailed phytochemical overview of this species. This work provides a useful tool for the further development and validation of appropriate analytical methods for routine quality control assessment of commercially available products containing the fruit of this species and further interpretation of their related pharmacological effects.

KEYWORDS. *Crescentia cujete*, Bignoniaceae, calabash, totumo, molecular networking, UPLC-MS/MS, *n*-alkyl glycosides, flavonoids, phenylethanoids, phenolic acids, iridoids glycosides.

The Abstract and Introduction need to be revised and written more clearly. You need to make the objectives more evident and what is novel in your work. What do you offer over and above the present data-bases? If there is not enough novelty, the paper is not likely to be acceptable.

1. INTRODUCTION

Crescentia is a Neotropical genus of Bignoniaceae represented by six species, ranging from Mexico to Amazon Brazil including West Indies (Gentry, 1980). It is also cultivated throughout tropical areas. *Crescentia cujete* L. is a small tree (up to 10 m high) with a small crown and long branches. Calabash fruit is seasonal appearing at the end of the dry season (Ejelonu et al., 2011; Gentry, 1980).

This species is mostly cultivated through tropical America, although its natural range average has not been clarified (Ambrósio Moreira et al., 2016). While it is widely distributed and cultivated from Mexico to South America and in the Caribbean region, it has been introduced and cultivated in tropical Africa (from Senegal to Cameroon) and South-East Asia (Ejelonu et al., 2011; Kaneko et al., 1997; Olaniyi et al., 2018; Wang et al., 2010). It is known vernacularly as 'calabash' (Jamaica, Bahamas, Grand Cayman, Bermuda, Belize), 'totumo' or 'jícaro' (Latin America) and 'Dao Tien' (Vietnam) (Gentry, 1980; Wang et al., 2010).

Calabash is currently important in folk medicine, since extracts of leaves, fruit, stem bark, flowers and seeds have been used for the treatment of various illnesses (Gentry, 1980; Wang et al., 2010). The pulp of the fruit of *C. cujete* mixed with sugar or honey is used as a medicine for alleviating respiratory conditions such as cold, cough and asthma, for the treatment of digestive system ailments (stomach pains, intestinal parasites) and in case of infertility (Das et al., 2014; Olaniyi et al., 2018). The fruit of this plant has been used since the eighteenth century in traditional Mexican medicine to prepare a tonic for the treatment of different respiratory infections, asthma, bronchitis, tuberculosis and breast pain (Valladares and Rios, 2007). In Panama, the fruit is used as febrifuge, laxative, emetic, vermifuge, and for inducing abortion (Gupta, 1987). In Vietnam, a fruit decoction is taken orally as an expectorant, antitussive, laxative and for the treatment of urethritis and stomach disorders (Kaneko et al., 1997; Wang et al., 2010). However detailed chemical studies on the active compounds of the fruit of *C. cujete* are scarce; a literature survey revealed that there are few studies addressing its phytochemical composition in a rather general way, such as the presence of tannins, phenols, flavonoids, saponins, alkaloids and hydrogen cyanide (Ejelonu et al., 2011). Additional studies showed the presence of iridoids (such as aucubin and ningpogenin), *n*-alkyl

glycosides, lignan glycosides (acanthoside D), and benzoyl derivatives (Kaneko et al., 1998, 1997; Wang et al., 2010).

Liquid chromatography tandem mass spectrometry (LC-MS), mainly through untargeted data-dependent MS/MS methods, is one of the most popular analytical methods in metabolomics since it provides a versatile approach of what a biological sample is composed of. Nevertheless, the interpretation of such complex data represents one of the major challenges in natural product research. Therefore, efficient handling of these data sets requires automated data treatment and comparison with spectral libraries relevant to natural products (Allard et al., 2016; De Oliveira et al., 2017). Recently, novel bioinformatics approaches such as molecular networking (MN) have emerged and provided new perspectives in order to facilitate the identification of known compounds in complex matrices. The Global Natural Product Social Molecular Networking (GNPS; https://gnps.ucsd.edu) web-based platform, an open-access database for tandem mass (MS/MS) spectrometry data, is a new approach that enables online dereplication and automated molecular networking for the analysis of LC-MS/MS-generated data sets (Kang et al., 2018; Wang et al., 2016). The main concept of molecular networking is the establishment of chemically similar networks of metabolites based on MS/MS spectral similarities, by categorizing secondary metabolites into clusters of related molecular structures (Kang et al., 2018). Thus, the use of metabolomics in combination with chemometric analysis is a powerful tool with great potential for both the phytochemical characterization and quality control of herbal products (Lee et al., 2017).

Currently, the quality of herbal derived products is usually assessed by the determination of one or two compounds as chemical markers using high performance liquid chromatography (HPLC) or gas chromatography (GC). However, herbal products contain numerous constituents, which implies that one or two chemical indicators would not provide an appropriate approach for quality control purposes. Therefore, the implementation and development of new quality strategies requires fundamental changes in order to improve quality standards. The aim of this investigation was to provide a comprehensive description of the phytochemical composition of the fruit pulp of *Crescentia cujete* collected in Panama by using a metabolomics approach combining UPLC-MS/MS-based molecular networking (generated at the GNPS website), with conventional isolation and NMR methods. This allowed the fast identification of known compounds, molecular families

and subsequent prioritization of the purification of compounds associated to unknown features. The chemical composition of *Crescentia cujete* is characterized according to the levels of confirmation as proposed by Schymanski et al., 2014: Level 1 (L1): Structure confirmed by reference standard or structure elucidation by NMR spectroscopy; level 2a (L2a): Probable structure by library spectrum match; level 3 (L3): tentative candidates based on MS, MS2 experimental data.

The metabolomics profiling information described in this investigation can subsequently be applied for the development and validation of appropriate quality control of food supplements derived from *Crescentia cujete*, such as antitussive oral liquid formulations, which are currently available on the market.

2. RESULTS AND DISCUSSION

In order to explore and achieve a holistic picture of the phytochemical composition of the fruit of *C. cujete*, its crude extract and fractions were analyzed by UPLC-HRMS with the automated acquisition of MS/MS spectra in a data-dependent analysis mode. A complementary characterization strategy was carried out using an MS/MS-based molecular networking approach and further isolation and structure elucidation of the main compounds was performed based on NMR analysis.

UPLC-HRMS profiles of *C. cujete* afforded 829 MS/MS spectra of parent ions (*m/z*) in negative ionization mode to develop a molecular network based on cosine similarity and visualized as nodes (Fig. 1). The molecular network was generated using the online workflow at the GNPS website. The edges represent the cosine similarities between different MS/MS spectra (>0.60). The MS/MS spectra in the network were then searched in an automated batch mode against available GNPS spectral libraries and hits were annotated as part of this dereplication strategy.

The first step consisted of visual inspection in order to select interesting clusters in the complete molecular network (Fig. 1). The exploration of the molecular network revealed that compounds were grouped according to their substituent similarities, for example in agreement with their aglycone and/or degree of glycosylation. By browsing through the dereplicated structures in the network, different clusters corresponding to *n*-alkyl glycosides (green), phenylethanoids (purple),

flavonoid glycosides (red), benzoyl derivatives (sky-blue), cinnamoyl derivatives (dark-blue) and iridoid glycosides (gold) were highlighted (Fig. 1).

The molecular network based on a fragmentation study, combined with NMR experiments, allowed the tentative characterization or full identification of 66 metabolites, including three undescribed iridoid glycoside derivatives (**13**, **21** and **24**). All identified metabolites with the confidence levels as proposed by Schymanski et al. can be found in Table 1. In parallel, analysis of the metabolites detected by UPLC-MS in positive ionization mode based on HRMS full spectra was performed in order to support and confirm our findings from the MS data recorded in negative ion mode (see supporting information: Table S1).

***2.1. n*-Alkyl glycosides.** In total, 9 compounds were identified as *n*-alkyl glycosides, most of which were detected as di-*O*-glycosides and *n*-alkyl-conjugates, such as pentanediol (**2**, **4**), pentane (**4**), octanol (**5**, **6**) and octane (**7**-**9**) derivatives. The MS/MS spectra of the precursor ions of compounds in cluster AG (see supplementary information Fig. S1), gave product ions at *m/z* 265.13 (**3**), 249.13 (**4**), 307.17 (**5**), 289.16 (**7**), 423.22 (**8**) and 291.18 (**9**), corresponding to a neutral loss of an anhydropentose residue $[M - H - 132.04]^-$, and additional ions at *m/z* at 161.04, characteristic of an anhydrohexose. Furthermore, compounds **2** and **3** were unambiguously identified as (2*R*,4*S*)-2-*O*-*β*-D-glucopyranosyl-2,4-pentanediol (Kaneko et al., 1998) and (2*R*,4*S*)-2-*O*-*β*-D-xylopyranosyl-(1→6)-*O*-*β*-D-glucopyranosyl-2,4-pentanediol (Kaneko et al., 1998), respectively, by comparison of their ^{1}H and ^{13}C-NMR spectral data (supporting information: Table S2 and Fig. S5-S8) with those previously reported for *C. cujete*. In addition, the presence of sucrose (**1**) was also confirmed by comparison of retention time and MS/MS fragmentation pattern with that of an authentic standard.

2.2. Phenylethanoid derivatives. MS^2 spectra of identified compounds from cluster FG1 (see Fig. 2) were characterized by a loss of 162.0 Da. They were clustered within the network due to the similarity of their fragmentation patterns, suggesting their structural similarity. Since an MS/MS fragment ion tolerance of 0.1 Da was used for the generation of the molecular network (as recommended), it was observed that the cleavage of isobaric hexosyl (162.05 Da) and caffeoyl (162.03 Da) residues (which are chemical substituents with the same nominal mass but with a different molecular formula), produced similar fragmentation patterns. As a result, FG1 was

observed as a heterogeneous cluster constituted of phenylethanoid, flavonoid and caffeoyl derivatives (Fig. 2A).

Unlike the majority of *n*-alkyl glycosides and iridoid glycosides, formic acid adducts $[M - H + FA]^-$ of these compounds were not observed. Compounds **25** (*m/z* 623.1981) and **26** (*m/z* 623.1980) were detected in cluster FG1 as two isomers displaying identical MS spectral properties, while only differing in their retention time (see Table 1). On the other hand, compounds **27** (*m/z* 637.2134) and **28** (*m/z* 651.2288) were located in cluster PE (Fig. 2A). All these compounds showed one main fragment $[M - H - 162.03]^-$, suggesting the loss of a caffeoyl moiety.

In order to obtain a better overview of the chemical structures of these molecules, full HRMS and MS^2 experiments were carried out in positive ionization mode (see supporting information: Fig. S9). For compounds **25** and **26**, the MS^2 spectra of the sodium adduct ion $[M + Na]^+$ at *m/z* 647.1952, gave product ions attributed to the neutral loss of 154.06 Da [M + H – anhydro-hydroxy-tyrosol]$^+$ at *m/z* 471.15, followed by a loss of 146.06 Da related to the loss of an anhydro-deoxyhexosyl residue observed at *m/z* 325.09. Further fragmentation led to the formation of a product ion at *m/z* at 163.04, suggesting the presence of an anhydro-caffeoyl moiety. The chemical structure of compound **25** was fully identified as acteoside (verbascoside) by comparison of its ^1H and ^{13}C NMR data with those published in the literature (supporting information: Table S3 and Fig. S10-S14) (Li et al., 2005; Xie et al., 2012). Low amounts of **26**, did not allow its identification by means of NMR spectroscopy, however, based on its identical fragmentation pattern in comparison with acteoside (**25**), it was tentatively characterized as its structural isomer isoacteoside (isoverbascoside) in which the caffeoyl moiety is attached to C-6 instead of C-4 of the central glucose.

Compounds **27** (*m/z* 661.2105 $[M + Na]^+$) and **28** (*m/z* 675.2272 $[M + Na]^+$) produced a similar fragmentation pattern as observed for acteoside (**25**) and isoacteoside (**26**) characterized by a major loss of 168.08 Da [M + H – anhydro-methoxy-hydroxy-tyrosol]$^+$ and consecutive cleavage of an anhydro-deoxyhexose (loss of 146.06 Da). Additional fragmentation in the MS^2 spectra of **27** and **28** yielded product ions at *m/z* at 163.04 and 177.05, suggesting that the main difference between these two compounds is the presence of a feruloyl residue in **28** and a caffeoyl unit in **27** at C-6 of the central glucose. Compounds **27** and **28** were tentatively characterized as cistanoside C and D,

respectively. To our knowledge, this is the first time that verbascoside (**25**) is indentified and **26-28** are tentatively characterized in the fruit of *C. cujete.*

2.3. Flavonoid glycosides. A detailed inspection of the molecular networking indicated the presence of three flavonoid glycoside clusters: FG1, FG2 and FG3 (Fig.3). As mentioned for the phenylethanoid derivatives, FG1 was observed as a heterogeneous cluster constituted by compounds containing isobaric substituents, such as hexosyl and caffeoyl residues. Five *O*-glycosyl flavonoids were detected in FG1 (Fig. 2A): **29** (*m/z* 447.0930), **30** (*m/z* 449.1087), **34** (*m/z* 433.1136), **36** (*m/z* 447.0927) and **41** (*m/z* 417.1192). MS^2 analysis of these five compounds produced aglycone ions (Y_0^-), attributed to the elimination of an anhydrohexose (loss of 162.05 Da) (see Table 1). Additional MS/MS fragments were not observed. Compounds **30**, **34**, **36** and **41** were tentatively identified as eriodictyol *O*-hexoside, naringenin *O*-hexoside, luteolin *O*-hexoside and pinocembrin *O*-hexoside, respectively. Compound **29** was unambiguously identified as luteolin 7-*O*-glucoside by comparison of its retention time and MS^2 spectra with that of an authentic reference standard.

Cluster FG2, which was also found to be related to flavonoid glyco-conjugates, contained di-*O*-glycosyl flavones and flavanones (Fig. 2A). Compounds **32**, **33** and **39** revealed molecular ions $[M - H]^-$ at *m/z* 577.1556, 579.1711 and 563.1765, respectively. The characteristic aglycone ions (Y_0^-) at *m/z* 269.04 (**32**), 271.06 (**33**) and 255.06 (**39**), which were proposed as deprotonated apigenin, naringenin and pinocembrin respectively, were formed by the neutral losses of 308.11 Da, attributed to the successive cleavage of a deoxyhexose and a hexose moiety from their corresponding parent ions in negative ion mode. Furthermore, compound **33** produced specific product ions at *m/z* 459.13 and 313.07, were formed by the loss of 120.03 Da $[M - H - C_8H_8O]^-$ and a successive loss of a deoxyhexose moiety $[M - H - C_8H_8O - 146.06]^-$. The loss of C_8H_8O is related to the retro Diels-Alder (RDA) reactions of the C–C bond at positions 1/3 of the C-ring ($^{1,3}A_0$) (Cuyckens and Claeys, 2004; Vukics and Guttman, 2010) (Fig. 2B). This fragmentation pattern of compound **33** was identical to that reported for naringin (Zeng et al., 2018; Zhao et al., 2013). From the previous results, compounds **32**, **33** and **39** were tentatively identified as rhoifolin, naringin and pinocembrin *O*-deoxyhexoside-*O*-hexoside, respectively.

