The elemental composition of chocolates is related to cacao content and origin: a multi-element fingerprinting analysis of single origin chocolates

Ruth VANDERSCHUEREN ^{a,*}, Daniela MONTALVO ^a, Bart DE KETELAERE ^b, Jan A. DELCOUR ^c and Erik SMOLDERS ^a

^a Division of Soil and Water Management, KU Leuven, Kasteelpark Arenberg 20, 3001 Leuven, Belgium

^b Division of Mechatronics, Biostatistics and Sensors (MeBioS), KU Leuven, Kasteelpark Arenberg 30, 3001 Leuven, Belgium

^c Centre for Food and Microbial Technology, KU Leuven, Kasteelpark Arenberg 20, 3001 Leuven, Belgium

* Corresponding author: Vanderschueren R., KU Leuven, Kasteelpark Arenberg 20, 3001 Leuven, Belgium. Phone: +32 16 37 33 73 E-mail: Ruth.Vanderschueren@kuleuven.be

ABSTRACT

Commercially available single origin chocolates (n=139) were analysed by ICP-MS to identify the potential of elemental fingerprinting for tracing cacao origin in chocolate and to compare chocolate composition relative to trace metal limits. Cadmium (Cd) concentrations exceeded the EU limit of 0.80 mg Cd kg⁻¹ in 16 samples, all produced with cacao from South or Central America. Six samples contained lead (Pb) concentrations > 0.10 mg kg⁻¹, the limit of the Codex Alimentarius for edible fats. Increasing cacao content was associated with increased element concentrations for most elements, indicating cacao as the main source of minerals and trace elements. Significant differences in elemental composition between origins (P value ≤ 0.05) were found for Ba, Cd, Mo and Sr. Classification and regression tree analysis (CART) resulted in a decision tree that could effectively classify chocolate samples by cacao origin continent (overall misclassification rate 23%) based on the concentrations of five elements (Ba, Cd, Mo, Sr and Zn). Samples of South America were classified based on their Cd concentration, indicating the geogenic origin of Cd.

1.INTRODUCTION

Chocolate is made from the seeds of the cacao tree (*Theobroma cacao* L.) through an extensive post-harvest process that includes fermentation, drying and roasting. Cacao is a perennial crop grown at low altitude in tropical climates (within 20° latitude from the equator). The plant originates from Central and South America where many cacao varieties can be found, and was exported to other suitable geographical areas such as Africa in the 19th century (Beckett, 2008). Several health benefits have been attributed to moderate cacao consumption mainly due to its high polyphenol content, which has been related to cardiovascular health (Corti et al., 2009; Hooper et al., 2012; Lee et al., 2003). However, cacao-derived products may contain potentially toxic trace elements, such as cadmium (Cd), nickel (Ni) and lead (Pb).

In 2014, the EU Commission approved Cd limits in chocolate that have been enforced since January 2019. These limits range from 0.10 mg Cd kg⁻¹ dry weight in chocolates with < 30% cacao solids to 0.80 mg Cd kg⁻¹ in cacao powder and chocolates with ≥ 50% cacao solids (European Commission, 2014). Recently, the Codex Alimentarius adopted similar limits for Cd in chocolate, ranging from 0.80 mg Cd kg⁻¹ for chocolates that contain between 50% and 70% cacao solids to 0.90 mg Cd kg⁻¹ for chocolates with ≥ 70% cacao solids (Codex Alimentarius Commission, 2018). The European Food Safety Authority reviewed the exposure of the European population to Ni in food. This study concluded that chronic dietary exposure to Ni is of concern for the general European population and proposed a tolerable daily intake level of 2.8 µg Ni kg⁻¹ bodyweight (EFSA CONTAM Panel, 2015). In addition, the review highlighted that cacao based products are among the main food sources with high levels of Ni, e.g. some types of chocolate (average 3.8 mg Ni kg⁻¹) and cacao beans and derived products (average 9.5 mg Ni kg⁻¹). This information prompted the European Commission to recommend that the member states monitor the Ni content of foods from 2016 until 2018, which may result in limits for Ni in chocolate in the future (European Commission, 2016). The EU regulation currently does not state explicit limits for Pb in chocolate, however the European Commission Regulation on maximum levels for contaminants in foodstuffs states that fats and oils (including cacao butter which is one of the main ingredients of chocolate) should not contain more than 0.10 mg Pb kg⁻¹ (European Commission, 2006). This limit of 0.10 mg Pb kg⁻¹ is endorsed by the Codex Alimentarius for edible fats (Codex Alimentarius Commission, 2017), and by the US Food and Drug Agency (US FDA) for candy destined for young children (U.S. FDA, 2006).

The presence of trace metals in chocolate may be influenced by several factors. The concentrations of different minerals in chocolate increase with increasing cacao content (Abt et al., 2018; Lo Dico

et al., 2018; Villa et al., 2014; Yanus et al., 2014). This indicates that these elements mainly originate from the cacao and not from other ingredients or contamination during processing. The mineral content of cacao is affected by its geographical origin. Notable geographical differences in the elemental composition of cacao beans have been reported (Bertoldi et al., 2016). Chocolate is often made using a mix of cacao beans from different origins. However, single origin chocolates are becoming increasingly popular and are often marketed as luxury products, which raises interest in the influence of cacao origin on the quality and food safety of the final product. The composition of single origin chocolates allows to trace back the origin of the cacao, provided that the geogenic signal in the composition is strong and conserved. This has already been proven with cacao beans: multi element fingerprinting analysis followed by principal component analysis (PCA) and discriminant analysis (DA) have been used to identify the geographical origin of cacao beans (Acierno et al., 2016; Acierno et al., 2017; Bertoldi et al., 2016; Caligiani et al., 2014; Caligiani et al., 2016; Cambrai et al., 2010; D'Souza et al., 2017; Diomande et al., 2015; Kumari et al., 2018). Much less information is available for the similar analysis of the final product, i.e. chocolate. Junior et al. (2018) used DA based on elemental composition to differentiate between organic and conventional chocolates and Bertoldi et al. (2016) used 13 chocolates as a validation set to test the efficiency of their DA model for the classification of cacao beans by geographical origin.

