Lipidomic and metabolomic profiles of *Coffea canephora* L. beans cultivated in South-western Nigeria

Chinyere F. Anagbogu[1,2,3*], Jiaqi Zhou[1], Festus O. Olasupo[2,3], Mohammed Baba-Nitsa[3] and Diane M. Beckles[1*]

[1] Department of Plant Sciences, University of California, Davis, CA 95616

[2] Department of Crop Protection and Environmental Biology, University of Ibadan, Ibadan, Nigeria

[3] Crop Improvement Division, Cocoa Research Institute of Nigeria, Ibadan, Nigeria

[*] Corresponding authors

E-mail: flora2na@yahoo.com (CFA)

E-mail: dmbeckles@ucdavis.edu (DMB)

Abstract

Coffee (*Coffea* spp.) is one of the most popular refreshments globally. Coffee lipid diversity has untapped potential for improving coffee marketability because lipids contribute significantly to both the health benefits and cup quality of coffee. However, there have not been extensive studies of lipids of *C. canephora* genotypes. In this study, Ultra-performance liquid chromatography coupled with mass spectrometry (UPLC–MS) profiling of lipid molecules was performed for 30 genotypes consisting of 15 cultivated and 15 conserved genotypes of *C. canephora* in Southwestern Nigeria. We identified nine classes of lipids in the 30 genotypes which belong to the 'Niaouli', 'Kouillou' and 'Java Robusta' group: among these, the most abundant lipid class was the triacylglycerols, followed by the fatty acyls group. Although 'Niaouli' diverged from the 'Kouillou' and 'Java Robusta' genotypes when their lipid profiles were compared, there was greater similarity in their lipid composition by multivariate analysis, compared to that observed when their primary metabolites and especially their secondary metabolite profiles were examined. However, distinctions could be made among genotypes. Members of the fatty acyls group had the greatest power to discriminate among genotypes, however, lipids that were low in abundance e.g. a cholesterol ester (20:3), and phosphotidylethanolamine (34:0) were also helpful to understand the relationships among *C. canephora* genotypes. The lipid diversity identified among the *C. canephora* genotypes examined correlated with their overall Single Nucleotide Polymorphism diversity assessed by genotype-by-sequencing, supporting the relevance of this study to coffee cup quality improvement.

Keywords: *Coffea canephora*; bean lipidomics; UPLC-MS; metabolites; SNP

Introduction

"Lipids" are any organic compounds that are insoluble in water, but soluble in organic solvents [1]. Represented in this group are a broad range of molecules such as fatty acids, triacylglycerols, and phospholipids. Lipids serve diverse and critical biological roles in plants such as maintaining cellular membrane integrity and homeostasis, acting as key components in cell signaling pathways and being used as an energy store [2]. In coffee beans, lipids are important reserve compounds and can make up to 10-17% of fresh weight [2,3].

The lipids accumulated in coffee beans are important contributors to beverage organoleptic properties [4-6], and are excellent discriminators of coffee quality [7-11]. Lipids determine beverage foam and emulsion formation and stabilization, which in turn influences flavour and aroma enhancement, especially in espresso brewing [4,12]. Coffee foam and emulsions are responsible for the characteristic pleasant aroma of coffee by trapping and retaining the volatile compounds [13]. Further, the foam created during coffee brewing also influences beverage creaminess and texture [14,15], which further enhances coffee sensory perception.

Coffee bean lipid composition is influenced by the environment but is largely genotype-dependent [11,16]. Differences are apparent in *C. arabica* and *C. canephora,* the two most consumed types of coffee [11]. Arabica coffees generally have lipid contents of 15% which is higher than that in Robusta which averages 10% [2,3]. Arabica also has higher palmitic (16:0), arachidic (20:0), stearic (18:0) and linolenic acids (18:3), but is lower in oleic acid (18:3) compared to *C. canephora* [17,18]. A recent genome wide association study of 107 *C. arabica* accessions found single nucleotide polymorphisms positioned within or near genomic regions coding for proteins involved in lipid and diterpenes metabolic pathways [19], indicating the potential genetic diversity of these key metabolites. In addition to genotype, coffee growth environment e.g. altitude, shade and temperature can also influence lipid composition [8,9,14,20].

Lipidomics in plants is still in its infancy compared to the broader metabolomic profiling, and the data for *C. canephora* is comparatively less than that available for *C . arabica. C. arabica* is known to have a high lipid content compared to *C. canephora* [2,3]. However, there have not been extensive studies of lipids of *C. canephora* genotypes, so the variability relative to *C. arabica* still remains largely undiscovered. This is important because although *C. canephora* is not as popular as *C. arabica*, its production and consumption is increasing, making it important to understand its organoleptic properties [21]. Further, genetic diversity among tetraploid *C. arabica* is not as high as diploid *C. canephora* [22] and this may be reflected in the metabolomic and lipid species diversity within these species.

The aim of this study was to determine the lipid and metabolomics profiles of *C. canephora* genotypes conserved in the Cocoa Research Institute of Nigeria (CRIN) coffee repository, and those used as a source of subsistence by farmers in the southwestern region of Nigeria. Because of the multifaceted role lipids play in determining coffee sensory quality, we wished to identify potential lipid markers differentiating among these *C. canephora* genotypes. These markers would be critical for future coffee improvement programs in Nigeria. We also wished to broaden the current understanding of lipid profiles in diverse coffee species for which there is relatively scarce data. An ultraperformance liquid chromatography method that incorporated a charged surface hybrid (CSH) solid phase was used to separate the different lipid molecules, followed by electrospray ionization (ESI) quadrupole time of flight (QTOF) tandem mass spectrometry (MS/MS) i.e. CSH-ESI QTOF MS/MS. These compounds were subjected to univariate and multivariate analyses to understand their relative abundance. They were also compared to the primary and secondary metabolites profiled in these same coffees, and their relatedness based on lipidomics was compared to that based on Genotyping-by-Sequencing data. We therefore were able to build a comprehensive, multidimensional overview of bean lipids in Southwestern Nigerian coffees.

Materials and methods

Plant material

Coffee beans (*Coffea canephora)* were harvested from coffee germplasm repository of Cocoa Research Institute of Nigeria (CRIN) and coffee farmers' field. Similar genotypes were used for metabolomics [23] consisting of 'Niaouli', 'Kouillou' and 'Java Robusta' varieties. They were classified into six groups based on the result from Single Nucleotide Polymorphism-Genotype by Sequencing analysis [23]. 'Niaouli' is comprised of three genotypes: Nia_1, Nia_2 and Nia_3, (respectively, classified as Groups 1, 2 and 3), 'Kouillou' is comprised of two genotypes: C111 and C36, (respectively, classified as Groups 4, and 5) and, 'Java Robusta' is comprised of one genotype: T1049, and was classified as Group 6. There were five replicates for each group (genotypes) giving a total of 30 samples (Table 1).

Table 1: Coffee genotypes used in this study and their symbols

Variety	**'Niaouli'**			**'Kouillou'**		**'Java Robusta'**
Group/ Genotype	**Group1 (Nia_1)**	**Group 2 (Nia_2)**	**Group 3 (Nia_3)**	**Group 4 (C111)**	**Group 5 (C36)**	**Group 6 (T1049)**
Samples' Symbols	Nia_11	Nia_21	Nia_31	C111_1	C36_1	T1049_1
	Nia_12	Nia_22	Nia_32	C111_2	C36_2	T1049_2
	Nia_13	Nia_23	Nia_33	C111_3	C36_3	T1049_3
	Nia_14	Nia_24	Nia_34	C111_4	C36_4	T1049_4
	Nia_15	Nia_25	Nia_35	C111_5	C36_5	T1049_5

Sample Preparation

Reddish mature (ripened), coffee beans of these genotypes were collected in ice bags and immediately transferred to -80^{o}C. The endosperms of the coffee bean were excised using sterile blade and re-transferred to -80^{o}C. These endosperms were lyophilized, ground into powder with Udy mill (Udy

Corporation) and sealed prior to lipidomic analysis. The lipid was extracted following the protocols according to Matyash *et al*., [24]. Dried extracts were resuspended using a mixture of methanol/toluene (9:1, *v*/*v*) (60 µL) containing an internal standard [12-[(cyclohexylamino)carbonyl] amino]-dodecanoic acid (CUDA)] used as a quality control.

Data Acquisition

Extracted lipids were separated on an Acquity UPLC CSH C18 column (100 × 2.1 mm; 1.7 µm) maintained at 65 °C. The mobile phases for positive mode consisted of 60:40 ACN:H_2O with 10 mM ammonium formate and 0.1% (v/v) formic acid (A) and 90:10 IPA:ACN with 10 mM ammonium formate and 0.1% (v/v) formic acid (B). For negative mode, the mobile phase modifier was 10 mM ammonium acetate instead. The gradient was as follows: 0 min 85% (A); 0–2 min 70% (A); 2–2.5 min 52% (A); 2.5–11 min 18% (A); 11–11.5 min 1% (A); 11.5–12 min 1% (A); 12–12.1 min 85% (A); and 12.1–15 min 85% (A). Sample temperature was maintained at 4 °C in the autosampler. 2 µL of sample was injected. Vanquish UHPLC system (ThermoFisher Scientific) was used. Thermo Q-Exactive HF Orbitrap MS instrument was operated in both positive and negative ESI modes respectively with the following parameters: mass range 120−1700 *m/z*; spray voltage 3.6kV (ESI+) and −3kV (ESI−), sheath gas (nitrogen) flow rate 60 units; auxiliary gas (nitrogen) flow rate 25 units, capillary temperature 320 °C, full scan MS1 mass resolving power 60,000, data-dependent MS/MS (dd-MS/MS) 4 scans per cycle, normalized collision energy at 20%, 30%, and 40%, dd-MS/MS mass resolving power 15,000. Thermo Xcalibur 4.0.27.19 was used for data acquisition and analysis. The instrument was tuned and calibrated according to the manufacturer's recommendations.