The presence of methoxy-flavonoids was also observed in the cluster FG3 from the molecular networking (Fig. 2A). Compounds **31**, **35** and **37** showed deprotonated molecules at *m/z* 447.1033

[M – H]⁻, 461.1088 [M – H]⁻ and 491.1193 [M – H]⁻, respectively. Their product ions at *m/z* 315.05, 299.05 and 329.07 (loss of 162.05 Da), resulted from an anhydrohexose cleavage (Y_0^-). Further fragmentation led to the formation of product ions at [M – H – CH_3]⁻ and [M – H – hexose – CH_3]⁻ of all three compounds, which are characteristic fragments formed by the loss of a methyl group (15.02 Da). These product ions were not observed in the MS/MS spectra of flavonoids from cluster FG1 and FG2. Based on their fragmentation pattern, compounds **31**, **35** and **37** were tentatively characterized as methoxy-hydroxy-luteolin-*O*-hexoside, methoxy-luteolin-*O*-hexoside and dimethoxy-hydroxy-luteolin-*O*-hexoside, respectively.

In addition, four aglycones were also detected in the fruit extract of *C. cujete*, which showed [M – H]⁻ ions at *m/z* 285.0403 (**38**), 287.0559 (**40**), 255.0663 (**42**) and 271.0613 (**43**) in the full MS spectra. Analysis of the MS^2 data revealed $^{0,3}A^-$ fragments related to RDA reaction (Cuyckens and Claeys, 2004; Vukics and Guttman, 2010): 151.00 [M – H – C_8H_6O]⁻ (**38**), 151.00 [M – H – $C_8H_8O_2$]⁻ (**40**), 151.00 [M – H – C_8H_8]⁻ (**42**) and 151.00 [M – H –120 (C_8H_8O)]⁻ (**43**), suggesting the presence of luteolin, eriodictyol, pinocembrin and naringenin, respectively. By comparison with reference standards, **38** and **43** were unambiguously identified as luteolin and naringenin.

2.4. Benzoyl and cinnamoyl derivatives. A detailed inspection of cluster BC (see supplementary information Fig. S2A) revealed that the metabolites assigned to this cluster were esterified with two different phenolic acids: benzoic acid or cinnamic acid. Five nodes (**18**, **50**, **52**, **54** and **55**) were assigned as benzoyl derivatives since the MS^2 spectra of their parent ion ([M – H]⁻) revealed a main product ion due to the neutral loss of 122.03 Da ([M – anhydro-benzoyl]⁻) and the presence of a fragment at *m/z* 121.03, attributed to an anhydrobenzoyl moiety. Similarly, eight compounds (**21**, **24**, **59**, **60**–**64**) within this cluster were assigned as cinnamoyl derivatives due to the neutral loss of 148.05 Da ([M – anhydro-cinnamoyl]⁻ and consecutive presence of a product ion at *m/z* 147.04, which were formed by the loss of a cinnamoyl substituent (Fig. S2B).

Compound **62** and **64** showed a deprotonated molecular ion [M − H]⁻ at *m/z* 471.1513, its formate adduct [M – H + COOH]⁻ at *m/z* 517.1565, and a molecular formula of $C_{21}H_{28}O_{12}$ (Table 1). The MS/MS spectra of their parent ions exhibited product ions at *m/z* 323.09 and 147.04, characteristic of cinnamoyl di-hexosides. Compound **62** was unequivocally identified as sibirioside A by comparing its NMR spectral data with those reported in the literature (Li et al., 2014) (supporting

information: Table S4 and Fig. S15-S16). Compound **64**, present in a relatively low amount in comparison with **62**, was tentatively characterized as a structural isomer of sibirioside A.

Furthermore, comparing the NMR spectral data with those in the literature, compounds **65** and **66**, also described as cinnamoyl derivatives, were unambiguously identified as 1-*O*-*trans*-cinnamoyl-*β*-D-glucopyranose (Michodjehoun-Mestres et al., 2009) and *trans*-cinnamic acid (Davidse et al., 1990), respectively (supporting information: Table S4 and Fig. S17-S19). As previously reported by Pereira et al., 2017, our results pointed out that cinnamic acid was found as a major compound from the ethanolic extract of *C. cujete* fruit pulp. Furthermore, compounds **21** and **24** were assigned as cinnamoyl-containing iridoid derivatives and are described in detail in the iridoid glycoside section.

2.5. Iridoid glycosides. All iridoids were present in the form of glucosides, and some of them were either esterified with phenolic acids, such as benzoic, hydroxyl-benzoic, methoxy-hydroxy-benzoic, cinnamic and ferulic acids (Table 1). Compounds **15** (*m/z* 481.1350), **16** (*m/z* 511.1457), and **22** (*m/z* 537.1610), were observed in cluster IG1 (Fig. S3A). Their MS^2 spectra showed main product ions Y_0^- at *m/z* 319.08 (**15**), 349.09 (**16**) and 375.11 (**22**), obtained by the neutral loss of an anhydrohexose $[M - H - 162.05]^-$, a typical fragmentation pattern observed for most iridoid glucosides (Friščić et al., 2016). The hemiacetal group (characteristic in catalpol derivatives), was easily isomerized into two aldehydes groups (see Fig. S3B). Consecutive losses of 114.03 Da lead to the formation of product ions *m/z* 205.05 (**15**), 235.07 (**16**) and 261.08 (**22**), attributed to the cleavage of the cyclopentane ring and subsequent loss of one aldehyde group ($^{6,9}X^-$) ion (Hong et al., 2007). Additional characteristic product ions (X_0^-) in MS^2, observed at *m/z* 137.02 (**15**), 167.03 (**16**) and 193.05 (**22**), revealed the presence of hydroxybenzoyl (**15**), methoxy-hydroxybenzoyl (**16**) and feruloyl (**22**) residues. According to the MS/MS fragmentation pattern of catalpol derivatives previously described (Hong et al., 2007), **15**, **16** and **22** were tentatively characterized as catalposide, picroside II and feruloyl catalpol, respectively. In addition, compound **10** (*m/z* 361.1137) was tentatively characterized as catalpol (7,8-epoxy-aucubin). It is suggested that hexose residues (glucose) are attached to the aglycone through glycosidic bonds at the *O*-1 position and that these compounds were 6-*O*-phenolic acid esters of catalpol by comparison to similar compounds previously isolated from *C. cujete* (Kaneko et al., 1997; Wang et al., 2010). It is also

suggested, that the presence of catalpol derivatives may be related to its biosynthetic precursor aucubin (Kumar et al., 2013) in the fruit of *C. cujete*.

A detailed analysis of cluster IG2 showed that compounds **17** (*m/z* 465.1401)**, 20** (*m/z* 495.1509) and **23** (*m/z* 521.1662), were a group of iridoid glycosides similar to those compounds from cluster IG1 (Fig. S3A). These compounds produced characteristic ions Y_0^- at *m/z* 303.09 (**17**), 333.10 (**20**) and 359.11 (**23**); and X_0^- at *m/z* 137.02 (**17**), 167.03 (**20**) and 193.05 (**23**). In addition, the MS^2 spectrum of **17**, **20** and **23** revealed the same product ion at *m/z* 165.05 [M – H – anhydrous glucose – phenolic acid]$^-$, which allowed the tentative identification of aucubigenin as its iridoid aglycone. Compounds **17**, **20** and **23** were thus tentatively identified as hydroxy-benzoyl aucubin, methoxy-hydroxy-benzoyl aucubin and feruloyl aucubin.

Additionally, fractionation and further purification of *C. cujete* extract resulted in the isolation of 6 iridoid glycosides and their structures were elucidated by 1D and 2D NMR experiments. Among these, three undescribed compounds were identified (**13**, **21** and **24**) (Fig. 3). The three known compounds were assigned as *epi*-aucubin (**11**) (Bianco and Passacantilli, 1982), aucubin (**12**) (Ersöz et al., 2002) and ajugol (**14**) (Nykmukanova et al., 2017) by comparing their NMR spectral data to those reported in the literature (supporting information: Table S5 and Fig. S20-S37).

Compounds **11** (*epi*-aucubin) was isolated as the main iridoid glycoside from the fruit pulp of *C. cujete*. Besides, an inseparable mixture of compound **13** and **11** was obtained as a white amorphous powder in a ratio of about 1:1 as estimated by signal intensities in ^{13}C-NMR spectra and UPLC-DAD analysis. The structure elucidation of **13** was performed as a mixture. The NMR data for **13** were assigned based on the exclusion of the NMR signals and correlations assigned to pure *epi*-aucubin (**11**) (Table 3 and supporting information: Fig. S38-S44). The formula of **13** was determined to be $C_{15}H_{22}O_9$ by ESI-HRMS in negative ionization mode, displaying deprotonated molecule peaks at *m/z* 345.1190 [M – H]$^-$ (calculated for $C_{15}H_{21}O_9$ [M – H]$^-$, 345.1195) and *m/z* 391.1245 [M – H + FA]$^-$ (calculated for $C_{16}H_{23}O_{11}$ [M – H + FA]$^-$, 391.1246). This formula indicated 5 degrees of unsaturation in the molecule. The ^{13}C-NMR and DEPT spectra showed 15 carbon resonances, which were identified as two methylenes, twelve methines, and one carbon with no attached hydrogen. From these resonances, a glucosyl unit and nine carbon signals for the aglycone moiety, which were consistent with an iridoid skeleton, were revealed. The ^{13}C-NMR

and HMBC spectrum also revealed four olefinic carbons at δ_C 139.2, 106.0, 135.8 and 135.2 assigned to C-3, C-4, C-6 and C-7. The molecular formula, degree of unsaturation and NMR data suggested that compound **13** possesses a monotropein-like structure (without the methyl ester group attached at C-4), with two double bonds at C-3 and C-6. Analysis of the ^{1}H-NMR spectrum of compound **13** showed an anomeric proton resonance at δ_H 4.65 (d, J = 7.9 Hz), as well as the three-bond HMBC correlations from H-1′ to C-1 and from H-1 to C-1′, which provided evidence for the presence of the glucosyl unit at the C-1 position of the aglycone.

In the COSY spectrum, H-9 (δ_H 2.50, dd, J = 8.4, 2.4 Hz) correlated to H-5 (δ_H 3.44, m) and H-1 (δ_H 5.60, d, J = 2.4 Hz). The absence of any other cross-peak for H-9 suggested an oxygenated carbon with no attached hydrogen at δ_C 86.6 (C-8). In the HMBC spectrum, correlations of proton signal of two germinal oxy-methylene groups ascribed at H-10 (δ_H 3.73, d, J = 11.4 Hz and 3.61, d, J = 11.4 Hz) to C-7, C-8 and C9, indicated the connection of C-10 to C-8. Based on the above information, compound **13** was assumed to be an iridoid glycoside, similar to eranthemoside previously isolated from *Eranthemum pulchellum* (Fischer et al., 1987). A well-known effect for the configuration of C-8 epimers previously observed in the ^{13}C NMR data of gardenoside and monotropein methyl ester (Chaudhuri et al., 1980) was evidenced by compound **13** and eranthemoside. The β-oriented OH group at C-8 in compound **13** was determined from the comparison of its ^{13}C NMR data with those reported for eranthemoside (Fischer et al., 1987; Kanchanapoom et al., 2004). It was observed that an 8β-hydroxy substituent in compound **13** exerts a deshielding effect of ~7 ppm to C-9 in respect to eranthemoside (Chaudhuri et al., 1980). Additionally, a different configuration of compound **13** at C-8 affected C-6 and C-7 chemical shifts, inducing deshielding (~2 ppm) and shielding (3 ppm), respectively.

The relative configuration of **13** was determined by NOESY experiments and comparison of NMR data with those of gardenoside (Pachaly and Klein, 1987). Since a NOESY correlation was observed between H-5 and H-9 (supporting information: Fig. S45), located on the same side of the molecule, a correlation is also to be expected between H-9 and H-10 if they were on the same side as well. Since this is not the case, the -CH_2-OH substituent (C-10) was tentatively positioned in α-orientation, and the OH substituent at C-8 in β-orientation. The β-configuration of H-5 and H-9, and the α-orientation of H-1 of compound **13** are suggested to be similar to that of gardenoside by comparison of the coupling constant of H-1 ($J_{1,9}$ = 2.4 Hz), H-6 ($J_{6,5}$ = 2.8 Hz), and H-9 ($J_{9,1}$ = 2.4

Hz) (Table 2). In addition, the relatively large coupling constant of H-1′ and H-2′ ($J_{1',2'}$ = 7.9 Hz) indicated a β-configuration of the glucosyl residue. Based on the above evidence, the structure of **13** was determined to be 8-*epi*-eranthemoside. Finally, *epi*-eranthemoside, similar to eranthemoside (Fischer et al., 1987), appears to be unstable and turns brownish a few days after its purification.

Compounds **21** and **24**, both found in the cluster BG, which is related to cinnamoyl derivatives (Fig. S2), were identified as a pair of isomers and as a yellow amorphous solid. Even though these two compounds were clearly separated and isolated from the semi-preparative HPLC system as individual constituents, they were obtained as two separate mixtures of isomers in a ratio of 4:1 (**21**:**24**) and 1:2 (**21**:**24**), respectively, as established by signal intensities in ^{13}C-NMR spectra. This may indicate a possible spontaneous epimerization after isolation.