The present study was set up to assess the elemental composition of single origin chocolates and to identify the potential for tracing cacao origin by element fingerprinting of chocolates. This information may shed light on the impact of trace metal regulations on the cacao market as well as on the role of geographical origin. Classification and regression tree analysis is used to identify elements that most strongly characterise the cacao origin.

2. MATERIALS AND METHODS

2.1. Sample collection and preparation

Between February and May 2018, 139 single origin chocolates [not including white chocolates and/or chocolates with inclusions (e.g. fruits or nuts) or noticeable food colouring] were purchased online and in commercial retail points from 47 different manufacturers in several countries (Belgium, n = 86; Colombia, n = 3; Ecuador, n = 25; France, n = 10; Italy, n = 3; Madagascar, n = 6; Mexico, n = 2; New Zealand, n = 2; Peru, n = 1 and Switzerland, n = 1). The samples were divided into four regions based on the cacao origin mentioned on the packaging: Africa (n = 34), Asia Pacific (n = 14), Central America (n = 22) and South America (n = 69) (Table 1). Chocolates were grated using a manual plastic grater and 100 mg of each sample was digested with 3.0 mL of concentrated Suprapur[®] nitric acid (HNO₃, 65% w/w; Merck, Darmstadt, Germany) in an open block acid digestion for 11 hours at 130 °C. The obtained digests were diluted five times with Milli-Q water (18.2 MΩ cm⁻¹) and the total elemental composition of macro (Ca, K, Mg, Na, P, S) and trace elements (Al, As, B, Ba, Be, Cd, Co, Cr, Cs, Cu, Fe, Ga, In, Li, Mn, Mo, Ni, Sb, Se, Sn, Sr, Pb, Ti, Tl, U, V, Zn, Zr) was determined by inductively coupled plasma mass spectrometry (ICP-MS, Agilent 7700x, Agilent Technologies, Santa Clara, USA). Blank samples were included in quadruplicate in each digestion batch and treated in the same way as the samples.

2.2. Instrument calibration and sample analysis

Multi element analysis was performed using ICP-MS equipped with a collision cell, in the No Gas mode for ²⁷Al, ¹¹B, ¹³⁷Ba, ⁹Be, ¹³³Cs, ¹¹⁵In, ⁷Li, ²⁰⁸Pb, ¹²¹Sb, ¹¹⁸Sn, ²⁰⁵Tl, ²³⁸U and ⁹⁰Zr; in the Helium mode (5 mL min⁻¹) for ⁷⁵As, ⁴⁴Ca, ¹¹¹Cd, ⁵⁹Co, ⁵²Cr, ⁶³Cu, ⁵⁶Fe, ⁶⁹Ga, ³⁹K, ²⁴Mg, ⁵⁵Mn, ⁹⁵Mo, ²³Na, ⁶⁰Ni, ⁷⁸Se, ⁸⁸Sr, ⁴⁷Ti, ⁵¹V and ⁶⁶Zn; and in the High Energy Helium mode (10 mL min⁻¹) for ³¹P and ³⁴S. An internal standard containing ⁷²Ge, ¹⁰³Rh and ¹⁹³Ir was added online to correct for measurement drift. Calibration for all elements was performed using Certipur[®] multi element

certified standard solutions (Merck, Darmstadt, Germany). Measurements were performed in triplicate and instrument calibration was validated using three certified reference solutions (Reference Material for Measurement of Elements in Surface Waters, Spectrapure Standards, Oslo Norway; SRM 1643f Trace Elements in Water, National Institute of Standards and Technology, Maryland USA; and TM-28.4 Trace Elements in Water, Environment and Climate Change, Canada). All solutions and samples for ICP-MS analyses were acidified to 1% HNO₃ (Suprapur®, Merck, Darmstadt, Germany). Measurements below the limit of quantification (LOQ) were reported as <LOQ. The LOQ is defined here as 10 times the standard deviation of the four procedural blanks, reconverted to a chocolate weight based concentration using equal sample weight to digest volume.

2.3. Quality assurance

Two chocolate certified reference materials were included in triplicate in each digestion batch as quality assurance; NIST 2384 baking chocolate certified for Ca, Cd, Cu, Fe, K, Mg, Mn, P, Pb and Zn; and ERM®-BD512 dark chocolate certified for Cd, Cu, Mn and Ni. Due to poor recovery of the reference materials, Fe was excluded from further analysis. For all other elements, recoveries ranged between 79% and 102% (Supplementary Information, Table S 1). To verify reproducibility, duplicate samples were included at intervals of 15 samples. The coefficient of variation (CV) was calculated for these duplicates (13 duplicate sets) and ranged between 1% and 23% depending on the element (overall average CV = 6%). Gallium was excluded as a strong correlation was found with Ba ($R^2 = 0.99$), likely caused by isobaric interference with the Ba²⁺ ion which is also detected at $m/z = 69$.

2.4. Statistical analysis

131 All statistical analysis was executed using JMP[®] Pro version 14.0.0 (SAS Institute 2018). Elements
132 for which more than 15% of the data points were below LOQ (Al 19%, As 100%, Be 96%, In 90%,
133 Li 81%, Na 84%, Pb 60%, Sb 86%, Se 41%, Sn 98%, Ti 38%, Tl 59%, U 81% and Zr 91%) were
134 excluded from statistical analysis, however the raw data are given in the Supplementary
135 Information (Table S 2). Statistical analyses were thus performed using 18 elements (B, Ba, Ca,
136 Cd, Co, Cr, Cs, Cu, K, Mg, Mn, Mo, Ni, P, S, Sr, V and Zn). Measurements below LOQ were
137 replaced by LOQ/2 and concentrations were log transformed prior to statistical analysis to ensure
138 normal distribution of the data, except for classification and regression tree (CART) analysis as
139 this technique does not require normal distribution of the data.