Data Processing

Raw data files were converted to the mzML format using the ProteoWizard MSConvert utility. For each m/z values ion chromatogram was extracted with *m/z* thresholds of 0.005 Da and retention time threshold of 0.10 min. Apex of the extracted ion chromatograph was used as peak height value and exported to a .txt file. Peak height files for all the samples were merged together to generate a data matrix. Targeted peak height signal extraction was performed using an R script that is available at https://github.com/barupal. Extracted ion chromatograms for each peak were saved as pictures. CSH-POS and CSH-NEG data matrices were generated. No normalization was applied as minimum signal drift was observed during analysis.

Statistical analysis

Chroma TOF 4.3X software of LECO Corporation and LECO-Fiehn Rtx5 database were used for raw peaks exacting, the data baselines filtering and calibration of the baseline, peak alignment, deconvolution analysis, peak identification and integration of the peak area [25]. The RI (retention time index) method was used in the peak identification, and the RI tolerance was 5000. Metabolite data were normalized by dividing each peak area value by the area of internal standard (Ribitol). Data were $\log_{10}$ transformed, mean-centered and divided by the standard deviation of each variable before performing statistical analysis. The statistical analyses, such as ANOVA, PCA, PLS-DA were performed by using MetaboAnalyst 3.0 [26]. Univariate and multivariate statistical approaches were performed with MetaboAnalystR [26]. One-way Analysis of Variance (ANOVA) test was performed to ascertain the significant variables, and they were expressed as *f* - and *p*-values. The level of statistical significance (Log_{10}(p)) was determined, followed by post-hoc analyses to correct the *p*-value and thus generate the False Discovery Rate (FDR). Fisher's least significant difference method (LSD) was used to identify groups that differ in their lipid profiles [26]. The Hierarchical Clustering Analysis (HCA) plots for SNPs and lipids

were generated under the R environment. Pearson correlative analysis was performed using lipid and metabolism parameters. Two packages in R were used: “psych” [27] and “reshape2” [28] for calculating Pearson’s correlation coefficient (PCC), *p*-value and *q*FDR. The correlation network was drawn using MetScape 3.1.3 [29] in the Cytoscape environment [30].

Results and discussion

The primary aim of this work was to determine the lipid profile and potential lipid markers differentiating *C. canephora* conserved and cultivated in southwestern Nigeria. The main classes of lipids, their relative abundance among the genotypes and their ability to discriminate among the genotypes were determined. Next, how the genotypes associated with each other based on their primary, secondary and lipid profiles. Finally, the relationship among genotypes based on lipidomics was compared to genotyping-by-sequencing SNP data, to determine if there was good accordance between genotypic and functional data i.e. lipids.

Identification of lipids

A total of 1824 lipid species consisted of identified (96) and unidentified (1728) lipids were detected with CSH-ESI QTOF MS/MS. The positive and negative ESI operating modes separated the lipids into 1171 positive and 653 negative lipids. Six lipid subclasses are typically recognized in living systems [31], five were found in the *C. canephora* genotypes studied (Fig 1A).

Each class of lipid contains multiple molecules of varying lengths and degree of saturation. Although these lipid classes are diverse, we constructed a simplistic figure of how they may be related to each other in a metabolic network (Fig 1B). We identified a cholesterol ester (20:3) in this work but these compounds have not been well studied in plants.

189 **Fig 1**. **Lipid classes and their interrelationship. A)** Lipid classes and subclasses detected in *C. canephora* beans.
190 Classification is based on the LIPID MAPS (Metabolites and Pathways Strategy) system, which separates lipids
191 based on their two fundamental "building blocks" - the ketoacyl group and the isoprene group [31]. Abbreviations:
192 FA- Fatty acyls; DG - Diacylglycerol; TG- Triacylglycerols; PE - Phosphatidylethanolamine; PC -Phosphatidylcholine;
193 LPC Lysophosphatidylcholine; SM-Sphingomyelins; CE- Cholesterol esters. **B**) A simplified view of a proposed
194 biochemical interrelationship among the lipid classes according to data in [32,33], as well as in the MetaCyc
195 database [34] derived from plants. The red dashed arrow indicates data obtained from human cells [35].
196

197 # Relative amount of lipids variability among genotypes studied

198 Each lipid group i.e. FA, TG, DG, PC, LPC, is made up of multiple chemicals of varying molecular
199 mass, degree of unsaturation etc. The triacylglycerol (TG) group contained the most chemicals, with 45
200 different species identified (Figure S1A-S1F). We also identified 13 different fatty acyls (FA), nine
201 phosphatidylcholines (PC), four lysophosphatidylcholine (LPC), three diacylglycerol (DG) species and one
202 each of a cholesterol ether, sphingomyelin and phosphotidylethanolamine.

203 The relative abundance of the chemicals within each lipid class was examined across genotypes.
204 The ratio of the compounds with the maximum and minimum concentration was determined and was
205 highest for the TG, FA and DG classes i.e. 2145-, 2923- and 20-fold respectively and lowest for PC, LPC,
206 CE, SM, and PE (Fig 2). Triacylglycerol (52:2) and was the most abundant lipid molecule, while
207 Sphingomyelin (SM) was the lowest (Fig 2). the TG and FA were the most abundant lipids in these
208 genotypes consistent with previous reports [2,3]. Genotype C36_1 ('Kouillou') had the highest amount
209 of TG while Nia_11 ('Niaouli') had the highest amount of FA. The 'Niaouli' (Nia) genotypes are cultivated
210 by farmers, while 'Kouillou' (C111 and C36) and 'Java Robusta' (T1049**)** are conserved in the CRIN
211 germplasm. To better understand the lipid composition in coffee beans, we examined the specific lipids
212 identified within eight of the nine classes (the unknown compounds were not discussed) and compared
213 them across genotypes. Each class was considered in turn, and all data referenced from 3.2.1 to 3.2.3 is
214 contained in S1 Figure.

215 **Fig 2. Relative lipid abundance in the *C. canephora* species studied**. We identified nine subclasses of lipids within
216 the five broader lipid classes described in LIPIDS MAPS. For each subclass, the relative abundance of compounds

among each genotype was assessed. Values shown are the maximum and minimum values within each compound class, and the label indicates the genotype in which the highest value was found.

Fatty-acyl group (FA).

The predominant fatty acyls among our *C. canephora* germplasm were linoleic acid (18:2) followed by palmitic acid (16:0) (S1A Figure), which together encompassed 50% of the total fatty acids. Linoleic and palmitic were also the predominant FAs in a range of *C. arabica* and *C. canephora* genotypes studied by Speer and Kölling-Speer [2]. The number of carbons in the fatty acyls in this study ranged from C15 to C28, also similar to the range i.e. C14 to C24, found by Speer and Kölling-Speer [2].

The proportion of the major FA among genotypes was also examined, because of their influence on coffee quality. These FAs are also members of highly related biosynthetic pathways (Fig 3). FA (16:0) i.e. palmitic acid, was higher in 'Kouillou' and 'Java Robusta' compared to 'Niaouli' (S1A Figure). This is of interest because palmitic acid was positively associated with the high hedonistic values in some specialty Brazilian *C. arabica* coffees [9]. Oleic, linoleic, and linolenic acid levels have also been identified as good markers for Brazilian *C. arabica* coffees which had lower acidity, fragrance, body and flavor [9]. Oleic acid (18:1) was 3-fold lower than linoleic acid (18:2) especially in Nia_11 (S1A Figure). Further, linolenic acid (18:3) was low (10-fold) relative to linoleic acid (18:2). It is tempting to speculate that there is differential regulation of the enzymes that catalyze these steps (Fig 3). In Arabidopsis these two fatty acid desaturases (FAD2 and FAD3) operate independently [36].

Fig 3. Desaturation of C18 fatty acids. This pathway is found in plastids (catalyzed by fatty acid desaturases - FAD6 and FAD7/8) and in the endoplasmic reticulum (catalyzed by FAD2 and FAD3). Arrows indicate the relative amount of each compound based on data derived from this study.

Triacylglycerol (TG).

This group can comprise up to 75% of the lipids in coffee beans [37], and during coffee roasting, they become the carriers of the emerging flavor volatiles [4,12]. There were 45 classes of TGs identified, varying in acyl carbons from C48 to C60 in this study. The farmer cultivated accessions i.e. the 'Niaouli' genotypes had a greater proportion of high-carbon triacylglycerols i.e. C56, C58 lipids compared to 'Kouillou' which had more C48-C52 compounds (S1**B** Figure). The C52 triacylglycerols showed great structural diversity containing up to six double chain triacylglycerols i.e. (52:1) to (52:6).

Cholesterol ester (CE).