Compounds **21** and **24** were assigned the molecular formula $C_{24}H_{32}O_{10}$ based on the deprotonated molecule at *m/z* 479.1918 $[M - H]^-$ (calculated for $C_{24}H_{31}O_{10}$ $[M - H]^-$, 479.1923) and *m/z* 525.1976 $[M - H + FA]^-$ (calculated for $C_{25}H_{33}O_{12}$ $[M - H + FA]^-$, 525.1977), obtained by HR-ESI-QTOF-MS in negative ionization mode. This molecular formula indicated nine degrees of unsaturation. The structural elucidation of both compounds from the mixture was possible since well-separated key chemical shifts were observed in the ^{1}H and ^{13}C NMR spectra (see Table 2 and 3, supporting information S46-S59). Compounds **21** and **24** revealed signals characteristic of a 3-hydroxyethyl-2,5-hydroxymethyl-3-hydroxy-cyclopentene unit in the molecule, with one olefinic, two methines, one hydroxyl-methine, one methylene, three hydroxyl-methylene groups and one carbon with no attached hydrogen. In addition to the nine carbon signals due to the aglycone, six chemical shift values corresponding to a glucopyranosyl group and nine carbon signals attributed to a *trans*-cinnamoyl residue were observed in the ^{13}C NMR spectra (Table 3).

Inspection of the ^{1}H-NMR spectra of compounds **21** and **24** revealed a high relative *J* coupling constant (*J* = 7.8 Hz) of the anomeric proton at δ_H 4.31, suggesting a β-configuration of the glucose residue. In addition, three-bond HMBC correlations (see supplementary information Fig. S4, S52 and S59) from H-1′ to C-3 and from H-3 to C-1′, confirmed the location of the glucosyl group at the C-3 position of the aglycone. The HMBC spectra of both compounds showed correlations from

the proton at δ_H 4.51/4.35 (H-6′a/ H-6′b) to the carbonyl carbon (δ_C 168.5, C-9) revealing the presence of the *trans*-cinnamoyl group located at C-6′ of the glucosyl moiety.

Moreover, ^{1}H and ^{13}C NMR spectral data of the aglycone of **24** were similar to those reported for isoeucommiol (Bianco et al., 1981), which has a *β*-oriented OH group at C-6. When comparing the chemical shifts of the carbon nuclei of **21** with those of **24**, we observed a characteristic long-ranged effect associated with 6-hydroxy epimeric substitution (Chaudhuri et al., 1980; Damtoft et al., 1981), similar to that reported for aucubin and *epi*-aucubin in literature (Bianco and Passacantilli, 1982) and also observed in this study (see supporting information: Table S5). It was evident that a 6*α*-hydroxy substituent in compound **21** caused a shielding effect of ~3 ppm and ~7 ppm to C-4 and C-6, respectively, with respect to compound **24**. Additionally, a difference of ~3 ppm of the chemical shift of C-8 was also observed for these two compounds. Taken together all data, the difference between **21** and **24** must reside at the hydroxyl configuration at C-6 position. The relative configuration of the cyclopentane ring for compounds **21** and **24** is suggested to be H-5/H-9 *β*-oriented since this is the most common for structural naturally occurring iridoids (Dinda et al., 2007). Moreover, because of biogenic considerations, *β*-configurations at C-5 and C-9 were assumed identical to *epi*-aucubin (**11**) and aucubin (**12**). In conclusion, compounds **21** and **24** were assigned as 3-(6′-*trans*-cinnamoyl-*β*-D-glycopyranosyl)-6-*epi*-isoeucomiol and 3-(6′-*trans*-cinnamoyl-*β*-D-glycopyranosyl)-isoeucomiol, named crescentiol A and B, respectively. These two compounds were found to be closely related to crescentin III, 6-*O*-*p*-hydroxybenzoyl-10-deoxy-eucommiol and 6-benzoyl-10-deoxyeucommiol, which were previously reported for *C. cujete* (Kaneko et al., 1997; Wang et al., 2010).

Biological activities have been reported for several major compounds, including phenolic acids, iridoids and phenylethanoids, suggesting them as the potential active constituents. For example, *trans*-cinnamic acid (**66**), catalpol (**10**), aucubin (**12**) and ajugol (**14**), exhibited anti-microbial and anti-inflammatory properties (Dong et al., 2011; Kim et al., 2000; Kupeli et al., 2007; Letsididi et al., 2018; Schink et al., 2018). Similarly, verbascoside (**25**) has been reported for its anti-microbial and anti-inflammatory properties (Frezza et al., 2018; Pardo et al., 1993). The presence of verbascoside together with the iridoids observed in *C. cujete* suggested a significant chemotaxonomic implication, considering that these two kinds of molecules showed concurrent limited distribution within the order Lamiales. This characteristic composition has been found in

species of Bignoniaceae and other related families (Martin et al., 2007), such as *Teucrium chamaedrys* (Lamiaceae) (Frezza et al., 2018), species of the genus *Verbascum* (Scrophulariaceae) (Akkol et al., 2007; Kahraman et al., 2018), and of the genus *Pedicularis* (Orobanchaceae) (Zhongjian et al., 1992; Zhongjian and Zimin, 1992; Zimin and Zhongjian, 1991). Noteworthy is the fact that the leaves and flowers of *Verbascum* species (which are prepared as a water infusion in Turkish folk medicine) (Akkol et al., 2007; Kahraman et al., 2010, 2018), and the fruit of *C. cujete* are both employed for relieving respiratory conditions such as colds, bronchitis and asthma. Accordingly, it is suggested that their similar phytochemical composition in terms of phenylethanoids and iridoids could explain (at least in part) their related medicinal properties.

3. CONCLUSION

The untargeted metabolomics approach combined with UPLC-MS/MS-based molecular networking, isolation and NMR analysis has proven to be a useful strategy to facilitate the phytochemical profiling of the fruit extract of *C. cujete*. Besides, manual annotation and comparison with reference standards of analogues for unequivocal identification significantly improved the dereplication efficiency of the molecular network. Our results revealed that the fruit of *C. cujete* contains a variety of constituents, and in particular *n*-alkyl glycosides, phenylethanoids, flavone and flavanone glycosides, phenyl acid derivatives and iridoids glycosides.

The findings of this study provide a comprehensive description of the phytochemical composition of the fruit of *Crescentia cujete.* This approach established a useful starting point for the selection of *trans*-cinnamic acid (**66**), benzoic acid (**49**), 6-epi-aucubin (**11**), aucubin (**12**), ajugol (**14**), verbascoside (**25**) as appropriate chemical markers. The development of analytical methods for the quality control of marketed commercial products containing this plant species is being conducted by our research group.

4. EXPERIMENTAL SECTION

4.1. Reagents. Methanol (MeOH) (HPLC grade), ethyl acetate (EtOAc) (HPLC grade), acetonitrile (ACN) (HPLC grade), chloroform (HPLC grade) and absolute ethanol (EtOH)

(analytical reagent grade) were acquired from Fisher Scientific (Leicestershire, UK). Formic acid (FA) (98+%, pure, analytical reagent grade) was obtained from Acros Organics (Geel, Belgium). Formic acid (FA) (99%, UPLC/MS grade) for LC-MS analysis was acquired from Biosolve Chimie (Dieuze, France). Ultrapure water with a resistivity of 18.2 MΩ·cm at 25 °C (Milli-Q, Waters, Milford, MA, USA) was used as the extraction solvent and for mobile phase preparation. For the NMR experiments, DMSO-d_6 (99.9% atom D) and methanol-d_4 (99.8% atom D) were purchased from Sigma-Aldrich USA (St. Louis, MO, USA).

Protocatechuic acid (97%), naringenin (98%) and *trans*-cinnamic acid (97%) were obtained from Sigma-Aldrich (St. Louis, MO, USA). Sucrose (cristal, 99%) was purchased from Fisher Scientific (Geel, Belgium), and luteolin (98%) from Santa Cruz Biotechnology (Dallas, Texas, USA). Luteolin-7-*O*-glucoside (95%) was acquired from Extrasynthese (Genay, France) and 4-hydroxy-benzoic acid (99%) from Sigma-Aldrich Chemie GmbH (Steinheim, Germany). All these compounds were used as reference standards.

4.2. General experimental procedures

Optical rotations were recorded on a Jasco P-2000 spectropolarimeter (Easton, MD, USA) at 20 °C with Spectra Manager™ software. The samples were dissolved in MeOH and specific rotation was determined at 589 nm with a path length of 50 mm.

NMR spectra were recorded on a Bruker DRX-400 instrument equipped with either a 3 mm inverse broadband (BBI) probe or a 5 mm dual $^{1}H/^{13}C$ probe using standard Bruker pulse sequences and operating at 400 MHz for ^{1}H and at 100 MHz for ^{13}C NMR spectra. The spectra were processed with Topspin version 1.3 and ACD/NMR Processor Academic Edition (version 12.01, Advance Chemistry Development, Inc.). NMR spectra were recorded in methanol-d_4. Raw and processed NMR data is available at https://doi.org/10.7910/DVN/MFB4KI.

TLC was performed on pre-coated silica gel F_{254} plates (Merck, Darmstadt, Germany), and the bands were observed under UV light (254 and 366 nm). TLC was performed with a mobile phase of EtOAc : MeOH : H_2O (70 : 25 : 5). TLC plates were examined under UV lights, and then sprayed with vanillin reagent (1 g of vanillin in 950 mL of ethanol 96% + 5 L of concentrated

sulphuric acid) and subsequently heated at 110 °C for 5 min. The TLC plates were then inspected visually.

The first isolation steps were performed by column chromatography with MCI Gel® CHP20P, a highly porous styrene-divinylbenzene polymer resin, particle size 75-150 µm (Supelco, Mitsubishi Chemical Corp., Bellefonte, USA). Silica gel particle size 40-63 µm (Merck, Darmstadt, Germany) was used as a stationary phase for F6 fractionation.

Semi-preparative HPLC system. The system is comprised of a sample manager, injector and collector (2767), a quaternary gradient module (2545), a System Fluidics Organizer, an HPLC-pump (515), a DAD (2998) and a Micromass Quattro mass spectrometer with TQD, all supplied by Waters (Milford, MA, USA). This system was equipped with a Kinetex® C_{18} column (250 x 10.0 mm, 5 µm) (Phenomenex, Torrance, CA, USA). As the mobile phase H_2O + 0.1% FA (A) and ACN + 0.1% FA (B) or MeOH + 0.1% FA (C) at a flow rate of 3.0 mL/min were used. A make-up flow containing 80% MeOH + 0.1% FA was employed. Detection was performed at negative ion mode under the following conditions: capillary voltage, 3.50 kV; Cone voltage, 50 V; extractor voltage, 3 V; source temperature, 140 °C; desolvation temperature, 400 °C; desolvation gas flow, 850 L/h. The UV detector was set at 280 nm. MassLynx version 4.1 was used to process the data.

During the runs, a gradient was set as follows: a) For sub-fraction $F2_E$ (min/B%): 0.0/4, 15.0/4, 25.0/10, 40.0/20, 45.0/100, 50.0/4, 60.0/4. b) For sub-fraction $F2_F$ (min/C%): 0.0/5, 12.0/5, 20.0/10, 40.0/20, 55.0/50, 65.0/100, 70.0/100, 75.0/5, 85.0/5. c) For sub-fractions $F6_{18\text{-}19}$ (min/B%): 0.0/10, 5.0/10, 15.0/22, 25.0/22, 30.0/100, 35.0/100, 40.0/10, 50.0/10. d) For sub-fraction $F6_{t26}$ (min/B%): 0.0/10, 5.0/10, 10.0/25, 30.0/25, 35.0/100, 40.0/100, 45.0/10, 55.0/10. e) For subtraction $F7_{EtOAc}$ (min/B%): 0.0/10, 5.0/10, 15.0/15, 20.0/15, 30.0/19, 50.0/19, 55.0/100, 60.0/100, 65.0/10, 75.0/10.

4.3. Plant material. The fruits of *Crescentia cujete* were collected in October 2015 in Panama Oeste, Republic of Panama (GPS 8°48'34" N, 79°53'31" W). The taxonomical classification was carried out by the botanist O.O. Ortiz (one of the co-authors) and deposited at the Herbarium of the University of Panama (PMA) (voucher specimen Ortiz, O.O. 2524). The plant material was freeze-dried in a lyophilizer (Labconco FreeZone 18 L, Kansas City, MO, USA).

4.4. Extraction and isolation. The dried fruits (740.0 g) were extracted with ethanol 70% (3 L x 4) at room temperature. The combined extract was concentrated under reduced pressure below 40 °C to give a crude residue (467.7 g). A portion of the dried extract (160.0 g) was suspended in 300 mL of H_2O and subjected to 400 g of MCI gel, using H_2O:MeOH (v/v) mixtures with increasing organic solvent concentrations (F1: 100:0; F2: 90:10; F3: 80:20; F4: 70:30; F5: 60:40; F6: 50:50, F7: 40:60; F8: 30:70; F9: 20:80, F10: 10:90; F11: 0:100). All fractions were dried under reduced pressure. One part was used for isolation and the other part was employed for UPLC-HRMS analysis.

4.5. UPLC-MS/MS analysis for molecular networking. For LC-MS analysis, the crude extract and fractions were dissolved in water:methanol (60:40) at a concentration of 1.0 mg/mL. Samples were further diluted with water to obtain a final concentration of 0.1 mg/L. Analyses were performed on an ACQUITY LC system (Waters, Milford, MA, United States) equipped with an Acquity UPLC Peptide BEH C_{18} column (2.1 × 100 mm; 1.7 µm, Waters, Milford, MA, United States). Elution was conducted with a mobile phase consisting of H_2O + 0.1% FA (A) and ACN (B), which were pumped at a rate of 0.4 mL/min. The gradient program was set as follows: 2% B (0-1 min), 2-26% B (1-10 min), 26-65% B (10-16 min), 65-100% B (16-18 min), 100% B (18-20 min), 100-2% B (20-21 min), 2% B (21-25 min). The column oven was set at 30 °C and the injection volume was 10 µL. UPLC-HR-MS/MS analyses were achieved by coupling the LC system to an orbitrap MS (Q ExactiveTM, Thermo Fisher Scientific, San Jose, CA, United States) equipped with an electrospray ionization source with Thermo Xcalibur software (version 3.0, Thermo Fisher Scientific, San Jose, CA, United States). The following parameters were used for data acquisition in negative ionization mode: capillary temperature at 350 °C, source voltage at 2.70 kV, source current at 100 µA, capillary voltage at –30 V, sheath gas flow rate at 40 L/min and auxiliary gas flow at 10 L/min. MS scans were operated in full scan mode from *m/z* 115 to 1000 with a mass resolution of 20 000 at *m/z* 922. MS^1 scan was followed by MS^2 scans of the five highest intensity ions above an absolute threshold of 3000 counts. Selected parent ions were fragmented at a fixed collision energy value of 35 eV and an isolation window of 1.3 amu.