140 A one factor Analysis of Variance (ANOVA) was used to investigate the effect of different origins
141 on the 18 elements considered (Kutner et al., 2005). After a positive omnibus F-test, a post-hoc
142 Tukey's Honestly Significant Difference test was used to identify significant differences between
143 the average elemental concentrations in chocolates from the different origins (P value ≤ 0.05). This
144 analysis only allows to consider each of the 18 elements separately, but does not give insight in
145 the overall differences in elemental composition. In order to overcome this limitation, CART
146 analysis was performed to provide a combined analysis that allows for an appealing visualisation
147 of the differences amongst the regions (Breiman et al., 1984). The CART method is a non-
148 parametric technique that identifies variables with the power to classify samples (in this case
149 according to the origin of cacao in chocolates) and uses these variables to build a decision tree.
150 Each node in the decision tree corresponds to a conditional IF-statement in the model and splits
151 the dataset in two subgroups based on the cut-off value of a single variable. Samples that are higher
152 than the cut-off value for this variable belong in one subgroup and samples with lower values
153 belong in the other. Splits (and cut-off values) are chosen so that the difference in the average

response (in this case the origin) between the subgroups is maximized, consecutive splits thus result in subgroups that are more homogeneous (with regard to origin). At each node and leaf of the decision tree, the model calculates probabilities for each category. Although this was not the prime goal of this work, these probabilities can then be used to predict the classification of unknown samples. The cacao content differed strongly between chocolate samples (Table 1) and as stated in the introduction, the elemental composition of chocolates is strongly influenced by their cacao content. Therefore, CART analysis was applied to the residuals of a linear regression of the considered element concentrations against the cacao content of the samples. This correction mimics the real-life situation in which the origin of a chocolate sample is to be determined given its known cacao content. To prevent over fitting of the CART model, a fivefold cross-validation was used.

3. RESULTS AND DISCUSSION

3.1. The elemental composition of chocolates and its relation to origin and cacao content

The 139 samples originated from 27 countries and were split in four regions (Table 1). The fractions of collected samples for each region were 24% (Africa), 10% (Asia Pacific), 16% (Central America) and 50% (South America). Globally, cacao production is highest in Africa (73%), followed by Central and South America (together 17%) and Asia Pacific (10%) (based on cacao production between October 2015 and September 2016, (ICCO, 2018)). This indicates a sample bias towards South America. However, that bias was likely related to the single origin market as South American cacao is renowned and respected for its high quality. For example, the Nacional variety native to Ecuador is known for its aromatic flavour (Afoakwa, 2010; Beckett, 2008). The mean cacao content (%) in the collected samples was unaffected by origin ($P > 0.05$).

Median, minimum and maximum concentrations of all measured elements can be found in the Supplementary Information (Table S 2). One sample (Tanzania, 70% cacao) was considered an outlier and excluded from analysis due to its extremely high Cr concentration (24 mg Cr kg⁻¹). Of all elements, the macro elements Ca (median 720 mg kg⁻¹), K (median 6437 mg kg⁻¹), Mg (median 1887 mg kg⁻¹), P (median 2748 mg kg⁻¹) and S (median 925 mg kg⁻¹) were present in chocolate in the largest concentrations, with the exception of Na which was below LOQ in 84% of the samples. The most prevalent micro element was Zn (median 27 mg kg⁻¹) followed by Mn (median 14 mg kg⁻¹) and Cu (median 13 mg kg⁻¹). Concentrations of Al, As, Be, In, Li, Na, Pb, Sb, Se, Sn, Ti, Tl, U and Zr were below LOQ in more than 15% of the samples and these elements were therefore excluded from statistical analysis.

Single correlation analyses between element concentrations and cacao content of the chocolates were significant for all analysed elements (Pearson's correlation coefficient, P value $\leq$ 0.05), except for Cr and V (Table 2), suggesting that cacao is the main origin of these minerals in chocolate. Calcium was the only element that negatively correlated to cacao content, which can be explained by the presence of Ca in milk powder used in the preparation of milk chocolates. Since most minerals originate from the cacao, it might be argued to normalize all data for cacao content (%) to identify the geogenic signature of, for example, 100% cacao solids. However, single correlations between normalized element concentrations (to 100% cacao solids) and cacao content were lower, but still significant (P value $\leq$ 0.05) after normalization of the data to 100 % cacao solids, for B (r = 0.54), Mg (r = 0.60), Mn (r = 0.49), Co (r = 0.17), Ni (r = 0.31), Cu (r = 0.45), Sr (r = 0.30) and Cd (r = 0.24). In the sections below, we first discuss the elemental concentration without normalisation since, for example, metal limits are expressed on the final product and since

the cacao contents in the samples were similar, e.g. the mean cacao content (%) was unaffected by origin. However, for CART analysis data were corrected for cacao as explained previously.

The average elemental concentrations for each region are reported in Table 3. The average concentrations of Ba, Cd, Mo, and Sr significantly differed between regions (P value ≤ 0.05). No significant differences were observed in the macro element (Ca, K, Mg, P and S) concentrations between the different regions.

3.2. Presence of potentially toxic trace elements in chocolate and its relation to cacao content and origin

Arsenic concentrations were below the LOQ ($0.08 \text{ mg As kg}^{-1}$) in all samples, which is in accordance with the results reported by Lo Dico et al. (2018). No regulations exist so far regarding the concentration of As in chocolate but measured values were below the lowest threshold value for As in rice ($0.10 \text{ mg As kg}^{-1}$, for rice for the production of food for infants or young children (European Commission, 2015)).