Cholesterol esters are cholesterol molecules with a long-chain fatty acid linked to the hydroxyl group [38]. Cholesterol was thought to only be important in animal metabolism with low amounts found in plants, however, this molecule has now been identified as a key precursor for thousands of bioactive plant metabolites [39]. Cholesterol ester (20:3) was very low in abundance in *C. canephora* relative to the other lipid classes (Fig 2), however, levels were notably higher in 'Niaouli' when compared to 'Kouillou' and 'Java Robusta' (S1B Figure).

Lysophosphatidylcholine LPC), Phosphatidylcholine (PC) and Phosphotidylethanolamine (PE).

Phospholipids are potentially important to food sensory perception [37,40]. PCs accumulated to higher levels than LPCs which makes sense given their relative position the lipid biosynthetic pathway (Fig 1B). Generally, the relative levels of these lipids did not vary among genotypes. PEs are usually found in plant membranes [41]. Their levels were low among the genotypes studied, but there was a significant difference found between 'Java Robusta' which did not accumulate as much PE as the other genotypes.

Discriminatory lipid markers

Univariate analysis.

Several studies indicate that lipids can be used as a marker to discriminate among coffee types [13,15,19,42]. One-way analysis of variance (p<0.05; Table 3), was used to identify lipid species that could broadly discriminate among coffees genotypes. Twenty-four lipids separated metabolites in the farmer's genotypes from those conserved at CRIN, since the data for 'Kouillou' and 'Java Robusta' were identical. 'Kouillou' and 'Java Robusta' were treated as a single group i.e. 'Kouillou/Java Robusta' because of the similarities detected. Fourteen of these lipid species were higher in 'Niaouli' and ten were higher in 'Kouillou/Java Robusta' (Table 3).

Table 3. **One-way Analysis of Variance and Fisher's Least Square Difference (LSD) for discriminating genotypes based on lipid content.** The Farmer's cultivated accessions ('Niaouli' i.e. Nia_1, Nia_2, Nia_3) and those from the conserved genotypes ('Kouillou' i.e. C111, C36 and 'Java Robusta' i.e. T1049) were compared. The *f*-value is derived from the *F*. statistic test for significance, the *p*-value tests variability between two groups, $-LOG_{10}(p)$ determines the significant levels, and FDR means False Discovery Rate.

Lipid species more abundant in 'Niaouli'	*f.value*	*p.value*	-LOG10(p)	FDR
TG (56:2)	15.454	7.86E-07	6.1044	3.20E-05
TG (58:4)A	15.260	8.79E-07	6.0562	3.20E-05
PE (34:2)	15.037	1.00E-06	6.0000	3.20E-05
CSH_negESI (19:2)	11.988	6.84E-06	5.1650	0.000131
TG (58:3)	11.987	6.84E-06	5.1648	0.000131
CE (20:3)	11.212	1.18E-05	4.9299	0.000188
CSH_posESI (11:6)	9.9478	3.00E-05	4.5229	0.000411
FA (28:0)	9.0851	5.94E-05	4.2263	0.000713
TG (58:2)	6.4147	0.000645	3.1901	0.006197
TG (56:3)	5.2852	0.002060	2.6862	0.014124
TG (56:1)	5.1318	0.002431	2.6142	0.014586
FA (22:0)	5.0736	0.002590	2.5867	0.014627
FA (20:1)	4.4342	0.005305	2.2754	0.025462
TG (56:4)	3.9802	0.009030	2.0443	0.037692
Lipid species more abundant in 'Kouillou' /'Java Robusta'				
FA (16:0)	7.3943	0.000255	3.5933	0.002721
FA (20:0)	6.2489	0.000760	3.1189	0.006637
TG (50:1)_1	5.9483	0.001029	2.9874	0.008235

CSH_posESI (17:6)	5.7749	0.00123	2.9102	0.009082
CSH_posESI (29:8)	5.1657	0.002343	2.6302	0.014586
CSH_posESI (04:1)	4.6549	0.004124	2.3847	0.021996
CSH_posESI (28:2)	4.5120	0.004852	2.3141	0.024513
TG (52:2)	4.2978	0.006211	2.2069	0.028392
TG (48:1)	4.1161	0.007685	2.1144	0.033535
CSH_posESI (14:1)	3.8540	0.010507	1.9785	0.042028

The lipid with the highest discriminatory power between the two groups of genotypes were TG (56:2), TG (58:4)A, and PE (34:2) which accumulated to high levels in 'Niaouli' relative to other genotypes. The lipid species which differentiates 'Kouillou/Java Robusta' from 'Niaouli' were two saturated fatty acids, palmitic (16:0) arachidic acids (20:0) and TG (50:1). Based on our data analysis we show in Table 4, the lipids that have a high discriminatory power in the coffees in our study, were compared with information from data that was published by others. Fatty acids were the most discriminatory of all lipids, although the types varied with respect to the species examined (Table 4).

Table 4. **Fatty acids with high discriminatory power among coffee genotypes.** Included are the two main varieties of *C. canephora* ('Niaouli' and 'Kouillou') used in this study, two cultivated *Coffea* species (*C. arabica* and *C. canephora*) and specialty or high-quality coffee. Citations to the referenced data are included.

'Niaouli'	**'Kouillou'**	***C. arabica***	***C. canephora***	**Specialty coffees**
▪ FA (22:0) Behenic acid. ▪ FA (20:1) Arachidic acid. ▪ FA (28:0) Montanic acid	▪FA (16:0) Palmitic acid ▪ FA (20:0) Arachidic acid	▪ FA (18:2) Linoleic acid [17] ▪ Polyunsaturated Fatty acids [16]	▪ FA (18:1) Oleic acid [17] ▪ Monounsaturated fatty acids [17]	▪ FA (14:0) Myristic acid [6][1] ▪ FA (20:0) Arachidic acid. ▪ FA (18:0) Stearic acid ▪ FA (16:0) Palmitic acid [9] [2]

1 *C. canephora*; 2: *C. arabica*

Multivariate analysis – Partial Least Square Projections

Variable influence (or importance) of projection (VIP) scores were used as another tool to identify the lipids that contribute most to differences among genotypes (Fig 4). This analysis predicted

299 that PE 34:2, a negative unknown lipid, CSH_negESI (19:2), and FA (28:0) i.e. dodecanoic acid could
300 distinguish 'Niaouli' from the conserved genotypes. The ANOVA data in Table 3, also indicated that these
301 three compounds accumulated to higher levels in the 'Niaouli' genotype.
302
303 **Fig 4. Potential lipidomic markers differentiating 'Niaouli' (genotypes groups 1–3) from 'Kouillou' (groups 4–5)**
304 **and 'Java' (group 6).** The variable importance in projection (VIP) scores on the x-axis provide an estimate of the
305 contribution of a given predictor (lipid species shown on the y-axis) to the Partial Least Square (PLS) regression
306 above. The higher the VIP score, the better the lipid species is as a predictor of the discrimination among the
307 genotypes.
308

309 Understanding the lipid networks in *C. canephora* coffee beans

310 We examined the lipid-to-lipid correlations among *C. canephora* beans. Many compounds in
311 related pathways tend to occur within defined stoichiometries and kinetics parameters [43,44].
312 Neighbouring compounds may be correlated with each other. This is also true of metabolites that are in
313 chemical equilibrium, show mass conservation, or are highly sensitive to a common parameter [43,45].
314 Using Pearson's correlation coefficient at a cut-off of ≥ | ± 0.75|, we identified those lipids that co-
315 occurred at similar relative levels across the *C. canephora* genotypes studied. Highly specific interactions
316 between lipids are essential for regulation of cell physiology [46,47] (Fig 5 and detailed in S2 Table).
317 The most notable observation was the very strong negative correlations between CE (20:3) and
318 other lipids, many of which were uncharacterized (Fig 5). Negative correlations were also found with CE
319 (20.3) and TG (54.1), TG (52.1), but most notably with TG (50.1). Based on the number of carbons in
320 their acyl chains, it seems likely that these three TGs are closely connected in the TG biosynthetic
321 pathway. Negative correlations among metabolites are sometimes due to regulatory mechanisms
322 controlling metabolites that have conserved moieties, or metabolites connected by enzymes with high
323 variance which can cause negative correlations between its substrate and product metabolites [43,48].
324 Based on the simplified pathway in Figure 1B, a relationship between the CE and TGs can be envisaged.

Not all of the relationships between CE (20:3) and other lipids were negative. There was a positive correlation between CE (20:3) and TG (58.2), TG (58.4), and FA (22.0) respectively. Most triacylglycerols were positively associated with other lipids except, TG (52:0) and TG (50:1), which showed a negative correlation (Fig 5).

Fig 5. Pearson's correlation network illustrating correlation patterns among lipids. Correlations found among lipid data are shown as attribute circular layout. The blue lines indicate lipids whose occurrence negatively correlate, while the pink lines denote lipids that show a positive correlation. Line thickness indicates the strength of correlation. A permissive threshold level of ≥ | ± 0.75| was set for the Pearson's correlation coefficient and a *q*FDR of <0.05 was chosen for the identification of significant correlations within the lipid data.

Comparing the primary, secondary metabolites and lipid profiles among *C. canephora*

Metabolomics and lipidomics are powerful strategies to qualitatively and quantitatively analyze a wide range of small molecules in a biological sample, which represent endpoints of genome expression [49]. The metabolomics data from our recent work [21] was subjected to PLS-DA analysis for both primary and secondary metabolites and compared to the lipidomics profile of these coffees. Our aim was to determine the diversity of metabolism across genotypes. We expected greater similarity in primary metabolism across genotypes, and that specialized metabolites would be more divergent [50]. Lipids are primary metabolites but in coffee beans, it is not known the extent to which they are part of the fundamental or specialized metabolic machinery.