4.6. Molecular Networking and Metabolomics Analysis. Raw data obtained from the UPLC–MS/MS system were converted to mzXML format using the ProteoWizard tool MSConvertGUI (version 3.0.18257-8e71123fb, Vanderbilt University, United States) (Kessner et al., 2008). The

MZmine workflow for feature-based molecular networking on GNPS was used. The MS data processing workflow comprises raw data file import, peak detection, peak list deisotoping, alignment and gap filling (Olivon et al., 2017; Pluskal et al., 2010). The MZmine parameter of the workflow for LC-MS was as follows: Feature extraction for MS^1 was achieved with a signal threshold of 1.0 $x10^4$, and 1.0 $x10^2$ for MS^2. A centroid algorithm was used. Chromatogram builder was run with a minimum height of 1.0 $x10^4$ and tolerance of 5 ppm. Chromatograms were deconvoluted with an *m/z* range for MS^2 pairing of 0.02 Da and retention time range for MS^2 of 0.05 min and a baseline cut-off algorithm of 1.0 $x10^3$. Detected peaks were aligned through Join Aligner Module considering mass (5 ppm) and retention time tolerance (0.1 min).

Processed files were uploaded to the GNPS website (http://gnps.ucsd.edu) and a molecular network was created with the feature-based molecular networking workflow (https://ccms-ucsd.github.io/GNPSDocumentation/featurebasedmolecularnetworking/) (Wang et al., 2016).The data were filtered by removing all MS/MS fragment ions within +/– 17 Da of the precursor *m/z*. MS/MS spectra were window filtered by choosing only the top 6 fragment ions in the +/- 50 Da window throughout the spectrum. The precursor ion mass tolerance was set to 0.05 Da and an MS/MS fragment ion tolerance of 0.1 Da. A network was then created where edges were filtered to have a cosine score above 0.6 and more than 1 matched peak. Further, edges between two nodes were kept in the network if, and only if each of the nodes appeared in each other's respective top 10 most similar nodes. Finally, the maximum size of a molecular family was set to 100, and the lowest-scoring edges were removed from molecular families until the molecular family size was below this threshold. The spectra in the network were then searched against the GNPS spectral libraries. The library spectra were filtered in the same manner as the input data. All matches kept between network spectra and library spectra were required to have a score above 0.7 and at least 1 matched peaks. The resulting molecular network is available at http://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=2eead4ea1dbc45838d209a9e9a6bc978. Automated compound annotation list is available at https://gnps.ucsd.edu/ProteoSAFe/result.jsp?task=3e4de316ba4f4ff795fb2a49573c9e16&view=view_all_annotations_DB. To visualize the data, the output was imported into Cytoscape version 3.6.1. (https://cytoscape.org/) (Shannon et al., 2003) and displayed using the settings “preferred layout” with “directed” style. Processed molecular network in Cytoscape is available at https://doi.org/10.7910/DVN/A45UIA. Manual annotation of peak ions was performed by

searching their predicted molecular formulae against databases such as Dictionary of Natural Products (DNP), Chemspider and SciFinder.

4.7. Purification of compounds. Fraction F2 (1.15 g) was submitted to flash chromatography on a Grace Reveleris X2 system (Columbia, MD, USA) using the Reveleris® Navigator™ software. The system was equipped with a binary pump, a UV detector set at 280 nm, a fraction collector, a pre-packed Flash Grace Reveleris silica cartridge (40 g) with particle size of 40 µm, and a mobile phase consisting of ethyl acetate (A) and methanol (B) at a flow rate of 25 mL/min. During the run, a gradient was set as follows (min/B%): a) For fraction F2: 0.0/0, 3.0/0, 13.0/10, 18.0/10, 28.0/20, 58.0/38, 64.0/50, 69.0/100, 75.0/100. In total, 120 test tubes (each containing 25 mL) were obtained. Tubes showing a similar pattern were combined based on TLC analysis. This resulted in nine fractions ($F2_A$- $F2_I$). Purification of $F2_E$ (266.6 mg) by semi-preparative HPLC (see general experimental procedures) afforded **11** (26.6 mg) and a mixture of **11**, **12** and **13** (7.8 mg). Similarly, $F2_F$ (310.7 mg) was purified to afford the previously isolated **11** (6.6 mg), an inseparable mixture of **11** and **13** (5.7 mg), **14** (17.3 mg), **2** (6.3 mg) and **3** (3.7 mg).

Fraction 6 (803.8 mg) was subjected to open silica gel CC (40.0 x 2.5 cm) using successive elutions of EtOAc–MeOH–H_2O (100:0:0, 92:5:3, 75:20:5, 55:40:5, 20:75:5, v/v/v) to yield 45 test tubes each containing 25 mL of solvent. Tubes showing a similar pattern were combined based on TLC analysis. This resulted in 28 fractions ($F6_1$- $F6_{28}$). Purification of $F6_{18}$ (28.8 mg) and $F6_{19}$ (16.7 mg) by semi-preparative HPLC (see general experimental procedures) afforded **25** (12.4 mg). Furthermore, $F6_{26}$ (118.0 mg) was purified to afford **62** (42.8 mg) and a mixture of **21** and **24** (4.3 mg) in a ratio of 4:1.

Fraction F7 (350.8 mg) was suspended in 20 mL of water and subsequently partitioned with chloroform and ethyl acetate to obtain $F7_{CHCl_3}$ (78.0 mg) and $F7_{EtOAc}$ (95.5 mg), respectively. $F7_{EtOAc}$ was purified by semi-preparative HPLC (see general experimental procedures) to afford a mixture of **21** and **24** (4.6 mg) in a ratio of 1:2.

Accurate mass measurements of the isolated compounds from the fruit of *C. cujete* were performed by using a Xevo G2-XS QTof spectrometer (Waters, Milford, MA, USA) coupled with an ACQUITY LC system equipped with MassLynx version 4.1 software. For analysis, 5 µL of samples were injected on an ACQUITY UPLC® BEH Shield RP18 column (100 mm × 2.10 mm,

1.7 μm, Waters, Milford, MA, USA). The mobile phase solvents consisted of H_2O + 0.1% FA (A) and ACN + 0.1% FA (B), and the gradient was set as follows (min/B%): 0.0/2, 1.0/2, 14.0/26, 24/65, 26.0/100, 29.0/100, 31.0/2, 36.0/2. The flow rate was 0.4 mL/min. The column oven was set at 40 °C. During the first analysis, full scan data were recorded in ESI− and ESI+ ionization mode from *m/z* 50 to 2000 and the analyzer was set in sensitivity mode (approximate resolution: 22,000 FWHM). The spray voltage was set either at +1.5 kV and –1.0 kV; cone gas flow and desolvation gas flow at 50.0 L/h and 1000.0 L/h, respectively; and source temperature and desolvation temperature at 120 °C and 550 °C, respectively. Data were also recorded using MS^e in the positive and negative ionization modes. Leucine-Encephalin was used as a lock mass. UV chromatograms were recorded at 200 and 280 nm.

*4.7.1. (2R,4S)-2-O-β-D-xylopyranosyl-(1→6)-β-D-glucopyranosyl-2,4-pentanediol (**2**).* Pale yellow amorphous powder; $[\alpha]_D^{25}$ = –92.6 (*c* 0.33, MeOH); HR-ESI-MS spectrum (negative ionization mode) displayed a molecular ion at *m/z* 397.1713 $[M - H]^-$ (calculated for $C_{16}H_{29}O_{11}$, 397.1715) and a FA adduct ion at *m/z* 443.1772 $[M - H + FA]^-$ (calculated for $C_{17}H_{31}O_{13}$, 443.1770); ^1H and ^{13}C NMR data: See supporting information (Table S2).

*4.7.2. (2R,4S)-2-O-β-D-glucopyranosyl-2,4-pentanediol (**3**).* Pale yellow amorphous powder; $[\alpha]_D^{25}$ = –99.7 (*c* 0.61, MeOH); HR-ESI-MS spectrum (negative ionization mode) displayed a molecular ion at *m/z* 265.1283 $[M - H]^-$ (calculated for $C_{11}H_{21}O_7$, 265.1293) and a FA adduct ion at *m/z* 311.1337 $[M - H + FA]^-$ (calculated for $C_{12}H_{23}O_9$, 311.1348); ^1H and ^{13}C NMR data: See supporting information (Table S2).

*4.7.3. Epi-aucubin (**11**).* White amorphous powder; $[\alpha]_D^{25}$ = –72.1 (*c* 2.37, MeOH); HR-ESI-MS spectrum (negative ionization mode) displayed a molecular ion at *m/z* 345.1188 $[M - H]^-$ (calculated for $C_{15}H_{21}O_9$, 345.1191) and a FA adduct ion at *m/z* 391.1244 $[M - H + FA]^-$ (calculated for $C_{16}H_{23}O_{11}$, 391.1246); ^1H and ^{13}C NMR data: See supporting information (Table S5).

*4.7.4. Ajugol (**14**).* Colourless amorphous solid; $[\alpha]_D^{25}$ = –188.7 (*c* 0.90, MeOH); HR-ESI-MS spectrum (negative ionization mode) displayed a molecular ion at *m/z* 347.1343 $[M - H]^-$ (calculated for $C_{15}H_{23}O_9$, 347.1348) and a FA adduct ion at *m/z* 393.1397 $[M - H + FA]^-$

(calculated for $C_{16}H_{23}O_{11}$, 393.1402); ^{1}H and ^{13}C NMR data: See supporting information (Table S5).

*4.7.5. Epi-aucubin (**11**) and epi-eranthemoside (**13**).* White amorphous powder; $[\alpha]_D^{25}$ = –195.9 (*c* 0.57, MeOH); HR-ESI-MS spectrum (negative ionization mode) for **4** displayed a molecular ion at *m/z* 345.1190 $[M - H]^-$ (calculated for $C_{15}H_{21}O_9$, 345.1191) and a FA adduct ion at *m/z* 391.1245 $[M - H + FA]^-$ (calculated for $C_{16}H_{23}O_{11}$, 391.1246); ^{1}H and ^{13}C NMR data: See Table 2 and 3.

*4.7.6. Crescentiol A (**21**) and Crescentiol B (**24**) ratio of 4:1.* Pale yellow amorphous solid; $[\alpha]_D^{25}$ = –71.9 (*c* 0.21, MeOH); HR-ESI-MS spectrum (negative ionization mode) for **5** displayed a molecular ion at *m/z* 479.1918 $[M - H]^-$ (calculated for $C_{24}H_{31}O_{10}$, 479.1923) and a FA adduct ion at *m/z* 525.1976 $[M - H + FA]^-$ (calculated for $C_{25}H_{33}O_{12}$, 525.1967); ^{1}H and ^{13}C NMR data: See Table 2 and 3.

*4.7.7. Crescentiol A (**21**) and Crescentiol B (**24**) ratio of 1:2.* Pale yellow amorphous solid; $[\alpha]_D^{25}$ = –71.2 (*c* 0.46, MeOH); HR-ESI-MS spectrum (negative ionization mode) for **6** displayed a molecular ion at *m/z* 479.1923 $[M - H]^-$ (calculated for $C_{24}H_{31}O_{10}$, 479.1923) and a FA adduct ion at *m/z* 525.1976 $[M - H + FA]^-$ (calculated for $C_{25}H_{33}O_{12}$, 525.1967); ^{1}H and ^{13}C NMR data: See Table 2 and 3.

*4.7.8. Acteoside (verbascoside) (**25**).* Yellow amorphous powder; $[\alpha]_D^{25}$ = –85.0 (*c* 0.57, MeOH); HR-ESI-MS spectrum (negative ionization mode) displayed a molecular ion at *m/z* 623.1974 $[M - H]^-$ (calculated for $C_{29}H_{35}O_{15}$, 623.1981); ^{1}H and ^{13}C NMR data: See supporting information (Table S3).

*4.7.9. Sibirioside A (**62**).* Pale yellow amorphous powder; $[\alpha]_D^{25}$ = +46.8 (*c* 2.26, MeOH); HR-ESI-MS spectrum (negative ionization mode) displayed a molecular ion at *m/z* 471.1508 $[M - H]^-$ (calculated for $C_{21}H_{27}O_{12}$, 471.1508) and a FA adduct ion at *m/z* 517.1561 $[M - H + FA]^-$ (calculated for $C_{22}H_{29}O_{14}$, 517.1563); ^{1}H and ^{13}C NMR data: See supporting information (Table S4).

*4.7.10. 1-O-trans-cinnamoyl-β-D-glucopyranose (**65**).* Pale yellow amorphous powder; $[\alpha]_D^{25}$ = +10.7 (*c* 0.47, MeOH); HR-ESI-MS spectrum (negative ionization mode) displayed a molecular ion at *m/z* 309.0970 $[M - H]^-$ (calculated for $C_{15}H_{17}O_7$, 309.0969) and a FA adduct ion at *m/z*

355.1026 [M – H + FA]$^-$ (calculated for $C_{16}H_{19}O_9$, 355.1024); ^{1}H and ^{13}C NMR data: See supporting information (Table S4).

ACKNOWLEDGMENTS

The authors gratefully acknowledge the scholarship received for Mr. Rivera's doctoral studies from the National Secretariat for Science, Technology and Innovation (SENACYT) and the Institute for Training and Development of Human Resources (IFARHU) of the Republic of Panama. The authors also thank SENACYT for financial support to Dr. Caballero-George through the incentive program of the National Investigation System (SNI) as well as through grant FID14-116. We would like to thank the MediHealth Project for providing Mr. Rivera with a secondment opportunity at the Department of Pharmacognosy and Natural Products Chemistry of the Kapodistrian University of Athens (Greece).