Nickel concentrations ranged between 0.3 and $12.5 \text{ mg Ni kg}^{-1}$, which is in accordance to values reported in literature by Bertoldi et al. (2016) (3.3 to $4.7 \text{ mg Ni kg}^{-1}$ depending on the origin of the chocolates), Junior et al. (2018) (on average 2.6 and $2.8 \text{ mg Ni kg}^{-1}$ in organic and conventional chocolates respectively), Mrmošanin et al. (2018) (between 1.90 to $5.90 \text{ mg Ni kg}^{-1}$ in ten chocolates purchased in Serbia) and Dahiya et al. (2005) (on average $2.76 \text{ mg Ni kg}^{-1}$ in cacao based candy). Nickel concentrations were strongly correlated with cacao content ($r = 0.55$) indicating that at least some part originates from the raw material (Table 3). This observation was previously corroborated by the European Food Safety Authority which reported on average higher

Ni concentrations in cacao beans and derived products ($9.5 \text{ mg Ni kg}^{-1}$) compared to the final chocolate products ($3.8 \text{ mg Ni kg}^{-1}$) (EFSA CONTAM Panel, 2015).

The Pb limit of 0.10 mg kg^{-1} as referred to in the introduction, was exceeded in six chocolate samples originating from India (0.11 mg kg^{-1} purchased in Belgium), Trinidad (0.12 mg kg^{-1} purchased in Belgium), Peru (0.68 mg kg^{-1} purchased in Belgium) and Ecuador (0.17 and 0.36 mg kg^{-1} purchased in Ecuador and 0.10 mg kg^{-1} purchased in Belgium). Values for Pb were below LOQ ($0.02 \text{ mg Pb kg}^{-1}$) in 60% of the samples and the median concentrations per region were also lower than the LOQ. To test whether there was a relationship between the Pb concentrations in the chocolates and their cacao content and/or origin, a categorical variable with two values ($>\text{LOQ}$ and $<\text{LOQ}$) was used. No significant relationship (X^2 statistic, $\alpha = 0.05$) was found between the proportions of detectable Pb concentrations and either cacao content or origin. This indicates that Pb in chocolates likely originates from the production process rather than the raw material (the cacao beans). The measured Pb concentrations were lower than those mentioned by Mrmošanin et al. (2018) (0.25 to $0.86 \text{ mg Pb kg}^{-1}$) but higher than the values reported by Bertoldi et al. (2016) (0.008 to $0.024 \text{ mg Pb kg}^{-1}$). Abt et al. (2018) reported significant differences in the average Pb concentrations measured in milk chocolate ($0.01 \text{ mg Pb kg}^{-1}$) compared to those in dark chocolate ($0.03 \text{ mg Pb kg}^{-1}$) and cacao powder ($0.11 \text{ mg Pb kg}^{-1}$) purchased on the US market. However, their results showed much lower Pb concentrations in the raw product ($0.003 \text{ mg Pb kg}^{-1}$ in cacao nibs), which again suggests that Pb in chocolate originates from the production process rather than the raw material. In contrast, Villa et al. (2014) did find a significant correlation between cacao content and Pb concentration in 30 chocolates purchased in Brazil.

From the total number of samples collected for this study, 16 chocolates (all $> 50\%$ cacao solids) exceeded the new EU regulation limit for Cd ($0.80 \text{ mg Cd kg}^{-1}$ for chocolate with $\geq 50\%$ cocoa

solids) (Figure 1). A correlation was found between cacao content and Cd concentration ($r = 0.45$, Table 3). This observation is in line with reports in literature (Abt et al., 2018; Lo Dico et al., 2018; Villa et al., 2014). Abt et al. (2018) measured on average higher Cd concentrations in cacao nibs ($0.62 \text{ mg Cd kg}^{-1}$) and cacao powder ($0.70 \text{ mg Cd kg}^{-1}$) in comparison to dark chocolate ($0.27 \text{ mg Cd kg}^{-1}$) and milk chocolate ($0.06 \text{ mg Cd kg}^{-1}$). Likewise, Lo Dico et al. (2018) reported median Cd concentrations increasing from chocolate with $< 30\%$ cacao solids ($< \text{LOQ}$) to chocolate with 30 to 50% cacao solids ($0.019 \text{ mg Cd kg}^{-1}$), chocolate with $> 50\%$ cacao solids ($0.116 \text{ mg Cd kg}^{-1}$) and finally cacao powder ($0.159 \text{ mg Cd kg}^{-1}$); and Villa et al. (2014) found a linear relationship between the cacao contents and the Cd concentrations measured in chocolates from Brazil.

From the chocolates collected for this study, samples originating from South America contained significantly higher concentrations of Cd compared to chocolates from Africa and Asia Pacific (Table 2) and the chocolates that exceed the new EU limits were all from Central and South America (Figure 1): Haiti ($1.44 \text{ mg Cd kg}^{-1}$, purchased in France); Peru (1.07 and $1.34 \text{ mg Cd kg}^{-1}$ purchased in Belgium and $0.81 \text{ mg Cd kg}^{-1}$ purchased in Peru); Colombia ($1.68 \text{ mg Cd kg}^{-1}$ purchased in Italy; 0.94 and $0.97 \text{ mg Cd kg}^{-1}$ purchased in Colombia; 1.22 , 2.30 and $2.61 \text{ mg Cd kg}^{-1}$ purchased in Belgium) and Ecuador (0.81 , 0.91 , 0.93 , 1.66 and $2.00 \text{ mg Cd kg}^{-1}$ purchased in Ecuador and $0.85 \text{ mg Cd kg}^{-1}$ purchased in Belgium). As shown by the results of our dataset, the concentration range for Cd was much larger in the South American sample set (from 0.02 up to $2.61 \text{ mg Cd kg}^{-1}$) compared to that from other origins (Supplementary Information, Table S2). This supports the hypothesis of Cd hotspot areas rather than a homogeneous geographical presence, where elevated Cd uptake by the cacao plant and consecutive accumulation in the cacao beans can result in high Cd chocolates. In a recent survey conducted in the cacao growing areas in Ecuador, Argüello et al. (2018) reported spatial variability on Cd concentrations in cacao beans (39%) and