We found that there was clearer differentiation between 'Niaouli' genotypes which clustered together and away from the 'Kouillou' and 'Java Robusta' when secondary metabolites were examined (Fig 6B) compared to primary metabolites (Fig 6A). No clear difference was observed among the groups with PLS-DA on lipid (Fig 6C) but SPLS-DA - (Fig 6D) was able to reveal more similarity between 'Java' and the 'Niaouli' types, while the 'Kouillou' genotypes were separate from the rest. This is noteworthy

because of the many studies that show 1) that lipids are valuable components of the sensory attributes of coffee and 2) that they are excellent at discriminating genotypes and growing environment [8,9,14].

Fig 6. **Differentiating among the *C. canephora* genotypes using multivariate analyses. A)** Primary metabolites analyzed by PLS-DA, **B)** Secondary metabolites analyzed by PLS-DA, **C)** lipids analyzed by PLS and **D)** lipids analyzed by Sparse-PLS. (PCA = Principal Component Analysis, PLS-DA = Partial Least Square –discriminate analysis. Genotypes 1-3 belong to the 'Niaouli' group i.e. Nia_1, Nia_2 and Nia_3, Genotypes 4-5 are members of the 'Kouillou' group i.e. C111 and C36, and Genotype 6 is 'Java Robusta' i.e. T1049. The primary and secondary metabolite data were derived from Anagbogu *et al*., 2019b [21].

Comparing the lipid variability among accessions with their Single Nucleotide Polymorphism diversity

In spite of the importance of lipids in coffee cup quality improvement, as far as we know, it has not been fully integrated into coffee breeding strategies. The DNA of coffees extracted in this study was analyzed using Genotype-by-Sequencing and their relatedness compared using hierarchical clustering. The lipidomics data was also analyzed similarly. Overall, there was good agreement between the DNA- and metabolite-based clustering (Fig 7A). From the genetic perpective, we observed three clusters I, II and III representing 'Niaouli', 'Kouillou', and 'Java' respectively (Fig 7A). The lipidomic analysis of the genotypes based on the mean values (Fig 7B) revealed two main clusters, separating the 'Niaouli' from 'Kouillou'/'Java Robusta' respectively (Fig 7B). However there were notable exceptions: SNP analysis indicated that C36 and C111 ('Kouillou') were highly related, while T1049 ('Java Robusta')was divergent, in contrast, lipidomics indicted that T1049 and C111 formed a distinguishable subgroup (cluster IIa) while C36 was more distant. (Fig 7C).

All lipid analyses (Figs 7B and C) revealed a clear disparity between the 'Niaouli' and 'Kouilou/ Java Robusta' genotypes. We suggest that there are two lipid diversity structures within the population

studied which can be used for future coffee improvement in Nigeria. From this study, FA(16:0) and FA(20:0) were higher (p<0.05) in the 'Kouillou' genotypes (Table 3), and these fatty acids are among the precursors of good quality coffee (Table 3). Incorporating these important traits into the genome of cultivated variety ('Niaouli') will be of great help towards its improvement. Also, further study on the SNP data to detect sources or causes of this great difference between 'Kouilou' and 'Niaouli' is needed.

Fig 7. Comparative analysis of the genomics and lipidomic analysis based on their hierarchical structure. **A)** genetic analysis based on ~100,000 SNP, **B)** lipidomic analysis indicating the mean value of each genotype from five biological replicates and **C)** lipidomic analysis including all replicates. The genotypes and replicates names were described in Table 1 as follows: Nia_11- Nia_15 were Genotype 1, Nia_21-Nia_25 were genotype 2, Nia_31-Nia35 were Genotype 3 and belong to the 'Niaouli' group; Members of the 'Kouilou' group have been referred to as Genotypes 4 and 5 in this study and included C111_1-C111_5 and C36_1 to C36_5. 'Java Robusta' coffee has been described as Genotype 6 included T1049_1-T1049_5.

Conclusion

There is much interest in using lipid profiles to identify biomarkers for coffee cup quality improvement. In this study, we used various data mining tools to better understand variations in lipid profiles of South western Nigerian coffee genotypes, and how these lipids relate to other metabolites, and to other studies done on coffee. Our findings can be summarized as following: 1) Lipids that could discriminate among the genotypes studied were identified, even though there was less variability for lipids among genotypes compared with other types of metabolites. 2) 'Java Robusta' and 'Kouillou' especially 'Kouillou' genotype C36, contained more saturated fatty acyls species such as FA(26:0), FA(16:0), FA(20:0), while 'Niaouli' contained more unsaturated FAs such as FA(16:1), FA(20:1). It should be interesting to study the SNPs of gene such as FADs which are responsible for desaturation in coffee bean. 3) Unexpected observations were made with respect to Cholesterol ester (20:3) which was one of

the least abundant lipids in our analysis. The levels of CE(20:3) and some TG compounds, showed some of the strongest negative correlations. Further, CE(20:3) can discriminate between 'Niaouli' and ''Kouillou' and Java Robusta' indicating its value as a genotypic marker. 4) Lipids as a class of compounds in this analysis did not discriminate among the genotype studied, as well as secondary metabolites but grouped genotypes into two lipid diversity structures. 5) There was a good correlation between classification of genotypes based on SNP variability and lipidomic profile.

Acknowledgements

Support for this research was provided in part by the Norman E. Borlaug Leadership Enhancement in Agriculture Program (Borlaug LEAP) through a grant to the University of California-Davis by the United States Agency for International Development. The opinions expressed herein are those of the authors and do not necessarily reflect the views of USAID. Funding was also provided by the Federal Ministry of Agriculture in Nigeria through a grant to the Cocoa Research Institute of Nigeria.

References

1. Atwood T, Campbell P, Parish H, Smith A, Vella F, Stirling J, editors. Oxford Dictionary of Biochemistry and Molecular Biology. Oxford, England: Oxford University Press; 2006.

2. Speer K, Kölling-Speer I. The lipid fraction of the coffee bean. Brazilian Journal of Plant Physiology. 2006;18:201-16.

3. Folstar P. Lipids. In: Clarke RJ, Macrae R, editors. Coffee. Vol I: Chemistry. Essex, England: Elsevier Applied Science; 1989. p. 203-22.

4. Navarini L, Ferrari M, Liverani FS, Liggieri L, Ravera F. Dynamic tensiometric characterization of espresso coffee beverage. Food Hydrocolloid. 2004;18(3):387-93. PubMed PMID: WOS:000220352600005.

5. Piazza L, Gigli J, Bulbarello A. Interfacial rheology study of espresso coffee foam structure and properties. J Food Eng. 2008;84(3):420-9. PubMed PMID: WOS:000250118300008.

6. Gamboa-Becerra R, Hernandez-Hernandez MC, Gonzalez-Rios O, Suarez-Quiroz ML, Galvez-Ponce E, Ordaz-Ortiz JJ, et al. Metabolomic Markers for the Early Selection of Coffea canephora Plants with Desirable Cup Quality Traits. Metabolites. 2019;9(10). PubMed PMID: WOS:000494273300022.

7. Amorim ACL, Hovell AMC, Pinto AC, Eberlin MN, Arruda NP, Pereira EJ, et al. Green and Roasted Arabica Coffees Differentiated by Ripeness, Process and Cup Quality via Electrospray Ionization Mass Spectrometry Fingerprinting. J Brazil Chem Soc. 2009;20(2):313-21. PubMed PMID: WOS:000264017600017.

8. Bertrand B, Villarreal D, Laffargue A, Posada H, Lashermes P, Dussert S. Comparison of the Effectiveness of Fatty Acids, Chlorogenic Acids, and Elements for the Chemometric Discrimination of Coffee (Coffea arabica L.) Varieties and Growing Origins. Journal of Agricultural and Food Chemistry. 2008;56(6):2273-80. doi: 10.1021/jf073314f.

9. Figueiredo LP, Borém FM, Ribeiro FC, Giomo GS, Henrique da Silva Taveira J, Malta MR. Fatty acid profiles and parameters of quality of specialty coffees produced in different Brazilian regions. African Journal of Agricultural Research. 2015;10(35):3484-93. doi: https://doi.org/10.5897/AJAR2015.9697.

10. Franca AS, Mendonca JCF, Oliveira SD. Composition of green and roasted coffees of different cup qualities. Lwt-Food Sci Technol. 2005;38(7):709-15. PubMed PMID: WOS:000230507000002.

11. Martin MJ, Pablos F, Gonzalez AG, Valdenebro MS, Leon-Camacho M. Fatty acid profiles as discriminant parameters for coffee varieties differentiation. Talanta. 2001;54(2):291-7. PubMed PMID: WOS:000168336400009.

12. Nebesny E, Budryn G. Evaluation of sensory attributes of coffee brews from robusta coffee roasted under different conditions. Eur Food Res Technol. 2006;224(2):159-65. PubMed PMID: WOS:000242295300003.

13. Farah A, Monteiro MC, Calado V, Franca AS, Trugo LC. Correlation between cup quality and chemical attributes of Brazilian coffee. Food Chem. 2006;98(2):373-80. PubMed PMID: WOS:000236527000024.

14. Mehari B, Redi-Abshiro M, Chandravanshi BS, Combrinck S, McCrindle R, Atlabachew M. GC-MS profiling of fatty acids in green coffee (Coffea arabica L.) beans and chemometric modeling for

tracing geographical origins from Ethiopia. Journal of the Science of Food and Agriculture. 2019;99(8):3811-23. doi: 10.1002/jsfa.9603.