APENDIX A. SUPPLEMENTARY DATA

Supplementary data to this article can be found online at

REFERENCES

Akkol, E.K., Tatli, I.I., Akdemir, Z.S., 2007. Antinociceptive and anti-inflammatory effects of saponin and iridoid glycosides from *Verbascum pterocalycinum* var. *mutense* Hub.-Mor. Zeitschrift fur Naturforsch. C 62, 813–820. https://doi.org/10.1515/znc-2007-11-1207

Allard, P.M., Péresse, T., Bisson, J., Gindro, K., Marcourt, L., Pham, V.C., Roussi, F., Litaudon, M., Wolfender, J.L., 2016. Integration of Molecular Networking and In-Silico MS/MS Fragmentation for Natural Products Dereplication. Anal. Chem. 88, 3317–3323. https://doi.org/10.1021/acs.analchem.5b04804

Ambrósio Moreira, P., Mariac, C., Scarcelli, N., Couderc, M., Rodrigues, D.P., Clement, C.R., Vigouroux, Y., 2016. Chloroplast sequence of treegourd (*Crescentia cujete*, Bignoniaceae) to study phylogeography and domestication. Appl. Plant Sci. 4, 1600048. https://doi.org/10.3732/apps.1600048

Bianco, A., Bonini, C., Guiso, M., Iavarone, C., Trogolo, C., 1981. ^{13}C NMR spectroscopy of cyclopentenpoliols. Tetrahedron 37, 1773–1777. https://doi.org/10.1016/S0040-

4020(01)98943-4

Bianco, A., Passacantilli, P., 1982. Isolation and partial synthesis of 6-epiaucubin, a new glucosidic iridoid. Tetrahedron 38, 359–362. https://doi.org/10.1016/0040-4020(82)80173-7

Chaudhuri, R.K., Afifi-Yazar, F.U., Sticher, O., Winkler, T., 1980. ^{13}C NMR spectroscopy of naturally occurring iridoids glucosides and their acylated derivatives. Tetrahedron 36, 2317–2326. https://doi.org/10.1016/0040-4020(80)80128-1

Cuyckens, F., Claeys, M., 2004. Mass spectrometry in the structural analysis of flavonoids. J. Mass Spectrom. 39, 1–15. https://doi.org/10.1002/jms.585

Damtoft, S., Rosendal, S., Jensen, Nielsen, B.J., 1981. ^{13}C and ^{1}H NMR spectroscopy as a tool in the configurational analysis of iridoid glucosides. Phytochemistry 20, 2717–2732. https://doi.org/10.1016/0031-9422(81)85275-2

Das, N., Islam, M.E., Jahan, N., Islam, M.S., Khan, A., Islam, M.R., Parvin, M.S., 2014. Antioxidant activities of ethanol extracts and fractions of *Crescentia cujete* leaves and stem bark and the involvement of phenolic compounds. BMC Complement. Altern. Med. 14, 45. https://doi.org/10.1186/1472-6882-14-45

Davidse, P.A., Dillen, J.L.M., Heyns, A.M., Modro, T.A., van Rooyen, P.H., 1990. Photochromic systems. Part 1. Structural and spectroscopic study of photochromically active products of Stobbe condensation. 2,3-Dibenzylidenesuccinic acid and its anhydride. Can. J. Chem. 68, 741–746. https://doi.org/10.1139/v90-117

De Oliveira, G.G., Carnevale Neto, F., Demarque, D.P., De Sousa Pereira-Junior, J.A., Sampaio Peixoto Filho, R.C., De Melo, S.J., Da Silva Almeida, J.R.G., Lopes, J.L.C., Lopes, N.P., 2017. Dereplication of Flavonoid Glycoconjugates from *Adenocalymma imperatoris-maximilianii* by Untargeted Tandem Mass Spectrometry-Based Molecular Networking. Planta Med. 83, 636–646. https://doi.org/10.1055/s-0042-118712

Dinda, B., Debnath, S., Harigaya, Y., 2007. Naturally Occurring Iridoids. A Review, Part 1. Chem. Pharm. Bull. 55, 159–222. https://doi.org/10.1248/cpb.55.159

Dong, J., Ma, X., Fu, Z., Guo, Y., 2011. Effects of microwave drying on the contents of functional constituents of *Eucommia ulmoides* flower tea. Ind. Crops Prod. 34, 1102–1110.

https://doi.org/10.1016/j.indcrop.2011.03.026

Ejelonu, B.C., Lasisi, A.A., Olaremu, A.G., Ejelonu, O.C., 2011. The chemical constituents of calabash (*Crescentia cujete*). African J. Biotechnol. 10, 19631–19636. https://doi.org/10.5897/ajb11.1518

Ersöz, T., Berkman, M.Z., Tasdemir, D., Calis, I., 2002. Iridoid and Phenylethanoid Glycosides from *Euphrasia pectinata*. Turk J Chem 26, 179–188.

Fischer, H., Jensen, H.F.W., Jensen, S.R., Nielsen, B.J., 1987. Eranthemoside, a new iridoid glucoside from Eranthemum pulchellum (Acanthaceae). Phytochemistry 26, 3353–3354. https://doi.org/10.1016/S0031-9422(00)82507-8

Frezza, C., Venditti, A., Matrone, G., Serafini, I., Foddai, S., Bianco, A., Serafini, M., 2018. Iridoid glycosides and polyphenolic compounds from *Teucrium chamaedrys* L. Nat. Prod. Res. 32, 1583–1589. https://doi.org/10.1080/14786419.2017.1392948

Friščić, M., Bucar, F., Kroata, H.P., 2016. LC-PDA-ESI-MSn analysis of phenolic and iridoid compounds from *Globularia spp.* J. Mass Spectrom. 51, 1211–1236. https://doi.org/10.1002/jms.3844

Gentry, A.H., 1980. Bignoniaceae: Part I (Crescentiae and Tourrrettiaeae). New York Botanical Garden Press, New York.

Gupta, M., 1987. Un estudio etnobotánico sobre plantas medicinales de Panamá, in: Escobar, N. (Ed.), Desarrollo de Las Ciencias Naturales y La Medicina En Panamá. Imprenta Universitaria, Panama, pp. 67–97.

Hong, J., Qin, X., Shu, P., Wu, G., Wang, Q., Qin, M., 2007. Analysis of catalpol derivatives by characteristic neutral losses using liquid chromatography combined with electrospray ionization multistage and time-of-flight mass spectrometry. Rapid Commun. Mass Spectrom. 24, 2680–2686. https://doi.org/10.1002/rcm.4676

Kahraman, Ç., Tatli, İ.İ., Kart, D., Ekizoğlu, M., Akdemir, Z.Ş., 2018. Structure elucidation and antimicrobial activities of secondary metabolites from the flowery parts of *Verbascum mucronatum* Lam. Turkish J. Pharm. Sci. 15, 231–237. https://doi.org/10.4274/tjps.50479

Kahraman, C., Tatli, I.I., Orhan, I.E., Akdemir, Z.S., 2010. Cholinesterase inhibitory and antioxidant properties of *Verbascum mucronatum* lam. and its secondary metabolites. Zeitschrift fur Naturforsch. 65, 667–674. https://doi.org/10.1515/znc-2010-11-1206

Kanchanapoom, T., Noiarsa, P., Ruchirawat, S., Kasai, R., 2004. Phenylethanoid and Iridoid Glycosides from the Thai Medicinal Plant, *Barleria strigosa*. Chem Pharm Bull 52, 612–614.

Kaneko, T., Ohtani, K., Kasai, R., Yamasaki, K., Neguyen M, D., 1998. *n*-Alkyl glycosides and p-hydroxybenzoyloxy glucose from fruits of *Crescentia cujete*. Phytochemistry 47, 259–263. https://doi.org/10.1016/S0031-9422(97)00409-3

Kaneko, T., Ohtani, K., Kasai, R., Yamasaki, K., Neguyen M, D., 1997. Iridoids and iridoid glucosides from fruits of *Crescentia cujete*. Phytochemistry 46, 907–910. https://doi.org/10.1016/S0031-9422(97)00375-0

Kang, K. Bin, Park, E.J., Da Silva, R.R., Kim, H.W., Dorrestein, P.C., Sung, S.H., 2018. Targeted Isolation of Neuroprotective Dicoumaroyl Neolignans and Lignans from *Sageretia theezans* Using in Silico Molecular Network Annotation Propagation-Based Dereplication. J. Nat. Prod. 81, 1819–1828. https://doi.org/10.1021/acs.jnatprod.8b00292

Kessner, D., Chambers, M., Burke, R., Agus, D., Mallick, P., 2008. ProteoWizard: Open source software for rapid proteomics tools development. Bioinformatics 24, 2534–2536. https://doi.org/10.1093/bioinformatics/btn323

Kim, D.H., Kim, B.R., Kim, J.Y., Jeong, Y.C., 2000. Mechanism of covalent adduct formation of aucubin to proteins. Toxicol. Lett. 114, 181–188. https://doi.org/10.1016/S0378-4274(99)00295-7

Kumar, V., Sood, H., Sharma, M., Chauhan, R.S., 2013. A proposed biosynthetic pathway of picrosides linked through the detection of biochemical intermediates in the endangered medicinal herb *picrorhiza kurroa*. Phytochem. Anal. 24, 598–602. https://doi.org/10.1002/pca.2437

Kupeli, E., Tatli, I.I., Akdemir, Z.S., Yesilada, E., 2007. Bioassay-guided isolation of anti-inflammatory and antinociceptive glycoterpenoids from the flowers of *Verbascum*

lasianthum Boiss. ex Bentham. J. Ethnopharmacol. 110, 444–450. https://doi.org/10.1016/j.jep.2006.10.004

Lee, K.M., Jeon, J.Y., Lee, B.J., Lee, H., Choi, H.K., 2017. Application of metabolomics to quality control of natural product derived medicines. Biomol. Ther. 25, 559–568. https://doi.org/10.4062/biomolther.2016.249

Letsididi, K.S., Lou, Z., Letsididi, R., Mohammed, K., Maguy, B.L., 2018. Antimicrobial and antibiofilm effects of trans-cinnamic acid nanoemulsion and its potential application on lettuce. Lwt - Food Sci. Technol. https://doi.org/10.1016/j.lwt.2018.04.018

Li, L., Tsao, R., Liu, Z., Liu, S., Yang, R., Young, J.C., Zhu, H., Deng, Z., Xie, M., Fu, Z., 2005. Isolation and purification of acteoside and isoacteoside from *Plantago psyllium* L. by high-speed counter-current chromatography. J. Chromatogr. A 1063, 161–169. https://doi.org/10.1016/j.chroma.2004.11.024

Li, X., Zhang, W., Qian, X., Tong, S., Yan, J., 2014. Separation and Purification of Harpagoside and Sibirioside A from Extract of *Scrophulariae Radix* by Macroporous Resin and High-speed Counter-current Chromatography. Chinese J. Mod. Appl. Pharm. 31, 1121–1124. https://doi.org/10.13748/j.cnki.issn1007-7693.2014.09.023

Martin, F., Hay, A.E., Corno, L., Gupta, M.P., Hostettmann, K., 2007. Iridoid glycosides from the stems of *Pithecoctenium crucigerum* (Bignoniaceae). Phytochemistry 68, 1307–1311. https://doi.org/10.1016/j.phytochem.2007.02.002

Michodjehoun-Mestres, L., Amraoui, W., Brillouet, J.M., 2009. Isolation, characterization, and determination of 1-*O*-trans-Cinnamoyl- β-D-glucopyranose in the epidermis and flesh of developing cashew apple (*Anacardium occidentale* L.) and four of its genotypes. J. Agric. Food Chem. 57, 1377–1382. https://doi.org/10.1021/jf803174c

Nykmukanova, M.M., Eskalieva, B.K., Burasheva, G.S., Iqbal Choudhary, M., Adhikari, A., Amadou, D., 2017. Iridoids from *Verbascum marschallianum*. Chem Nat Compd 53, 491–492. https://doi.org/10.1007/s10600-017-2056-6

Olaniyi, M.B., Lawal, I.O., Olaniyi, A.A., 2018. Proximate, phytochemical screening and mineral analysis of *Crescentia cujete* L . leaves. J. Med. Plants Econ. Dev. 2, a28.

https://doi.org/10.4102/ jomped.v2i1.28

Olivon, F., Grelier, G., Roussi, F., Litaudon, M., Touboul, D., 2017. MZmine 2 Data-Preprocessing to Enhance Molecular Networking Reliability. Anal. Chem. 89, 7836–7840. https://doi.org/10.1021/acs.analchem.7b01563

Pachaly, P., Klein, M., 1987. Constituents of *Andromeda polifolia*. Planta Med. 53, 442–444. https://doi.org/10.1055/s-2006-962769

Pardo, F., Perich, F., Villarroel, L., Torres, R., 1993. Isolation of verbascoside, an antimicrobial constituent of *Buddleja globosa* leaves. J. Ethnopharmacol. 39, 221–222. https://doi.org/10.1016/0378-8741(93)90041-3

Pereira, S.G., de Araújo, S.A., Guilhon, G.M.S.P., Santos, L.S., Junior, L.M.C., 2017. *In vitro* acaricidal activity of *Crescentia cujete* L. fruit pulp against *Rhipicephalus microplus*. Parasitol. Res. 116, 1487–1493. https://doi.org/10.1007/s00436-017-5425-y

Pluskal, T., Castillo, S., Villar-Briones, A., Orešič, M., 2010. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics 11. https://doi.org/10.1186/1471-2105-11-395

Schink, A., Naumoska, K., Kitanovski, Z., Kampf, C.J., Fröhlich-Nowoisky, J., Thines, E., Pöschl, U., Schuppan, D., Lucas, K., 2018. Anti-inflammatory effects of cinnamon extract and identification of active compounds influencing the TLR2 and TLR4 signaling pathways. Food Funct. 9, 5950–5964. https://doi.org/10.1039/c8fo01286e

Schymanski, E.L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H.P., Hollender, J., 2014. Identifying small molecules via high resolution mass spectrometry: Communicating confidence. Environ. Sci. Technol. 48, 2097–2098. https://doi.org/10.1021/es5002105

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res. 13, 2498–2504. https://doi.org/10.1101/gr.1239303

Valladares, M.G., Rios, M.Y., 2007. Iridoids from *Crescentia alata*. J. Nat. Prod. 70, 100–102. https://doi.org/10.1021/np060499w

Vukics, V., Guttman, A., 2010. Structural characterization of flavonoid glycosides by multi-stage mass spectrometry. Mass Spectrom. Rev. 1–16. https://doi.org/10.1002/mas

Wang, G., Yin, W., Zhou, Z.Y., Hsieh, K.L., Liu, J.K., 2010. New iridoids from the fruits of *Crescentia cujete*. J. Asian Nat. Prod. Res. 12, 770–775. https://doi.org/10.1080/10286020.2010.503189