cacao cultivating soils (32%) within a large number of surveyed fields and identified clear cacao bean Cd hotspot areas (of geogenic origin) based on prediction maps. Both the measured concentrations and geographical differences observed in the present study are in accordance to findings reported in literature for unprocessed and processed cacao. For example, Bertoldi et al. (2016) reported Cd concentrations ranging from 0.09 to 1.39 mg Cd kg⁻¹ in cacao beans and 0.14 to 0.62 mg Cd kg⁻¹ in chocolates from different origins. They found that Cd concentrations in cacao beans from South America were on average four times higher than those in cacao beans from Asia and more than ten times higher than cacao beans from West Africa. Mounicou et al. (2003) also reported high average Cd concentrations in cacao powders from Venezuela and different provinces in Ecuador (0.533 to 1.833 mg Cd kg⁻¹) compared to Ivory Coast (0.094 mg Cd kg⁻¹) and Ghana (0.133 mg Cd kg⁻¹).

To obtain a desirable product (with regard to flavour and cost), cacao from different geographical origins is often mixed during the chocolate production process. The large differences in Cd concentration in cacao from different origins indicates that this mixing strategy may be a suitable technique to ensure that the final product complies with the new Cd regulations. For this study, single origin chocolates were selected specifically since they will likely be most affected by the new EU Cd regulation, as they can only be produced using cacao beans from the indicated country. The large variation within the regions also indicates that blending strategies within countries may be a suitable technique to ensure that single origin chocolates comply with the EU regulation.

3.3. Classification and regression tree (CART) analysis

Due to the large number of unknowns and complexity of the recipes, fingerprinting of the final chocolate product requires robust methods. Classification and regression tree (CART) analysis is a non-parametric technique that has been used to verify the origin of food products based on their

288 elemental composition, e.g. to verify the origin of wine (Capron et al., 2007; González et al.,
289 2009). Here, fivefold cross-validated CART analysis was used to identify elements that most
290 strongly characterise chocolates from different origins and to generate a decision tree for the
291 classification of chocolates by origin, based on their elemental composition. The CART method
292 was chosen over other chemometric techniques that are often used to generate elemental
293 fingerprints in food (e.g. principal component analysis and discriminant analysis) because it does
294 not rely on the normality assumption and is more robust to the influences of the unknown variables
295 in the chocolate recipe. As mentioned previously, the elemental composition of chocolate bars is
296 significantly correlated to their cacao content. To correct for this relation, CART analysis was
297 applied to the residuals of the linear regression fit of each element to the cacao content of the
298 sample. The parameter estimates of these regressions can be found in the Supplementary
299 Information (Table S 3).

300 Five elements (Ba, Cd, Mo, Sr and Zn) capable of separating the dataset using seven splitting
301 nodes (Cd and Mo were both repeated in two nodes, Figure 2) were identified by CART analysis.
302 The first node of the decision tree was determined by Cd and divided the dataset in two subgroups.
303 The subgroup with high Cd residuals (which corresponds to a higher Cd concentration than what
304 would be expected based on the cacao content of the sample) was dominated by South American
305 samples, whereas its counterpart was made up of a mixture of origins. As such, the analysis
306 identified Cd concentration as the first variable capable of discriminating South American
307 chocolates from other origins, which is in agreement with the statistically higher average Cd
308 concentration that was measured in South American chocolates (Table 2). The left branch was then
309 split by Mo (node 2), further separating South America from Central America. Similarly, node 3
310 also separated South American samples from Central American and Asia Pacific samples based

on Mo. This was in accordance with the statistically higher Mo concentration measured in South American chocolates ($0.24 \text{ mg Mo kg}^{-1}$) compared to samples from Central America ($0.10 \text{ mg Mo kg}^{-1}$) and Asia Pacific ($0.10 \text{ mg Mo kg}^{-1}$) (Table 2). The fourth node was determined by Ba and separated African samples from a mixture of origins. This was again in accordance with the results of the Tukey test (Table 2), which showed that the average Ba concentration in African samples was statistically higher compared to chocolates from the other regions. In node 5, the dataset was split based on Sr, which separated African chocolates but divided the group of South American chocolates evenly, and the final node (node 6) was determined by Zn although overall no significant difference between the average Zn concentrations per origin was identified (Table 2). Bertoldi et al. (2016) showed similar differences in the average concentrations of Cd and Ba in cacao beans from different origins (Cd concentrations were significantly higher in Central and South American cacao beans compared to Asian and West African cacao, and Ba was reported significantly higher in cacao beans from East African cacao beans compared to all other origins). However, they measured higher Mo concentrations in East Africa compared to Asia, Central America and West Africa, which was not observed in the present study. The elements identified by the CART model (Cd, Ba, Mo, Sr and Zn) were also reported by Bertoldi et al. (2016) as part of their discriminant model (obtained through stepwise discriminant analysis) for the separation of cacao beans based on their geographical origin.