15. Novaes FJM, Oigman SS, de Souza ROMA, Rezende CM, Neto FRD. New approaches on the analyses of thermolabile coffee diterpenes by gas chromatography and its relationship with cup quality. Talanta. 2015;139:159-66. PubMed PMID: WOS:000353857500023.

16. Agatonovic-Kustrin S, Morton DW, Smirnov V, Petukhov A, Gegechkori V, Kuzina V, et al. Analytical Strategies in Lipidomics for Discovery of Functional Biomarkers from Human Saliva. Dis Markers. 2019;2019. PubMed PMID: WOS:000503382800001.

17. Cossignani L, Montesano D, Simonetti MS, Blasi F. Authentication of Coffea arabica according to Triacylglycerol Stereospecific Composition. J Anal Methods Chem. 2016. PubMed PMID: WOS:000381189000001.

18. Romano R, Santini A, Le Grottaglie L, Manzo N, Visconti A, Ritieni A. Identification markers based on fatty acid composition to differentiate between roasted Arabica and Canephora (Robusta) coffee varieties in mixtures. J Food Compos Anal. 2014;35(1):1-9. PubMed PMID: WOS:000340978700001.

19. Sant'Ana GC, Pereira LFP, Pot D, Ivamoto ST, Domingues DS, Ferreira RV, et al. Genome-wide association study reveals candidate genes influencing lipids and diterpenes contents in Coffea arabica L. Sci Rep-Uk. 2018;8. PubMed PMID: WOS:000419941000006.

20. Joet T, Laffargue A, Descroix F, Doulbeau S, Bertrand B, de Kochko A, et al. Influence of environmental factors, wet processing and their interactions on the biochemical composition of green Arabica coffee beans. Food Chem. 2010;118(3):693-701. PubMed PMID: WOS:000270492800032.

21. Anagbogu CF, Ilori CO, Bhattacharjee R, Olaniyi OO, Beckles DM. Gas Chromatography-Mass Spectrometry and Single Nucleotide Polymorphism-Genotype-By-Sequencing Analyses Reveal the Bean Chemical Profiles and Relatedness of Coffea canephora Genotypes in Nigeria. Plants-Basel. 2019;8(10). PubMed PMID: WOS:000498270000071.

22. Scalabrin S, Toniutti L, Di Gaspero G, Scaglione D, Magris G, Vidotto M, et al. A single polyploidization event at the origin of the tetraploid genome of Coffea arabica is responsible for the extremely low genetic variation in wild and cultivated germplasm. Sci Rep-Uk. 2020;10(1):4642. doi: 10.1038/s41598-020-61216-7.

23. Anagbogu CF, Bhattacharjee R, Ilori C, Tongyoo P, Dada KE, Muyiwa AA, et al. Genetic diversity and re-classification of coffee (Coffea canephora Pierre ex A. Froehner) from South Western Nigeria through genotyping-by-sequencing-single nucleotide polymorphism analysis. Genet Resour Crop Ev. 2019;66(3):685-96. PubMed PMID: WOS:000461347800009.

24. Matyash V, Liebisch G, Kurzchalia TV, Shevchenko A, Schwudke D. Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. J Lipid Res. 2008;49(5):1137-46. PubMed PMID: WOS:000255753700026.

25. Kind T, Wohlgemuth G, Lee DY, Lu Y, Palazoglu M, Shahbaz S, et al. FiehnLib: Mass Spectral and Retention Index Libraries for Metabolomics Based on Quadrupole and Time-of-Flight Gas Chromatography/Mass Spectrometry. Anal Chem. 2009;81(24):10038-48. PubMed PMID: WOS:000272562200025.

26. Xia JG, Sinelnikov IV, Han B, Wishart DS. MetaboAnalyst 3.0-making metabolomics more meaningful. Nucleic Acids Res. 2015;43(W1):W251-W7. PubMed PMID: WOS:000359772700039.

27. Revelle WR. psych : Procedures for Personality and Psychological Research. 2017.

28. Wickham H. Reshaping data with the reshape package. J Stat Softw. 2007;21(12):1-20. PubMed PMID: WOS:000252429400001.

29. Basu S, Duren W, Evans CR, Burant CF, Michailidis G, Karnovsky A. Sparse network modeling and metscape-based visualization methods for the analysis of large-scale metabolomics data. Bioinformatics. 2017;33(10):1545-53. PubMed PMID: WOS:000402130700015.

30. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498-504. PubMed PMID: WOS:000186357000016.

31. Fahy E, Cotter D, Sud M, Subramaniam S. Lipid classification, structures and tools. Bba-Mol Cell Biol L. 2011;1811(11):637-47. PubMed PMID: WOS:000297399000002.

32. Du Z-Y, Benning C. Triacylglycerol Accumulation in Photosynthetic Cells in Plants and Algae. In: Nakamura Y, Li-Beisson Y, editors. Lipids in Plant and Algae Development. Cham: Springer International Publishing; 2016. p. 179-205.

33. Zhou X-R, Callahan DL, Shrestha P, Liu Q, Petrie JR, Singh SP. Lipidomic analysis of Arabidopsis seed genetically engineered to contain DHA. Front Plant Sci. 2014;5(419). doi: 10.3389/fpls.2014.00419.

34. Caspi R, Billington R, Fulcher CA, Keseler IM, Kothari A, Krummenacker M, et al. The MetaCyc database of metabolic pathways and enzymes. Nucleic Acids Res. 2017;46(D1):D633-D9. doi: 10.1093/nar/gkx935.

35. Yang H, Suh DH, Kim DH, Jung ES, Liu K-H, Lee CH, et al. Metabolomic and lipidomic analysis of the effect of pioglitazone on hepatic steatosis in a rat model of obese Type 2 diabetes. British Journal of Pharmacology. 2018;175(17):3610-25. doi: 10.1111/bph.14434.

525 36. Chen MJ, Thelen JJ. ACYL-LIPID DESATURASE2 Is Required for Chilling and Freezing Tolerance in
526 Arabidopsis. Plant Cell. 2013;25(4):1430-44. PubMed PMID: WOS:000319450500020.

527 37. Zhou L, Khalil A, Bindler F, Zhao MJ, Marcic C, Ennahar S, et al. Effect of heat treatment on the
528 content of individual phospholipids in coffee beans. Food Chem. 2013;141(4):3846-50. PubMed
529 PMID: WOS:000324848000074.

530 38. Korber M, Klein I, Daum G. Steryl ester synthesis, storage and hydrolysis: A contribution to sterol
531 homeostasis. Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids.
532 2017;1862(12):1534-45. doi: https://doi.org/10.1016/j.bbalip.2017.09.002.

533 39. Sonawane PD, Pollier J, Panda S, Szymanski J, Massalha H, Yona M, et al. Plant cholesterol
534 biosynthetic pathway overlaps with phytosterol metabolism. Nat Plants.
535 2017;3:https://doi.org/10.1038/nplants.2016.205. PubMed PMID: WOS:000395899200007.

536 40. de Lima MSA, Rocha LA, Molina EF, Caetano BL, Marcal L, Mello C, et al. Thermoanalysis of
537 soybean oil extracted by two methods. Quim Nova. 2008;31(3):527-9. PubMed PMID:
538 WOS:000256956100012.

539 41. Liu Y, Wang GL, Wang XM. Role of Aminoalcoholphosphotransferases 1 and 2 in Phospholipid
540 Homeostasis in Arabidopsis. Plant Cell. 2015;27(5):1512-28. PubMed PMID:
541 WOS:000357679900010.

542 42. Scholz MBD, Kitzberger CSG, Pagiatto NF, Pereira LFP, Davrieux F, Pot D, et al. Chemical
543 composition in wild ethiopian Arabica coffee accessions. Euphytica. 2016;209(2):429-38.
544 PubMed PMID: WOS:000375064500011.

545 43. Camacho D, de la Fuente A, Mendes P. The origin of correlations in metabolomics data.
546 Metabolomics. 2005;1(1):53-63. doi: 10.1007/s11306-005-1107-3.

547 44. Stamova BS, Roessner U, Suren S, Laudencia-Chingcuanco D, Bacic A, Beckles DM. Metabolic
548 profiling of transgenic wheat over-expressing the high-molecular-weight Dx5 glutenin subunit.
549 Metabolomics. 2009;5(2):239-52. doi: 10.1007/s11306-008-0146-y. PubMed PMID:
550 WOS:000265686800008.

551 45. Steuer R. Review: On the analysis and interpretation of correlations in metabolomic data.
552 Briefings in Bioinformatics. 2006;7(2):151-8. doi: 10.1093/bib/bbl009.

553 46. Guan XL, Souza CM, Pichler H, Dewhurst G, Schaad O, Kajiwara K, et al. Functional Interactions
554 between Sphingolipids and Sterols in Biological Membranes Regulating Cell Physiology. Mol Biol
555 Cell. 2009;20(7):2083-95. PubMed PMID: WOS:000264752100018.

556 47. Shevchenko A, Simons K. Lipidomics: coming to grips with lipid diversity. Nat Rev Mol Cell Bio.
557 2010;11(8):593-8. PubMed PMID: WOS:000280201100015.