Wang, M., Carver, J.J., Phelan, V.V., Sanchez, L.M., Garg, N., Peng, Y., Nguyen, D.D., Watrous, J., Kapono, C.A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A. V., Meehan, M.J., Liu, W.T., Crüsemann, M., Boudreau, P.D., Esquenazi, E., Sandoval-Calderón, M., Kersten, R.D., Pace, L.A., Quinn, R.A., Duncan, K.R., Hsu, C.C., Floros, D.J., Gavilan, R.G., Kleigrewe, K., Northen, T., Dutton, R.J., Parrot, D., Carlson, E.E., Aigle, B., Michelsen, C.F., Jelsbak, L., Sohlenkamp, C., Pevzner, P., Edlund, A., McLean, J., Piel, J., Murphy, B.T., Gerwick, L., Liaw, C.C., Yang, Y.L., Humpf, H.U., Maansson, M., Keyzers, R.A., Sims, A.C., Johnson, A.R., Sidebottom, A.M., Sedio, B.E., Klitgaard, A., Larson, C.B., Boya, C.A.P., Torres-Mendoza, D., Gonzalez, D.J., Silva, D.B., Marques, L.M., Demarque, D.P., Pociute, E., O'Neill, E.C., Briand, E., Helfrich, E.J.N., Granatosky, E.A., Glukhov, E., Ryffel, F., Houson, H., Mohimani, H., Kharbush, J.J., Zeng, Y., Vorholt, J.A., Kurita, K.L., Charusanti, P., McPhail, K.L., Nielsen, K.F., Vuong, L., Elfeki, M., Traxler, M.F., Engene, N., Koyama, N., Vining, O.B., Baric, R., Silva, R.R., Mascuch, S.J., Tomasi, S., Jenkins, S., Macherla, V., Hoffman, T., Agarwal, V., Williams, P.G., Dai, J., Neupane, R., Gurr, J., Rodríguez, A.M.C., Lamsa, A., Zhang, C., Dorrestein, K., Duggan, B.M., Almaliti, J., Allard, P.M., Phapale, P., Nothias, L.F., Alexandrov, T., Litaudon, M., Wolfender, J.L., Kyle, J.E., Metz, T.O., Peryea, T., Nguyen, D.T., VanLeer, D., Shinn, P., Jadhav, A., Müller, R., Waters, K.M., Shi, W., Liu, X., Zhang, L., Knight, R., Jensen, P.R., Palsson, B., Pogliano, K., Linington, R.G., Gutiérrez, M., Lopes, N.P., Gerwick, W.H., Moore, B.S., Dorrestein, P.C., Bandeira, N., 2016. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat. Biotechnol. 34, 828–837. https://doi.org/10.1038/nbt.3597

Xie, J., Tan, F., Zhu, J., Yue, C., Li, Q., 2012. Separation, purification and quantification of verbascoside from *Penstemon barbatus* (Cav.) Roth. Food Chem. 135, 2536–2541. https://doi.org/10.1016/j.foodchem.2012.07.021

Zeng, X., Su, W., Zheng, Y., Liu, H., Li, P., Zhang, W., Liang, Y., Bai, Y., Peng, W., Yao, H., 2018. UFLC-Q-TOF-MS/MS-based screening and identification of flavonoids and derived metabolites in human urine after oral administration of exocarpium *Citri grandis* extract. Molecules 23, 895. https://doi.org/10.3390/molecules23040895

Zhao, H.Y., Fan, M.X., Wu, X., Wang, H.J., Yang, J., Si, N., Bian, B.L., 2013. Chemical profiling of the chinese herb formula Xiao-Cheng-Qi decoction using liquid chromatography coupled with electrospray ionization mass spectrometry. J. Chromatogr. Sci. 51, 273–285. https://doi.org/10.1093/chromsci/bms138

Zhongjian, J., Zimin, L., 1992. Phenylpropanoid and iridoid glycosides from *Pedicularis longiflora*. Phytochemistry 31, 3125–3127. https://doi.org/10.1016/0031-9422(92)83458-B

Zhongjian, J., Zimin, L., Changzeng, W., 1992. Phenylpropanoid and iridoid glycosides from *Pedicularis lasiophrys*. Phytochemistry 31, 263–266. https://doi.org/10.1016/0031-9422(91)83050-U

Zimin, L., Zhongjian, J., 1991. Phenylpropanoid and iridoid glycosides from *Pedicularis striata*. Phytochemistry 30, 1341–1344. https://doi.org/10.1016/S0031-9422(00)95234-8

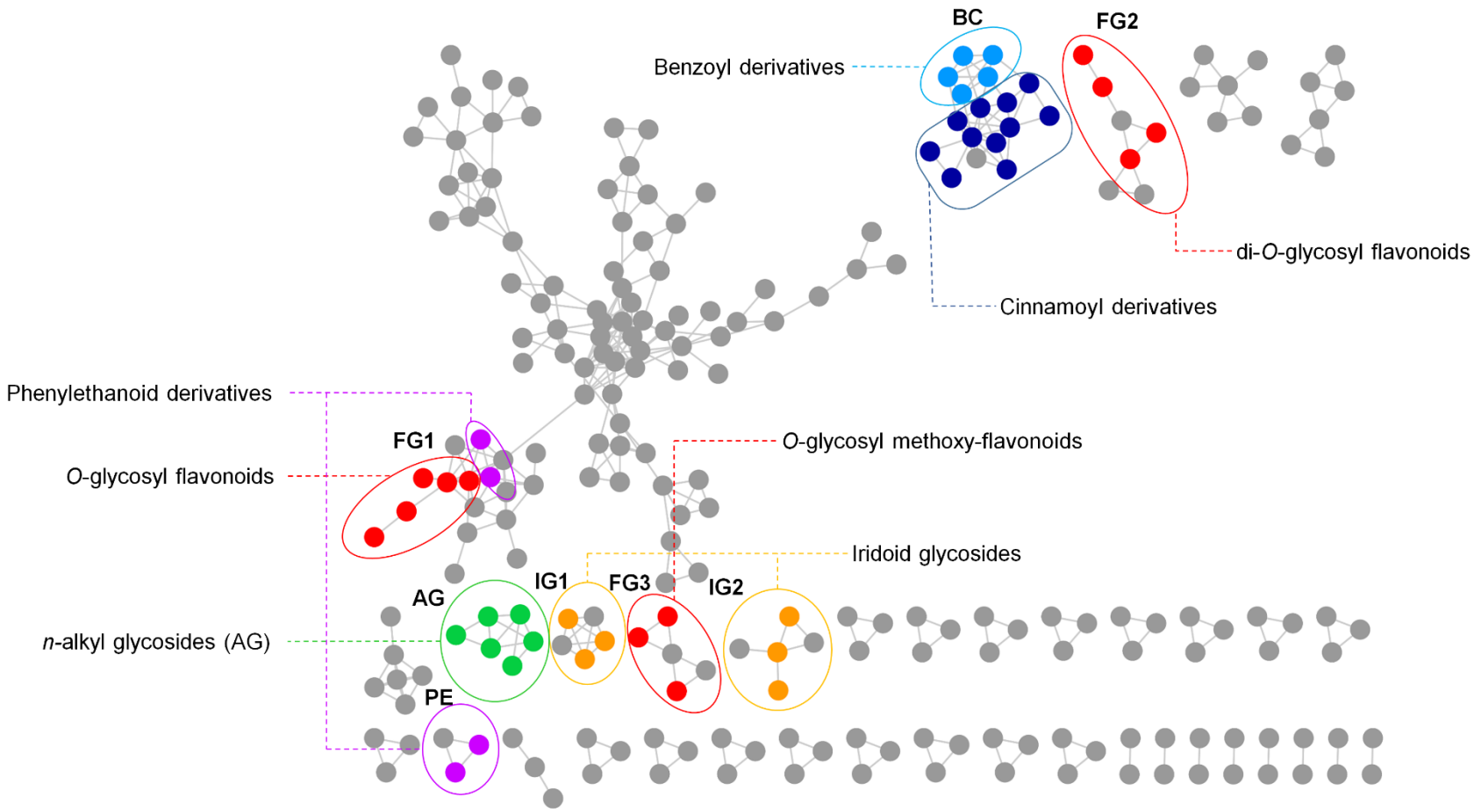

Figure 1. UPLC-MS/MS based molecular networking in negative ionization mode of the fruit extract of *Crescentia cujete*. AG: alkyl glycosides, BC: benzoyl and cinnamoyl derivatives, FG1-3: flavonoid glucosides, PE: phenylethanoid derivatives, IG1-2: iridoids glycosides. Node text indicates the parent ion, node color shows the chemical group (green: *n*-alkyl sugars, sky blue: benzoyl derivatives, dark blue: cinnamoyl derivatives, red: flavonoids glycosides, purple: phenylpropanoids derivatives, and gold: iridoid glycosides).

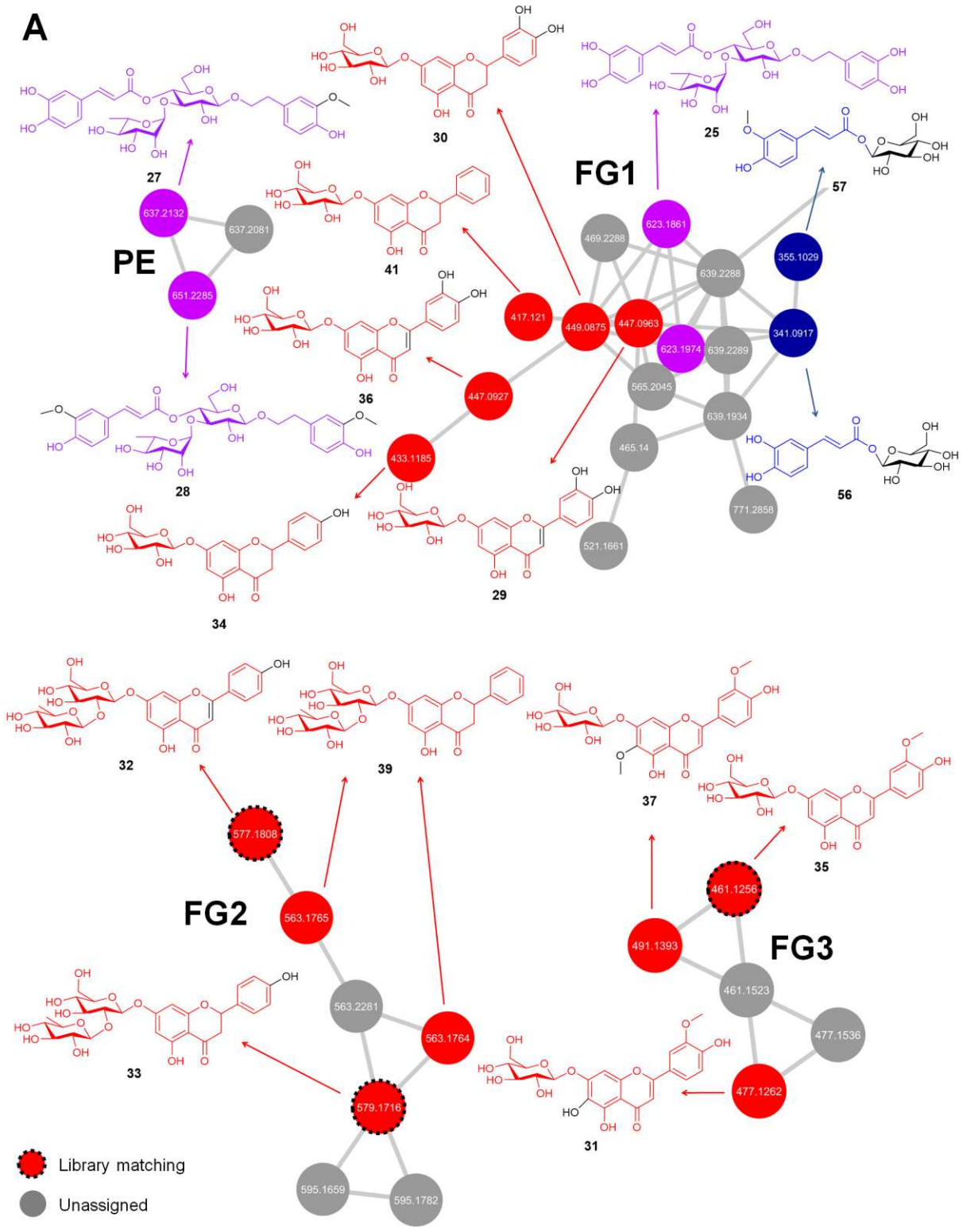

Figure 2. (A) Molecular network of the molecular family of flavonoid glycosides and phenylethanoid extracted from the MN of the fruit extract of *Crescentia cujete*.

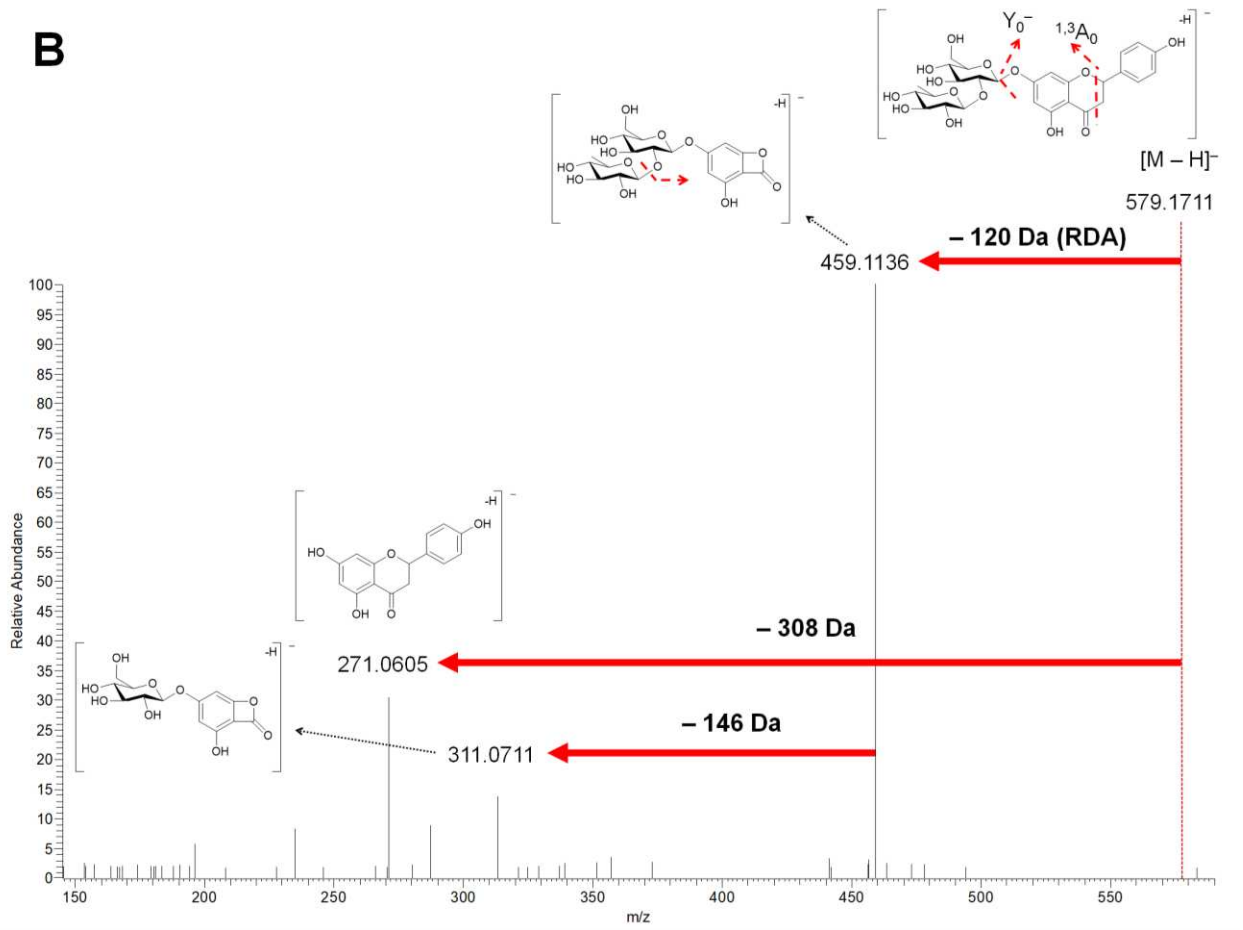

Figure 2. Continuation (B) Proposed fragmentation pathway observed in the MS/MS spectrum naringin (**33**).