The CART model was evaluated through several aspects. First, the fivefold cross-validated misclassification rate was 23%. This represents the fraction of the data for which the predicted category (or origin) was not equal to the observed category (the origin mentioned on the sample packaging). The confusion matrix displays more details regarding these misclassified samples (Figure 2 B). Second, the probabilities of each origin in the leaves (= endpoints) of the decision

tree provide an additional indication of how well the model was able to separate the origins from each other. High probabilities in the leaves were obtained for South American (0.99 in leaf b) and African (0.83 in leaf g) samples. The efficiency of the model for South America and Africa was corroborated by the low misclassification rates for these regions (only 10% of the African samples and 13% of the South American samples were misclassified). Separation of the Asia Pacific samples from the rest of the dataset was unsuccessful as all Asia Pacific samples were misclassified. This may be explained by the large geographical heterogeneity and corresponding differences in soil compositions and climate within the Asia Pacific sample group as it is comprised of a combination of islands and mainland countries. This likely resulted in a larger variety of elemental compositions complicating the separation of these samples from the bulk data set. Also, the small sample size of this group may have caused it to be underrepresented in our dataset. As the CART method strives to minimize the overall error rate, underrepresentation of a sample group may cause it to be poorly predicted by the model. Smaller misclassification rates may be reached using other parameters instead of total element concentrations, e.g. stable isotope ratios (Diomande et al., 2015; Perini et al., 2016).

4.CONCLUSION

This study revealed that 12% of the 139 collected single origin chocolates (15 samples from South America and one from Central America) exceeded the Cd limit for trade in the European market. A significant and strongly positive correlation was found between the Cd concentration measured in the chocolates and the cacao content, suggesting that the trace element most likely comes from the raw material (the cacao beans) rather than other ingredients or processing. The average Cd concentration was significantly higher in South American chocolates compared to chocolates from Africa and Asia Pacific, indicating the geogenic origin of the element as argued in several cacao

surveys. This observation was validated by the CART model, which identified Cd as the first node, splitting the dataset in a group that comprised mostly South American samples and a group with samples from all four regions. The recommended maximum Pb concentration of 0.10 mg Pb kg⁻¹ was exceeded in six samples originating from Ecuador, India, Peru and Trinidad but data suggest Pb concentrations were unrelated to the raw cacao material and were likely the result of post-harvest processing. This is not unexpected given the ubiquity of Pb in the environment. Classification and regression tree analysis proved a suitable technique to trace the origin of chocolates based on their elemental composition. It provided both a decision tree to classify unknown samples and information regarding the elements that most characterise chocolates from each origin. The total misclassification rate was 23% and was mostly caused by samples from Asia Pacific which were not separated from the rest of the data set.

SUPPORTING INFORMATION

The supplementary information includes the recoveries of the certified reference materials (Table S 1) and median, min and max concentrations of all elements measured in chocolates from four different origins (Africa, Asia Pacific, Central America and South America).(Table S 2).

FUNDING

This work was supported by the Research Foundation Flanders (FWO-Vlaanderen). RV is the recipient of a predoctoral fellowship from FWO-Vlaanderen.

CONFLICT OF INTEREST

The authors declare that there was no conflict of interest in the creation of this manuscript.

REFERENCES

- Abt, E., Sam, J. F., Gray, P., & Robin, L. P. (2018). Cadmium and lead in cocoa powder and chocolate products in the U.S. market. *Food Additives & Contaminants: Part B*, 11(2), 92–102. <https://doi.org/10.1080/19393210.2017.1420700>
- Acierno, V., Yener, S., Alewijn, M., Biasioli, F., & van Ruth, S. (2016). Factors contributing to the variation in the volatile composition of chocolate: Botanical and geographical origins of the cocoa beans, and brand-related formulation and processing. *Food Research International*, 84, 86–95. <https://doi.org/10.1016/j.foodres.2016.03.022>
- Acierno, V., Alewijn, M., Zomer, P., & van Ruth, S. M. (2017). Making cocoa origin traceable: fingerprints of chocolates using Flow Infusion - Electro Spray Ionization - Mass Spectrometry. *Food Control*, 85, 245–252. <https://doi.org/10.1016/j.foodcont.2017.10.002>
- Afoakwa, E. O. (2010). *Chocolate science and technology*. Wiley-Blackwell. <https://doi.org/10.1002/9781444319880>
- Argüello, D., Chavez, E., Lauryssen, F., Vanderschueren, R., Smolders, E., & Montalvo, D. (2018). Soil properties and agronomic factors affecting Cd concentration in cacao beans: A nationwide survey in Ecuador. *Science of the Total Environment*, 649, 120–127. <https://doi.org/10.1016/j.scitotenv.2018.08.292>
- Beckett, S. T. (2008). *Industrial chocolate manufacture and use*. Wiley-Blackwell. <https://doi.org/10.1002/9781444301588>
- Bertoldi, D., Barbero, A., Camin, F., Caligiani, A., & Larcher, R. (2016). Multielemental fingerprinting and geographic traceability of Theobroma cacao beans and cocoa products. *Food Control*, 65, 46–53. <https://doi.org/10.1016/j.foodcont.2016.01.013>

400 Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression*
 401 *trees*. Taylor & Francis.

402 Caligiani, A., Palla, L., Acquotti, D., Marseglia, A., & Palla, G. (2014). Application of ¹H NMR
 403 for the characterisation of cocoa beans of different geographical origins and fermentation
 404 levels. *Food Chemistry*, *157*, 94–99. <https://doi.org/10.1016/j.foodchem.2014.01.116>

405 Caligiani, A., Marseglia, A., Prandi, B., Palla, G., & Sforza, S. (2016). Influence of fermentation
 406 level and geographical origin on cocoa bean oligopeptide pattern. *Food Chemistry*, *211*, 431–
 407 439. <https://doi.org/10.1016/j.foodchem.2016.05.072>

408 Cambrai, A., Marcic, C., Morville, S., Sae Houer, P., Bindler, F., & Marchioni, E. (2010).
 409 Differentiation of chocolates according to the cocoa's geographical origin using
 410 chemometrics. *Journal of Agricultural and Food Chemistry*, *58*, 1478–1483.
 411 <https://doi.org/10.1021/jf903471e>

412 Capron, X., Smeyers-Verbeke, J., & Massart, D. L. (2007). Multivariate determination of the
 413 geographical origin of wines from four different countries. *Food Chemistry*, *101*, 1585–1597.
 414 <https://doi.org/10.1016/j.foodchem.2006.04.019>

415 Codex Alimentarius Commission. (2017). *Joint FAO/WHO Food standards programme and*
 416 *Codex Alimentarius Commission, 40th session*. Geneva, Switzerland.