48. Morgenthal K, Wienkoop S, Scholz M, Selbig J, Weckwerth W. Correlative GC-TOF-MS-based metabolite profiling and LC-MS-based protein profiling reveal time-related systemic regulation of metabolite-protein networks and improve pattern recognition for multiple biomarker selection. Metabolomics. 2005;1(2):109-21. PubMed PMID: WOS:000204301700002.

49. Nicholson JK, Connelly J, Lindon JC, Holmes E. Metabonomics: a platform for studying drug toxicity and gene function. Nat Rev Drug Discov. 2002;1(2):153-61. PubMed PMID: WOS:000178731000018.

50. Beckles DM, Roessner U. 5 - Plant metabolomics: Applications and opportunities for agricultural biotechnology. In: Altman A, Hasegawa PM, editors. Plant Biotechnology and Agriculture. San Diego: Academic Press; 2012. p. 67-81.

Supporting information

S1 Fig. The relative amounts of the different groups of lipids in the coffee.

S1 Table. The lipid species identified with corresponding charge mode

S2 Table. Pearson's Correlation Coefficient Values, P-values and Adjusted P-values for Lipid-to-Lipid Associations.

S3 Table. Pearson's Correlation Coefficient Values, P-values and Adjusted *P*-values for Metabolite-to-Lipid Associations.

S4 Table. Pearson's Correlation Coefficient Values, P-values and Adjusted P-values for all Metabolites vs. All Metabolites Associations.

FIGURES

Lipidomic and metabolomic profiles of *Coffea canephora* L. beans cultivated and harvested from South-western Nigeria

Chinyere F. Anagbogu[1,2,3*], Jiaqi Zhou[1], Festus O. Olasupo[2,3], Mohammed Baba-Nitsa[3] and Diane M. Beckles[1*]

[1] Department of Plant Sciences, University of California, Davis, CA 95616

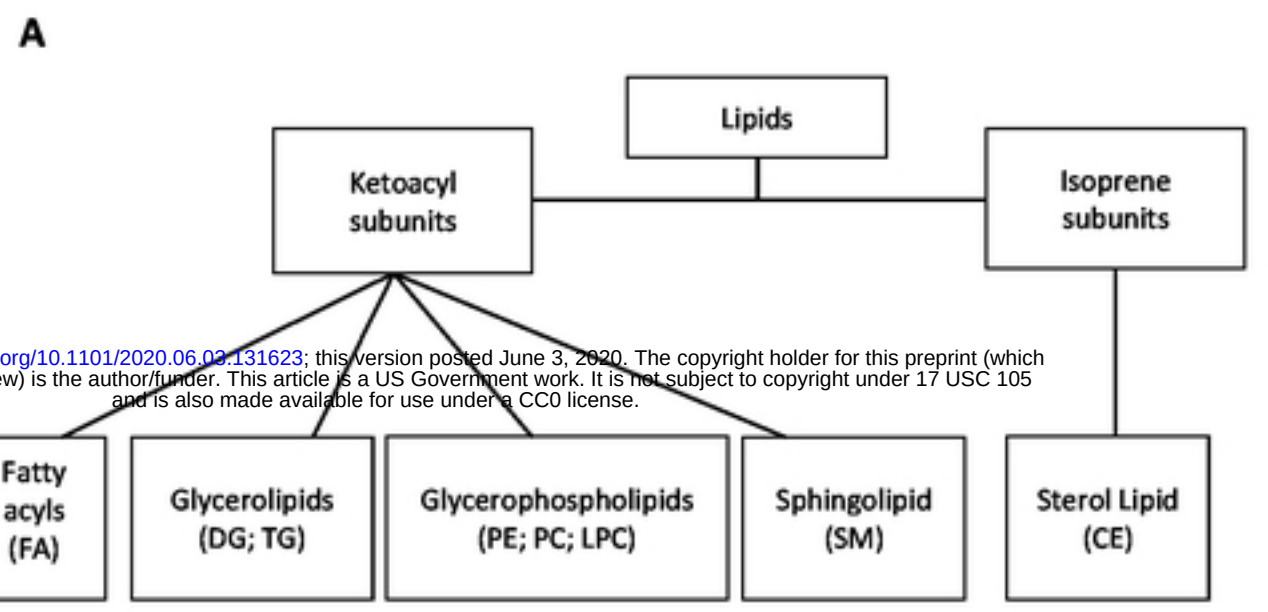
[2] Department of Crop Protection and Environmental Biology, University of Ibadan, Ibadan, Nigeria

[3] Crop Improvement Division, Cocoa Research Institute of Nigeria, Ibadan, Nigeria

[*] Corresponding authors

E-mail: flora2na@yahoo.com (CFA)

Email: dmbeckles@ucdavis.edu (DMB)



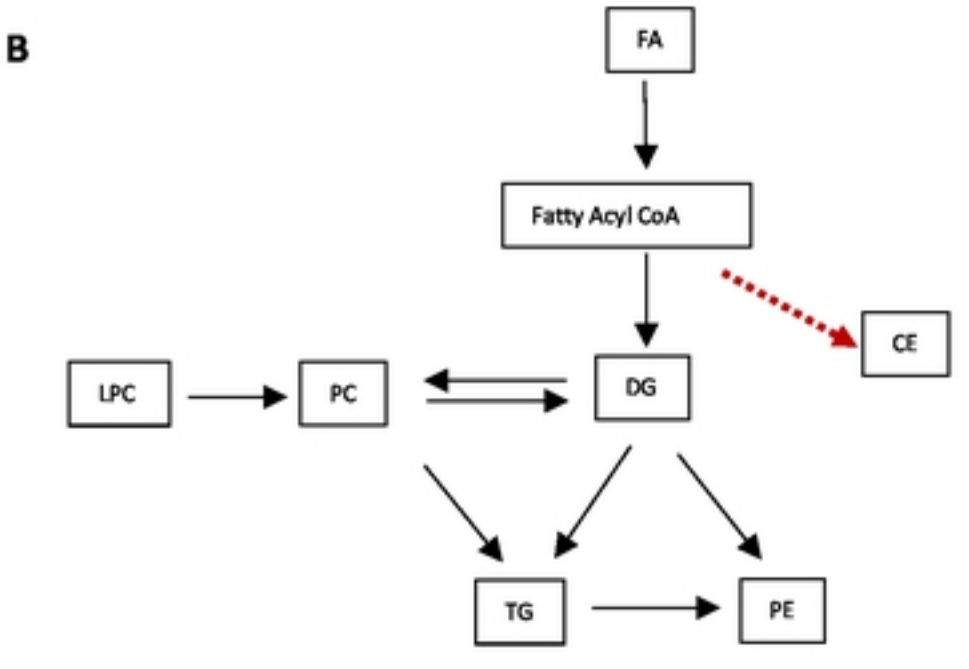


Fig 1. Lipid classes and their interrelationship. A) Lipid classes and subclasses detected in *C. canephora* beans. Classification is based on the LIPID MAPS (Metabolites and Pathways Strategy) system, which separates lipids based on their two fundamental "building blocks" - the ketoacyl group and the isoprene group [31]. Abbreviations: FA- Fatty acyls; DG - Diacylglycerol; TG- Triacylglycerols; PE - Phosphatidylethanolamine; PC -Phosphatidylcholine; LPC Lysophosphatidylcholine; SM-Sphingomyelins; CE- Cholesterol esters. **B)** A simplified view of a proposed biochemical interrelationship among the lipid classes according to data in [32,33], as well as in the MetaCyc database [34] derived from plants. The red dashed arrow indicates data obtained from human cells [35].