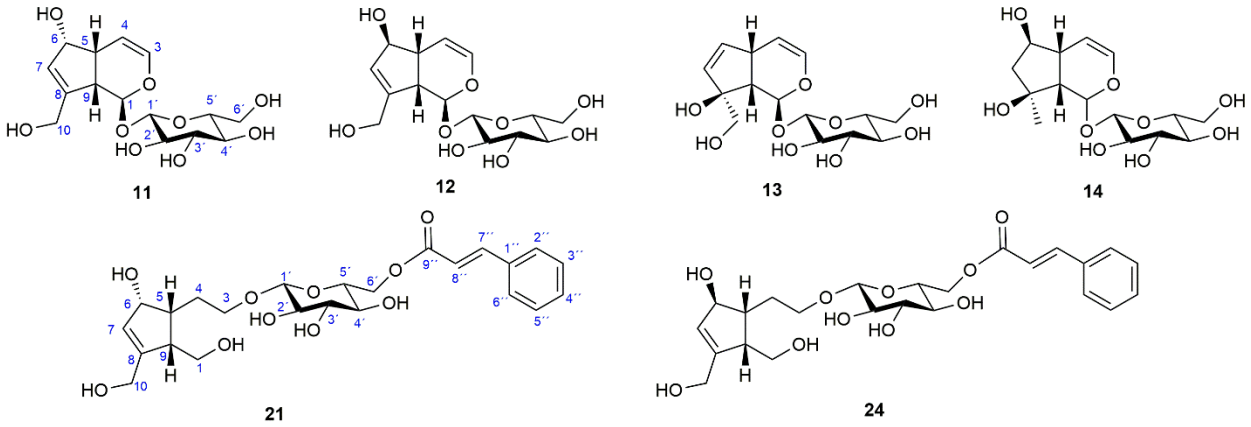

Figure 3. Chemical structures of iridoid glycosides (**11**-**14**, **21** and **24)** isolated from the fruit of *Crescentia cujete*.

Table 1. Detected compounds analyzed by UPLC-ESI-MS/MS from the fruit extract of *Crescentia cujete*. Level of confirmation as proposed by Schymanski et al., 2014: Level 1 (L1): Structure confirmed by reference standard or structure elucidation by NMR spectroscopy; level 2a (L2a): Probable structure by library spectrum match; level 3 (L3): tentative candidates based on MS, MS^2 experimental data.

No.	Compound identification	RT	Formula	Identification confidence	ESI negative full	Delta ppm	ESI negative mode MS/MS
	***n*-Alkyl sugars and sugars**						
1	Sucrose[b]	2.43	$C_{12}H_{22}O_{11}$	L1	341.1090 $[M–H]^-$, 387.1140 $[M–H+FA]^-$	0.1, -0.4	179.06 $[M–H–162(Fru)]^-$, 161.04 $[M–H–162(Fru)–H_2O]^-$
2	(2R,4S)-2-*O*-*β*-D-glucopyranosyl-2,4-pentanediol[c]	6.60	$C_{11}H_{22}O_7$	L1	265.1292 $[M–H]^-$, 311.1345 $[M–H+FA]^-$	-0.4, -0.8	161.04 $[M–H–104(1,2\text{-}PD)]^-$
3	(2R,4S)-2-O-*β*-D-xylopyranosyl-(1→6)-*O*-*β*-D-glucopyranosyl-2,4-pentanediol[c]	7.15	$C_{16}H_{30}O_{11}$	L1	397.1714 $[M–H]^-$, 443.1769 $[M–H+FA]^-$	-0.4, -0.3	265.13 $[M–H–132(Xyl)]^-$, 161.04 $[M–H–104(1,2\text{-}PD)]^-$
4	Pentanyl-*O*-pentoside-*O*-hexoside	10.43	$C_{16}H_{30}O_{10}$	L3	381.1762 $[M–H]^-$, 427.1816 $[M–H+FA]^-$	-1.6, -1.3	249.13 $[M–H–132(Pen)]^-$, 161.04 $[M–H–88(C_5H_{12}O)]^-$
5	Hydroxyoctanyl-*O*-pentoside-*O*-hexoside 1	12.24	$C_{19}H_{36}O_{11}$	L3	439.2180 $[M–H]^-$, 485.2234 $[M–H+FA]^-$	-1.1, -1.1	307.17 $[M–H–132(Pen)]^-$, 161.04 $[M–H–146(C_8H_{18}O_2)]^-$
6	Hydroxyoctanyl-*O*-pentoside-*O*-hexoside 2	12.71	$C_{19}H_{36}O_{11}$	L3	439.2180 $[M–H]^-$, 485.2234 $[M–H+FA]^-$	-1.1, -1.1	307.17 $[M–H–132(Pen)]^-$, 161.04 $[M–H–146(C_8H_{18}O_2)]^-$
7	Octenyl-*O*-pentoside-*O*-hexoside	14.59	$C_{19}H_{34}O_{10}$	L3	421.2079 $[M–H]^-$, 467.2134 $[M–H+FA]^-$	-0.0, -0.1	289.16 $[M–H–132(Pen)]^-$, 161.04 $[M–H–128(C_8H_{16}O)]^-$
8	Octanyl-di-*O*-pentoside-*O*-hexoside	14.87	$C_{24}H_{44}O_{14}$	L3	555.2654 $[M–H]^-$, 601.2708 $[M–H+FA]^-$	-0.7, -0.9	423.22 $[M–H–132(Pen)]^-$, 291.18 $[M–H–132(Pen)]^-$
9	Octanyl-*O*-pentoside-*O*-hexoside	14.95	$C_{19}H_{36}O_{10}$	L3	423.2233 $[M–H]^-$, 469.2287 $[M–H+FA]^-$	-0.7, -0.7	291.18 $[M–H–132(Pen)]^-$, 161.04 $[M–H–130(C_8H_{18}O)]^-$
	Iridoid glycosides						
10	Catalpol (7,8-epoxy-aucubin)	4.33	$C_{15}H_{22}O_{10}$	L3	361.1137 $[M–H]^-$, 407.1190 $[M–H+FA]^-$	-1.0, -1.3	199.06 $[M–H–162(Hex)]^-$
11	6-*epi*-aucubin[c]	5.31	$C_{15}H_{22}O_9$	L1	345.1189 $[M–H]^-$, 391.1244 $[M–H+FA]^-$	-0.5, -0.4	183.07 $[M–H–162(Glu)]^-$, 137.06 $[M–H–162(Glu)–46(CH_2O_2)]^-$
12	Aucubin[c]	5.31	$C_{15}H_{22}O_9$	L1	345.1189 $[M–H]^-$, 391.1244 $[M–H+FA]^-$	-0.5, -0.4	183.07 $[M–H–162(Glu)]^-$, 137.06 $[M–H–162(Glu)–46(CH_2O_2)]^-$
13	8-*epi*-eranthemoside[c]	6.26	$C_{15}H_{22}O_9$	L1	345.1189 $[M–H]^-$, 391.1244 $[M–H+FA]^-$	-0.7, -0.7	183.07 $[M–H–162(Glu)]^-$, 165.05 $[M–H–162(Glu)–H_2O]^-$
14	Ajugol[c]	6.83	$C_{15}H_{24}O_9$	L1	347.1346 $[M–H]^-$, 393.1398 $[M–H+FA]^-$	-0.5, -1.2	167.07 $[M–H–162(Glu)–H_2O]^-$, 149.06 $[M–H–162(Glu)–2H_2O]^-$, 121.06 $[M–H–162(Glu)–2H_2O–CO]^-$
15	Catalposide (hydroxybenzoyl catalpol)[a]	11.76	$C_{22}H_{26}O_{12}$	L2a	481.1349 $[M–H]^-$	-0.5	319.08 $[M–H–162(Glu)]^-$, 205.05 $[M–H–162(Glu)–114(C_5H_6O_3)]^-$, 137.02 $[HBZ–H]^-$
16	Picroside II[a] (Methoxy-hydroxybenzoyl catalpol)	12.13	$C_{23}H_{28}O_{13}$	L2a	511.1457 $[M–H]^-$	-0.1	349.09 $[M–H–162(Glu)]^-$, 235.06 $[M–H–162(Glu)–114(C_5H_6O_3)]^-$, 167.03 $[MHBZ–H]^-$
17	Hydroxybenzoyl-aucubin or aucubin	12.25	$C_{22}H_{26}O_{11}$	L3	465.1401 $[M–H]^-$	-0.3	327.11 $[M–H–138(HBZ)]^-$, 303.08 $[M–H–162(Glu)]^-$, 285.08 $[M–H–180(Glu)]^-$, 165.05 $[M–H–138(HBZ)–162(Glu)]^-$, 137.02 $[HBZ–H]^-$
18	Isoeucomiol benzoyl *O*-hexoside	12.32	$C_{22}H_{30}O_{10}$	L3	453.1765 $[M–H]^-$, 499.1822 $[M–H+FA]^-$	-0.2, 0.1	331.14 $[M–H–122(BZ)]^-$, 295.12 $[M–H–122(BZ)–2H_2O]^-$
19	Hydroxybenzoyl-ajugol	12.42	$C_{22}H_{28}O_{11}$	L3	467.1559 $[M–H]^-$	0.1	305.10 $[M–H–162(Glu)]^-$, 287.09 $[M–H–162(Glu)–H_2O]^-$, 137.02 $[HBZ–H]^-$
20	Methoxy-hydroxybenzoyl aucubin	12.48	$C_{23}H_{28}O_{12}$	L3	495.1509 $[M–H]^-$	0.1	333.10 $[M–H–162(Glu)]^-$, 327.11 $[M–H–168(MHBZ)]^-$, 167.03 $[MHBZ–H]^-$
21	Crescentiol A[c]	13.56	$C_{24}H_{32}O_{10}$	L1	479.1924 $[M–H]^-$, 525.1979 $[M–H+FA]^-$	0.1, 0.2	331.14 $[M–H–148(CM)]^-$, 147.04 $[CM–H]^-$
22	Feruloyl catalpol	13.66	$C_{25}H_{30}O_{13}$	L3	537.1610 $[M–H]^-$	-0.8	375.11 $[M–H–162(Glu)]^-$, 261.08 $[M–H–162(Glu)–114(C_5H_6O_3)]^-$, 193.05 $[FL–H]^-$
23	Feruloyl aucubin	13.73	$C_{25}H_{30}O_{12}$	L3	521.1662 $[M–H]^-$	-0.479	359.11 $[M–H–162(Glu)]^-$, 327.11 $[M–H–194(FL)]^-$, 193.05 $[FL–H]^-$

Table 1. Continued.