417 Codex Alimentarius Commission. (2018). *Report of the 12th session of the codex committee on*
 418 *contaminants in foods*. Rome, Italy.

419 Corti, R., Flammer, A. J., Hollenberg, N. K., & Lüscher, T. F. (2009). Cocoa and cardiovascular
 420 health. *Circulation*, *119*, 1433–1441.

421 <https://doi.org/10.1161/CIRCULATIONAHA.108.827022>

422 D'Souza, R. N., Grimbs, S., Behrends, B., Bernaert, H., Ullrich, M. S., & Kuhnert, N. (2017).
 423 Origin-based polyphenolic fingerprinting of Theobroma cacao in unfermented and fermented
 424 beans. *Food Research International*, 99, 550–559.
 425 <https://doi.org/10.1016/j.foodres.2017.06.007>

426 Dahiya, S., Karpe, R., Hegde, A. G., & Sharma, R. M. (2005). Lead, cadmium and nickel in
 427 chocolates and candies from suburban areas of Mumbai, India. *Journal of Food Composition*
 428 *and Analysis*, 18, 517–522. <https://doi.org/10.1016/j.jfca.2004.05.002>

429 Diomande, D., Antheaume, I., Leroux, M., Lalande, J., Balayssac, S., Remaud, G. S., & Tea, I.
 430 (2015). Multi-element, multi-compound isotope profiling as a means to distinguish the
 431 geographical and varietal origin of fermented cocoa (Theobroma cacao L.) beans. *Food*
 432 *Chemistry*, 188, 576–582. <https://doi.org/10.1016/j.foodchem.2015.05.040>

433 EFSA CONTAM Panel. (2015). Scientific opinion on the risks to public health related to the
 434 presence of nickel in food and drinking water. *EFSA Journal*, 13(2).
 435 <https://doi.org/10.2903/j.efsa.2015.4002>

436 European Commission. (2006). Commission regulation (EC) No 1881/2006 of 19 December 2006
 437 setting maximum levels for certain contaminants in foodstuffs. *Official Journal of the*
 438 *European Union*.

439 European Commission. (2014). Commission Regulation (EU) No 488/2014 of 12 May 2014
 440 amending Regulation (EC) No 1881/2006 as regards maximum levels of cadmium in
 441 foodstuffs. *Official Journal of the European Union*.

442 European Commission. (2015). Commission regulation (EU) 2015/1006 of 25 June 2015
 443 amending Regulation (EC) No 1881/2006 as regards maximum levels of inorganic arsenic in
 444 foodstuffs. *Official Journal of the European Union*. [https://doi.org/http://eur-](https://doi.org/http://eur-lex.europa.eu/pri/en/oj/dat/2003/l_285/l_28520031101en00330037.pdf)
 445 [lex.europa.eu/pri/en/oj/dat/2003/l_285/l_28520031101en00330037.pdf](https://doi.org/http://eur-lex.europa.eu/pri/en/oj/dat/2003/l_285/l_28520031101en00330037.pdf)

446 European Commission. (2016). Commission recommendation (EU) 2016/1111 of 6 July 2016 on
 447 the monitoring of nickel in food. *Official Journal of the European Union*.
 448 <https://doi.org/10.2903/j.efsa.2015.4002>

449 Gonzálvez, A., Llorens, A., Cervera, M. L., Armenta, S., & de la Guardia, M. (2009). Elemental
 450 fingerprint of wines from the protected designation of origin Valencia. *Food Chemistry*, 112,
 451 26–34. <https://doi.org/10.1016/j.foodchem.2008.05.043>

452 Hooper, L., Kay, C., Abdelhamid, A., Kroon, P. A., Cohn, J. S., Rimm, E. B., & Cassidy, A.
 453 (2012). Effects of chocolate, cocoa, and flavan-3-ols on cardiovascular health : a systematic
 454 review and meta-analysis of randomized trials. *The American Journal of Clinical Nutrition*,
 455 95, 740–751. <https://doi.org/10.3945/ajcn.111.023457>.

456 ICCO. (2018). *Quarterly Bulletin of Cocoa Statistics Cocoa* (Vol. XLIV).

457 Junior, A. C. M., Maione, C., Barbosa, R. M., Gallimberti, M., Paulelli, A. C. C., Segura, F. R.,
 458 Souza, V. C. O., Batista, B. L. & Barbosa Jr, F. (2018). Elemental fingerprint profiling with
 459 multivariate data analysis to classify organic chocolate samples. *Journal of Chemometrics*,
 460 32, 1–10. <https://doi.org/10.1002/cem.3036>

461 Kumari, N., Grimbs, A., D'Souza, R. N., Verma, S. K., Corno, M., Kuhnert, N., & Ullrich, M. S.
 462 (2018). Origin and varietal based proteomic and peptidomic fingerprinting of Theobroma
 463 cacao in non-fermented and fermented cocoa beans. *Food Research International*, 111, 137–

147. <https://doi.org/10.1016/j.foodres.2018.05.010>

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied Linear Statistical Models* (5th edition). McGraw-Hill/Irwin.