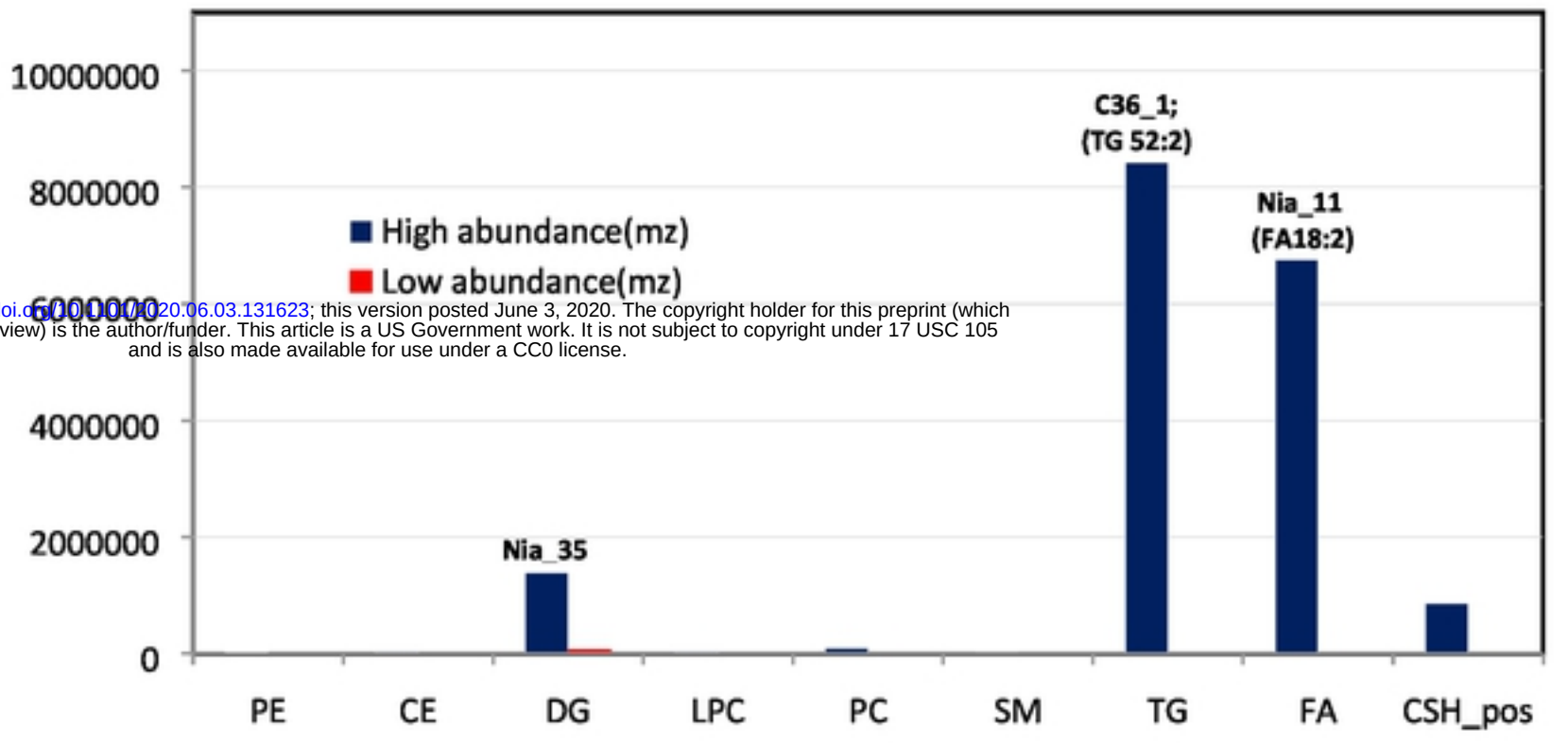


Fig 2. Relative lipid abundance in the *C. canephora* species studied. We identified nine subclasses of lipids within the five broader lipid classes described in LIPIDS MAPS. For each subclass, the relative abundance of compounds among each genotype was assessed. Values shown are the maximum and minimum values within each compound class, and the label indicates the genotype in which the highest value was found

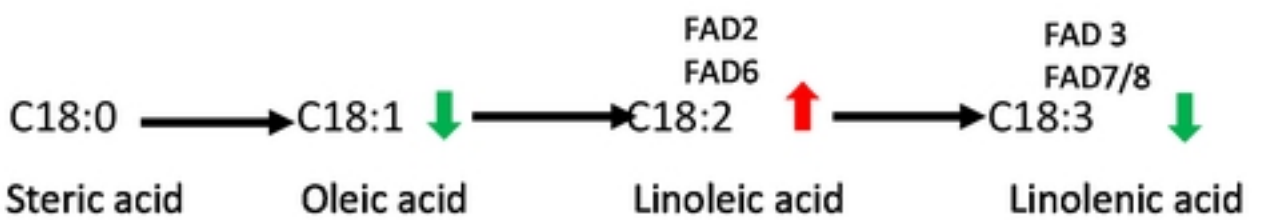


Fig 3. Desaturation of C18 fatty acids. This pathway is found in plastids (catalyzed by fatty acid desaturases - FAD6 and FAD7/8) and in the endoplasmic reticulum (catalyzed by FAD2 and FAD3). Arrows indicate the relative amount of each compound based on data derived from this study.

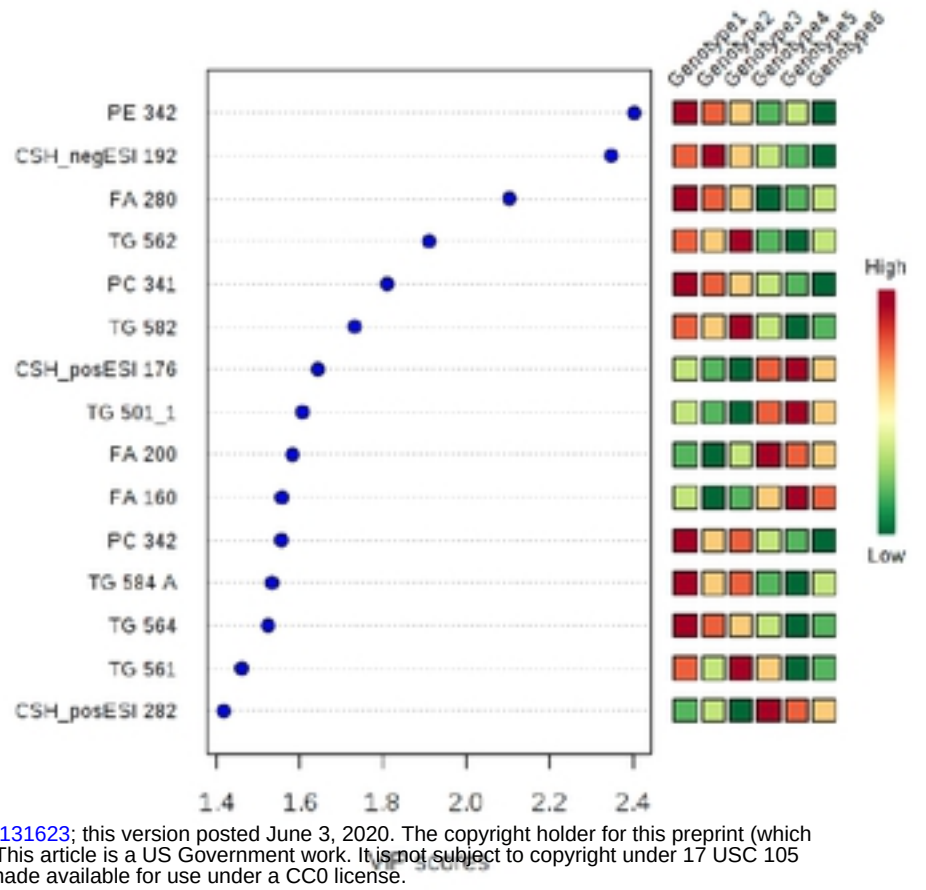


Fig 4. Potential lipidomic markers differentiating 'Niaouli' (genotypes groups 1–3) from 'Kouillou' (groups 4–5) and 'Java' (group 6). The variable importance in projection (VIP) scores on the x-axis provide an estimate of the contribution of a given predictor (lipid species shown on the y-axis) to the Partial Least Square (PLS) regression above. The higher the VIP score, the better the lipid species is as a predictor of the discrimination among the genotypes.

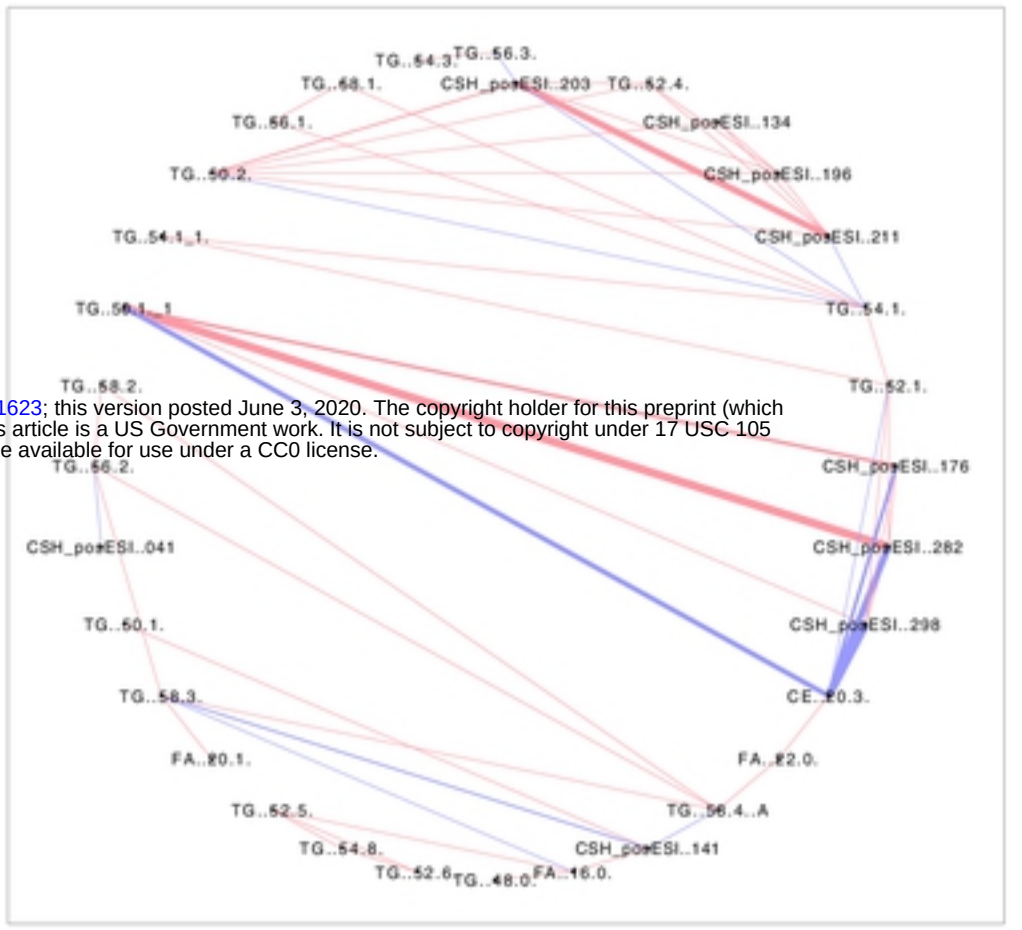


Fig 5. Pearson's correlation network illustrating correlation patterns among lipids. Correlations found among lipid data are shown as attribute circular layout. The blue lines indicate lipids whose occurrence negatively correlate, while the pink lines denote lipids that show a positive correlation. Line thickness indicates the strength of correlation. A permissive threshold level of ≥ | ± 0.75| was set for the Pearson's correlation coefficient and a *q*FDR of <0.05 was chosen for the identification of significant correlations within the lipid data.