No.	Compound identification	RT	Formula	Identification confidence	ESI negative full	Delta ppm	ESI negative mode MS/MS
24	Crescentiol B[c]	14.07	$C_{24}H_{32}O_{10}$	L1	479.1924 $[M–H]^–$, 525.1979 $[M–H+FA]^–$	0.1, 0.2	331.14 $[M–H–148(CM)]^–$, 147.04 $[CM–H]^–$
	Phenylethanoid derivatives						
25	Acteoside (Verbascoside)[c]	11.91	$C_{29}H_{36}O_{15}$	L1	623.1981 $[M–H]^–$	1.7	461.16 $[M–H–162(CF)]^–$
26	Isoacteoside (Isoverbascoside)	12.49	$C_{29}H_{36}O_{15}$	L3	623.1980 $[M–H]^–$	1.5	461.16 $[M–H–162(CF)]^–$
27	Cistanoside C	13.48	$C_{30}H_{38}O_{15}$	L3	637.2134 $[M–H]^–$	-0.6	475.18 $[M–H–162(CF)]^–$
28	Cistenoside D	14.26	$C_{31}H_{40}O_{15}$	L3	651.2288 $[M–H]^–$	-1.0	475.18 $[M–H–176(FL)]^–$
	Flavonoids						
29	Luteolin 7-*O*-glucoside[b]	11.98	$C_{21}H_{20}O_{11}$	L1	447.0930 $[M–H]^–$	1.7	285.04 $[M–H–162(Glu)]^–$
30	Eriodictyol *O*-hexoside	12.07	$C_{21}H_{22}O_{11}$	L3	449.1087 $[M–H]^–$	-0.5	287.05 $[M–H–162(Hex)]^–$
31	Methoxy-tetrahydroxyflavone *O*-hexoside (Methoxy-hydroxy-luteolin *O*-hexoside)	12.20	$C_{22}H_{22}O_{12}$	L3	477.1033 $[M–H]^–$	-1.2	462.08 $[M–H–CH_3]^–$, 315.05 $[M–H–162(Hex)]^–$, 300.03 $[M–H–162(Hex)–CH_3]^–$
32	Rhoifolin[a] (Apigenin *O*-deoxyhexoside *O*-hexoside)	12.85	$C_{27}H_{30}O_{14}$	L2a	577.1556 $[M–H]^–$	-1.4	269.04 $[M–H–162(Hex)–146(Deo)]^–$
33	Naringin[a] (Naringenin *O*-deoxyhexoside *O*-hexoside)	13.05	$C_{27}H_{32}O_{14}$	L2a	579.1711 $[M–H]^–$	-1.4	459.13 $[M–H–120(C_8H_8O)]^–$, 313.07 $[M–H–120(C_8H_8O)–146(Deo)]^–$, 271.06 $[M–H–162(Hex)–146(Deo)]^–$
34	Naringenin *O*-hexoside	13.34	$C_{21}H_{22}O_{10}$	L3	433.1136 $[M–H]^–$	-1.1	271.06 $[M–H–162(Hex)]^–$
35	Methoxy-trihydroxyflavone *O*-hexoside[a] (Methoxy-luteolin *O*-hexoside)	13.41	$C_{22}H_{22}O_{11}$	L3	461.1088 $[M–H]^–$	-0.2	446.08 $[M–H–CH_3]^–$, 299.05 $[M–H–162(Hex)]^–$ 284.03 $[M–H–162(Hex) – CH_3]^–$
36	Luteolin *O*-hexoside	13.59	$C_{21}H_{20}O_{11}$	L3	447.0927 $[M–H]^–$	-1.4	285.04 $[M–H–162(Hex)]^–$
37	Dimethoxy-hydroxy-luteolin *O*-hexoside	13.64	$C_{23}H_{24}O_{12}$	L3	491.1193 $[M–H]^–$	-0.4	476.10 $[M–H–CH3]^–$, 329.06 $[M–H–162(Hex)]^–$, 314.04 $[M–H–162(Hex) – CH3]^–$
38	Luteolin[b]	15.03	$C_{15}H_{10}O_6$	L1	285.0403 $[M–H]^–$	-0.1	151.00 $[M–H–118(C_8H_6O)]^–$
39	Pinocembrin *O*-deoxyhexoside *O*-hexoside	15.03	$C_{27}H_{32}O_{13}$	L3	563.1765 $[M–H]^–$, 609.1819 $[M–H+FA]^–$	-0.9, -1.0	443.13 $[M–H–120(C_8H_8O)]^–$, 297.08 $[M–H–120(C_8H_8O)–146(Deo)]^–$, 255.06 $[M–H–162(Hex)–146(Deo)]^–$
40	Eriodictyol	15.04	$C_{15}H_{12}O_6$	L3	287.0559 $[M–H]^–$	-0.8	151.00 $[M–H–136(C_8H_8O_2)]^–$
41	Pinocembrin *O*-hexoside	15.36	$C_{21}H_{22}O_9$	L3	417.1192 $[M–H]^–$	0.3	255.06 $[M–H–162(Hex)]^–$
42	Pinocembrin[a]	15.83	$C_{15}H_{12}O_4$	L2a	255.0663 $[M–H]^–$	0.1	151.00 $[M–H–104(C_8H_8)]^–$
43	Naringenin[b]	16.02	$C_{15}H_{12}O_5$	L1	271.0613 $[M–H]^–$	0.1	151.00 $[M–H–120(C_8H_8O)]^–$
	Benzoyl and hydroxy benzoyl derivatives						
44	Hydroxybenzoyl *O*-hexoside	7.38	$C_{13}H_{16}O_8$	L3	299.0772 $[M–H]^–$	-0.203	239.05 $[M–H–60(Hex)]^–$, 209.04 $[M–H–90(Hex)]^–$, 179.03 $[M–H–120(Hex)]^–$, 137.02 $[M–H–162(Hex)]^–$
45	Protocatehuic acid (3,4-dihydrobenzoic acid)[b]	7.43	$C_7H_6O_4$	L1	153.0197 $[M–H]^–$	0.398	109.03 $[M–H–44(CO_2)]^–$

Table 1. Continued.

No.	Compound identification	RT	Formula	Identification confidence	ESI negative full	Delta ppm	ESI negative mode MS/MS
46	Hydroxy-methoxy-Benzoyl *O*-hexoside	7.92	$C_{14}H_{18}O_9$	L3	329.0873 $[M–H]^–$	-1.4	269.07 $[M–H–60(Hex)]^–$, 239.05 $[M–H–90(Hex)]^–$, 209.04 $[M–H–120(Hex)]^–$, 167.03 $[M–H–162(Hex)]^–$
47	Sibiricose A3[a] (Hydroxybenzoyl fructoside)	7.96	$C_{19}H_{26}O_{13}$	L2a	461.1298 $[M–H]^–$	-0.6	323.10 $[M–H–138(HBZ)]^–$, 299.08 $[M–H–162(Hex)]^–$, 281.07 $[M–H–180(Hex–H_2O)]^–$, 239.05 $[M–H–162(Hex)–60(Hex)]^–$, 209.04 $[M–H–162(Hex)–90(Hex)]^–$, 137.02 $[M–H–162(Hex)–162(Hex)]^–$
48	Hydroxy-di-methoxybenzoyl-*O*-hexoside	8.22	$C_{15}H_{20}O_{10}$	L3	359.0977 $[M–H]^–$	1.2	299.08 $[M–H–60(Hex)]^–$, 269.07 $[M–H–90(Hex)]^–$, 239.05 $[M–H–120(Hex)]^–$, 197.04 $[M–H–162(Hex)]^–$
49	4-Hydroxy benzoic acid[b]	9.18	$C_7H_6O_3$	L1	137.0248 $[M–H]^–$	2.6	93.03 $[M–H–44(CO_2)]^–$
50	Benzoyl hexopyranosyl hexitol	9.69	$C_{19}H_{28}O_{12}$	L3	447.1505 $[M–H]^–$, 493.1557 $[M–H+FA]^–$	-0.7, -1.1	325.11 $[M–H–122\ (BZ)]^–$, 121.03 $[BZ–H]^–$
51	Benzoyl hexopyranosyl hexitol	9.73	$C_{19}H_{28}O_{12}$	L3	447.1502 $[M–H]^–$	-1.5	325.11 $[M–H–122(BZ)]^–$
52	Benzoyl pentapyranosyl *O*-hexopyranoside 1	10.14	$C_{18}H_{24}O_{11}$	L3	415.1245 $[M–H]^–$, 461.1300 $[M–H+FA]^–$	-0.3, -0.1	293.09 $[M–H–122\ (BZ)]^–$, 121.03 $[BZ–H]^–$
53	Benzoyl pentapyranosyl *O*-hexopyranoside 2	10.25	$C_{18}H_{24}O_{11}$	L3	415.1245 $[M–H]^–$, 461.1300 $[M–H+FA]^–$	-0.3, -0.1	293.09 $[M–H–122\ (BZ)]^–$, 121.03 $[BZ–H]^–$
54	Benzoyl fructofuranosyl *O*-glucopyranoside 1	10.37	$C_{19}H_{26}O_{12}$	L3	445.1350 $[M–H]^–$, 491.1406 $[M–H+FA]^–$	-0.3, -0.1	323.10 $[M–H–122\ (BZ)–H_2O]^–$, 121.03 $[BZ–H]^–$
55	Benzoyl fructofuranosyl *O*-glucopyranoside 2	10.86	$C_{19}H_{26}O_{12}$	L3	445.1348 $[M–H]^–$, 491.1403 $[M–H+FA]^–$	-0.9, -0.7	323.10 $[M–H–122\ (BZ)–H_2O]^–$, 121.03 $[BZ–H]^–$
	Cinnamoyl and hydroxy cinnamoyl derivatives						
56	Dihydroxy-cinnamoyl *O*-hexopyranoside (Caffeoyl *O*-hexoside)	8.11	$C_{15}H_{18}O_9$	L3	341.0876 $[M–H]^–$	-0.5	179.03 $[M–H–162(Hex)]^–$, 161.02 $[M–H–180(Hex)]^–$, 135.04 $[M–H–162(Hex)–44(CO_2)]^–$
57	Feruloyl *O*-hexoside 1	9.83	$C_{16}H_{20}O_9$	L3	355.1032 $[M–H]^–$	-0.7	193.05 $[M–H–162(Hex)]^–$, 175.04 $[M–H–162(Hex)–H_2O]^–$
58	Feruloyl *O*-hexoside 2	10.19	$C_{16}H_{20}O_9$	L3	355.1030 $[M–H]^–$	-1.4	193.05 $[M–H–162(Hex)]^–$, 175.04 $[M–H–162(Hex)–H_2O]^–$
59	Cinnamoyl hexopyranosyl hexitol	12.14	$C_{21}H_{30}O_{12}$	L3	473.1666 $[M–H]^–$	0.3	325.11 $[M–H–148(CM)]^–$, 147.04 $[CM–H)]^–$
60	Cinnamoyl hexosypyranosyl pentitol	12.41	$C_{20}H_{28}O_{11}$	L3	443.1560 $[M–H]^–$	0.2	295.10 $[M–H–148(CM)]^–$, 147.04 $[CM–H]^–$
61	Cinnamoyl pentapyranosyl *O*-hexopyranoside	12.61	$C_{20}H_{26}O_{11}$	L3	441.1401 $[M–H]^–$, 487.1457 $[M–H+FA]^–$	-0.4, -0.1	293.08 $[M–H–148(CM)]^–$, 147.04 $[CM–H]^–$
62	Sibirioside A[c]	12.70	$C_{21}H_{28}O_{12}$	L1	471.1513 $[M–H]^–$, 517.1565 $[M–H+FA]^–$	1.0, 0.4	323.10 $[M–H–148(CM)]^–$, 147.04 $[CM–H]^–$
63	Cinnamoyl *O*-hexanopyranosyl deoxyhexitol	12.99	$C_{21}H_{30}O_{11}$	L3	457.1712 $[M–H]^–$	-0.7	309.12 $[M–H–148(CM)]^–$, 147.04 $[CM–H]^–$
64	Cinnamoyl *O*-fructofuranosyl hexopyranoside / Sibirioside A isomer	13.24	$C_{21}H_{28}O_{12}$	L3	471.1513 $[M–H]^–$, 517.1565 $[M–H+FA]^–$	1.0, 0.4	323.10 $[M–H–148(CM)]^–$, 147.04 $[CM–H]^–$
65	1-*O*-trans-cinnamoyl-*β*-D- glucopyranose[a,c]	13.24	$C_{15}H_{18}O_7$	L1	309.0981 $[M–H]^–$, 355.1036 $[M–H+FA]^–$	0.2, 0.3	147.04 $[CM–H]^–$
66	Trans-cinnamic acid[b,c]	15.97	$C_9H_8O_2$	L1	147.0456 $[M–H]^–$	0.4	103.05 $[M–H–44(CO_2)]^–$

Footnote: Identification result obtained by GNPS annotation[a], comparing the with authentic reference standards[b], metabolite isolated and elucidated by 1D and 2D NMR experiments[c]. Compound with no superscript were manually annotated. FA = formic acid, Fru = fructose, Xyl = xylose, Glu = glucose, 1,2-PD = 1,2-pentadediol, Hex = hexoside, Pen = pentoside, Deo = deoxyhexoside, HBZ = hydroxybenzoyl moiety, MHBZ = methoxy-hydroxybenzoyl moiety, BZ = benzoyl moiety, CM = cinnamoyl moiety, FL = feruloyl moiety, CF = caffeoyl moiety.

Table 2. ^{1}H NMR spectral data for compounds **13, 21, 24** (methanol-d_4, δ in ppm, J in Hz).

Position	13	21	24
Aglycone			
1	5.60 (d, 2.4)	3.59 (dd, 11.0, 2.5), 3.48 (dd, 11.0, 2.8)	3.66 (dd, 11.3, 3.8), 3.53 (dd, 11.0, 5.1)
3	6.13 (dd, 6.2, 2.1)	3.96 (m), 3.72 (ddd, 9.6, 6.4, 6.4)	3.97 (m), 3.73 (ddd, 9.6, 6.4, 6.4)
4	4.87 (dd, 6.4, 2.8)	1.84 (m), 1.77 (m)	1.93 (m), 1.9 (m)
5	3.44 (m)	2.37 (m, 7.8, 7.8, 6.5, 6.5)	2.17 (m, 7.8, 7.8, 7.5, 7.5)
6	5.88 (dd, 5.7, 2.8)	4.33 (m)	4.51 (m)
7	5.72 (dd, 5.7, 1.7)	5.87 (d, 1.5)	5.68 (br. s)
9	2.50 (dd, 8.4, 2.4)	2.63 (d, 7.3)	2.74 (m)
10	3.73 (d, 11.4), Ha	4.14 (d, 14.5), Ha	4.14 (d, 15.6), Ha
	3.61 (d, 11.4), Hb	4.10 (d, 14.5), Hb	4.11 (d, 15.6), Hb
Glucose			
1′	4.65 (d, 7.9)	4.31 (d, 7.8)	4.31 (d, 7.8)
2′	3.25 (m)	3.20 (t, 9.0, 9.0)	3.21 (t, 8.3, 8.3)
3′	3.36 (m)	3.77 (m)	3.37 (m)
4′	3.31 (m)	3.77 (m)	3.37 (m)
5′	3.31 (m)	3.54 (m)	3.54 (m)
6′	3.86 (m), Ha	4.51 (dd, 11.7, 2.1), Ha	4.52 (dd, 11.8, 2.1), Ha
	3.65 (dd, 11.9, 5.3), Hb	4.35 (m), Hb	4.35 (m), Hb
Cinnamic acid			
2″	-	7.61 (dd, 6.0, 3.5)	7.61 (dd, 6.5, 3.0)
3″	-	7.40 (m)	7.40 (m)
4″	-	7.40 (m)	7.40 (m)
5″	-	7.40 (m)	7.40 (m)
6″	-	7.61 (dd, 6.0, 3.5)	7.61 (dd, 6.5, 3.0)
7″	-	7.71 (d, 16.1)	7.71 (d, 16.0)
8″	-	6.56 (d, 16.1)	6.57 (d, 16.0)

Table 3. ^{13}C NMR spectral data for compounds **13, 21, 24** (MeOH-d_4, δ in ppm).

Position		4	5	6
Aglycone				
	1	93.2	58.0	61.8
	3	139.2	69.9	70.4
	4	106.0	26.8	29.7
	5	38.9	44.4	50.2
	6	135.8	75.3	82.1
	7	135.2	129.6	131.9
	8	86.6	151.0	148.3
	9	52.6	49.8	50.2
	10	67.4	60.8	61.3
Glucose				
	1′	99.4	104.5	104.4
	2′	74.8	75.1	75.3
	3′	78.0	77.9	78.1
	4′	71.6	71.7	71.8
	5′	78.2	75.2	75.5
	6′	62.8	64.8	65.0
Cinnamic acid				
	1″	-	135.7	135.9
	2″	-	129.3	129.5
	3″	-	130.1	130.2
	4″	-	131.6	131.7
	5″	-	130.1	130.2
	6″	-	129.3	129.5
	7″	-	146.6	146.7
	8″	-	118.7	118.9
	9″	-	168.5	168.7