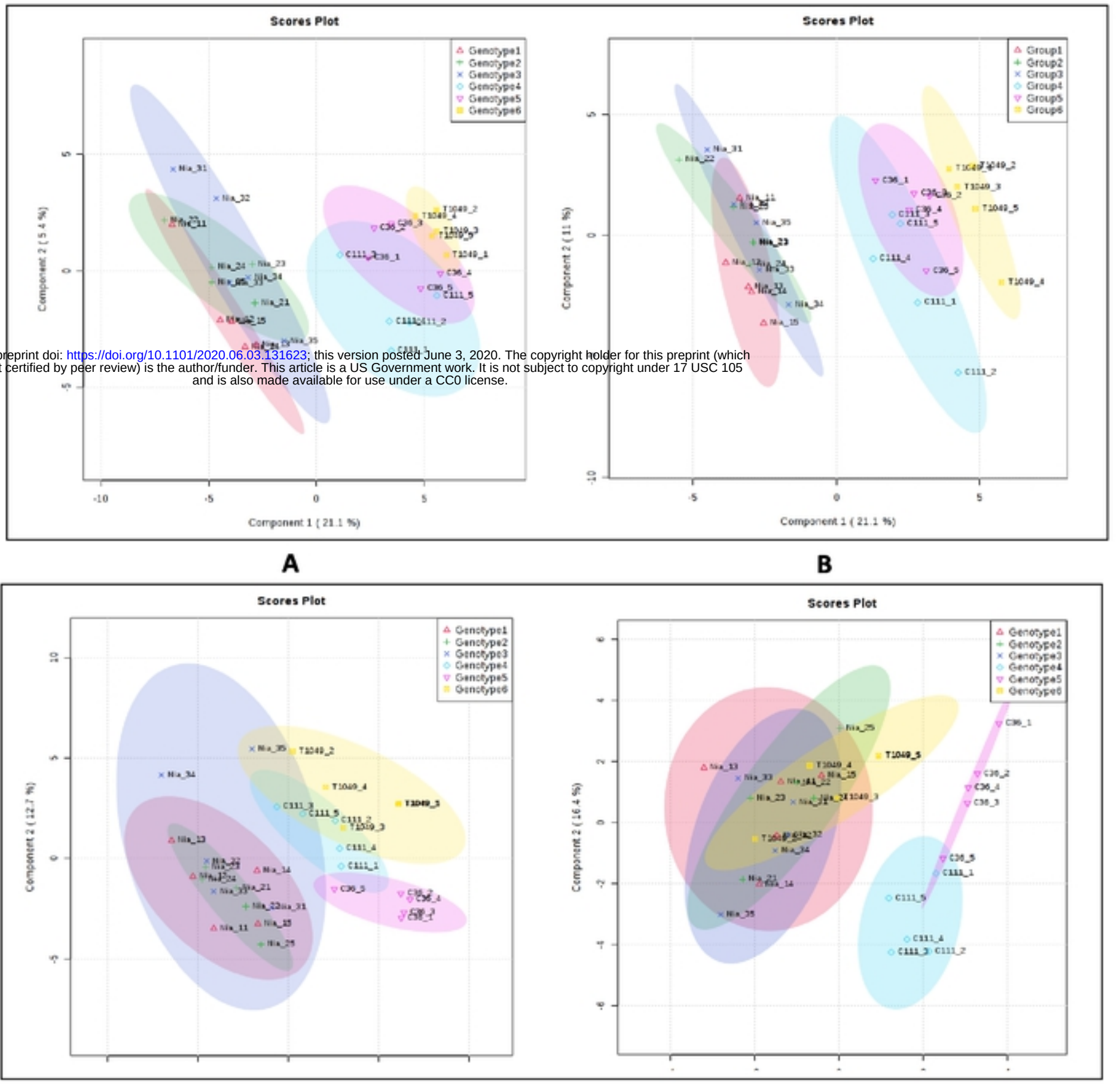


Fig 6. Differentiating among the *C. canephora* genotypes using multivariate analyses. **A)** Primary metabolites analyzed by PLS-DA, **B)** Secondary metabolites analyzed by PLS-DA, **C)** lipids analyzed by PLS and **D)** lipids analyzed by Sparse-PLS. (PCA = Principal Component Analysis, PLS-DA = Partial Least Square –discriminate analysis. Genotypes 1-3 belong to the 'Niaouli' group i.e. Nia_1, Nia_2 and Nia_3, Genotypes 4-5 are members of the 'Kouillou' group i.e. C111 and C36, and Genotype 6 is 'Java Robusta' i.e. T1049. The primary and secondary metabolite data were derived from Anagbogu *et al.*, 2019b [21].

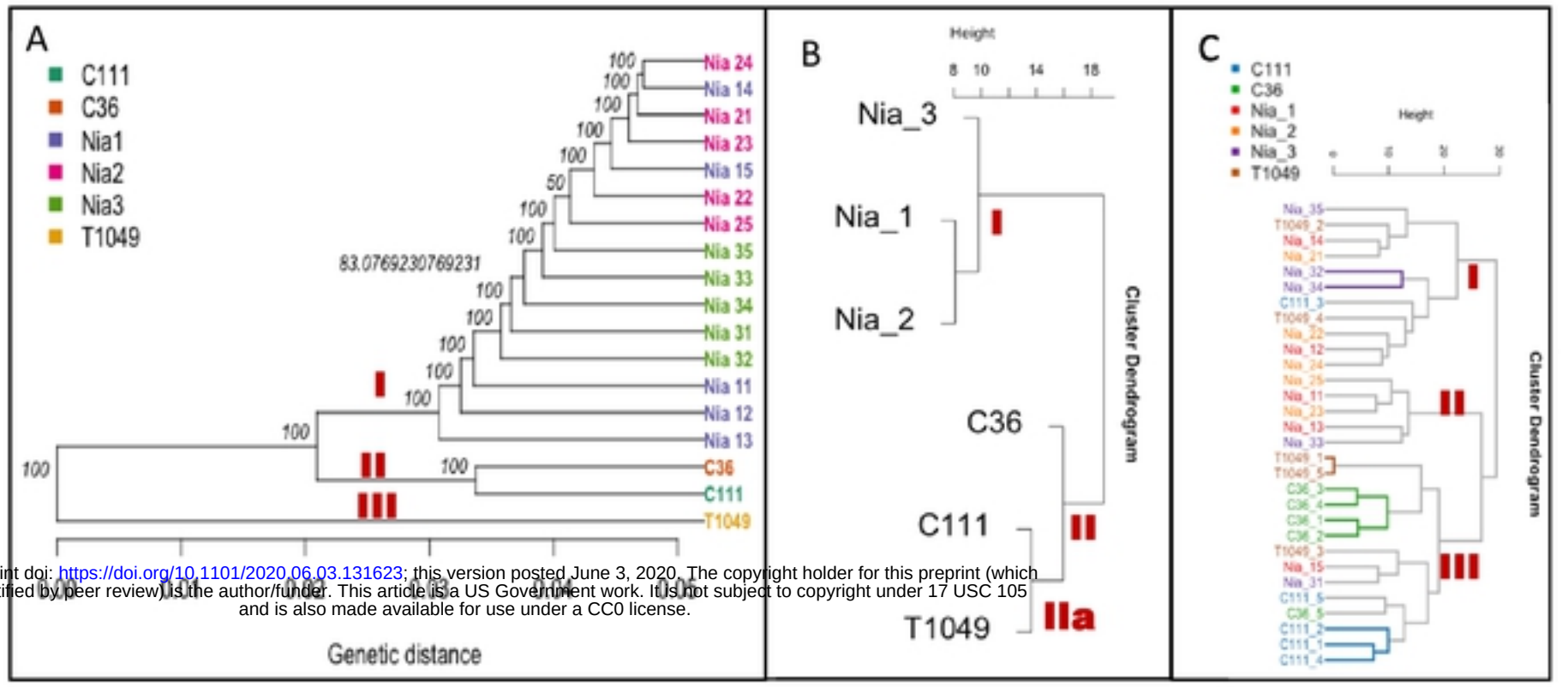


Fig 7. Comparative analysis of the genomics and lipidomic analysis based on their hierarchical structure. **A)** genetic analysis based on ~100,000 SNP, **B)** lipidomic analysis indicating the mean value of each genotype from five biological replicates and **C)** lipidomic analysis including all replicates. The genotypes and replicates names were described in Table 1 as follows: Nia_11- Nia_15 were Genotype 1, Nia_21-Nia_25 were genotype 2, Nia_31-Nia35 were Genotype 3 and belong to the 'Niaouli' group; Members of the 'Kouilou' group have been referred to as Genotypes 4 and 5 in this study and included C111_1-C111_5 and C36_1 to C36_5. 'Java Robusta' coffee has been described as Genotype 6 included T1049_1-T1049_5.