Lee, K. W., Kim, Y. J., Lee, H. J., & Lee, C. Y. (2003). Cocoa has more phenolic phytochemicals and a higher antioxidant capacity than teas and red wine. *Journal of Agricultural and Food Chemistry*, 51, 7292–7295. <https://doi.org/10.1021/jf0344385>

Lo Dico, G. M., Galvano, F., Dugo, G., D’ascenzi, C., Macaluso, A., Vella, A., Giangrosso, G., Cammilleri, G. & Ferrantelli, V. (2018). Toxic metal levels in cocoa powder and chocolate by ICP-MS method after microwave-assisted digestion. *Food Chemistry*, 245, 1163–1168. <https://doi.org/10.1016/j.foodchem.2017.11.052>

Mounicou, S., Szpunar, J., Andrey, D., Blake, C., & Lobinski, R. (2003). Concentrations and bioavailability of cadmium and lead in cocoa powder and related products. *Food Additives and Contaminants*, 20(4), 343–352. <https://doi.org/10.1080/0265203031000077888>

Mrmošanin, J. M., Pavlović, A. N., Krstić, J. N., Mitić, S. S., Tošić, S. B., Stojković, M. B., Micić, J. N. & Đorđević, M. S. (2018). Multielemental quantification in dark chocolate by ICP OES. *Journal of Food Composition and Analysis*, 67, 163–171. <https://doi.org/10.1016/j.jfca.2018.01.008>

Perini, M., Bontempo, L., Ziller, L., Barbero, A., Caligiani, A., & Camin, F. (2016). Stable isotope composition of cocoa beans of different geographical origin. *Journal of Mass Spectrometry*, 51, 684–689. <https://doi.org/10.1002/jms.3833>

U.S. FDA. (2006). Guidance for industry: lead in candy likely to be consumed frequently by small

485 children. Retrieved August 15, 2018, from
486 [https://www.fda.gov/Food/GuidanceRegulation/GuidanceDocumentsRegulatoryInformation](https://www.fda.gov/Food/GuidanceRegulation/GuidanceDocumentsRegulatoryInformation/ucm077904.htm)
487 [/ucm077904.htm](https://www.fda.gov/Food/GuidanceRegulation/GuidanceDocumentsRegulatoryInformation/ucm077904.htm)

488 Villa, J. E. L., Peixoto, R. R. A., & Cadore, S. (2014). Cadmium and lead in chocolates
489 commercialized in Brazil. *Journal of Agricultural and Food Chemistry*, 62, 8759–8763.
490 <https://doi.org/10.1021/jf5026604>

491 Yanus, R. L., Sela, H., Borojovich, E. J. C., Zakon, Y., Saphier, M., Nikolski, A., Gutflais, E.,
492 Lorber, A. & Karpas, Z. (2014). Trace elements in cocoa solids and chocolate: An ICPMS
493 study. *Talanta*, 119, 1–4. <https://doi.org/10.1016/j.talanta.2013.10.048>

494

495 Figure 1: Cadmium concentrations in chocolate classified by region of origin and cacao content
496 (%). The horizontal line is the EU limit (2018) for Cd in chocolates with > 50% cacao content
497 (0.80 mg Cd kg⁻¹). Sixteen samples from South America (n=15) and Central America (n = 1)
498 exceed this limit.

499
500

501 Figure 2: A) Decision tree for the classification of chocolates by cacao origin. Classification is
502 based on fivefold cross-validated classification and regression tree analysis (CART) applied to the
503 residuals of the individual regressions of the elemental concentrations of chocolate to their cacao
504 content (parameter estimates can be found in Supplementary Information, Table S 3). Nodes are
505 indicated by numbers and leaves (end points) are indicated by lowercase letters. For each node and
506 leaf, the cut-off value of the residuals, the composition of the subgroup and the probabilities for
507 each origin are given. B) Confusion matrix for the CART model showing that 60 of the 69 South
508 American samples were correctly classified and that the overall misclassification rate was 23%.
509 The model was not able to correctly classify Asia Pacific samples. A = Africa, AP = Asia Pacific,
510 CA = Central America, SA = South America.

512 Table 1: Geographical origin and cacao content (%) of single origin chocolate samples (n = 139).
 513 The range of cacao contents of collected samples was not different between the regions.

Region	n	Cacao content ^a (%) min.-max. (median)	Countries of origin (n)
Africa	34	32–100 (72)	Cameroon (1), D.R. Congo (4), Ivory Coast (1), Madagascar (19), Tanzania ^b (5), Uganda (4)
Asia Pacific	14	42–73 (68)	India (4), Indonesia (3), Papua New Guinea (1), Samoa (1), Vanuatu (2), Vietnam (3)
Central America	22	35–100 (71)	Belize (1), Costa Rica (5), Cuba (2), Dominican Republic (5), Grenada (2), Haiti (1), Mexico (2), Nicaragua (2), Trinidad (2)
South America	69	35–100 (70)	Bolivia (1), Brazil (3), Colombia (12), Ecuador (34), Peru (16), Venezuela (3)

514 ^a From package information; ^b One sample from Tanzania (70% cacao) was considered an outlier
 515 and excluded from statistical analysis due to its extremely high Cr concentration (24 mg Cr kg⁻¹).
 516

517 Table 2: Pearson's correlation coefficients between log transformed element concentrations and
518 cacao content (%) in single origin chocolates.

	Cacao content (%)				
	All regions	Africa (n = 33)	Asia Pacific (n = 14)	Central America (n = 22)	South America (n = 69)
B	0.84*	0.87*	0.85*	0.92*	0.81*
Ba	0.68*	0.87*	0.60*	0.46*	0.65*
Ca	-0.34*	-0.30	-0.84*	-0.43*	-0.21
Cd	0.45*	0.75*	0.53	0.76*	0.31*
Co	0.54*	0.64*	0.65*	0.33	0.57*
Cr	0.11	0.04	0.06	0.20	0.13
Cs	0.24*	0.26	0.36	0.17	0.30*
Cu	0.79*	0.78*	0.91*	0.75*	0.95*
K	0.79*	0.87*	0.76*	0.72*	0.78*
Ni	0.55*	0.76*	0.56*	0.74*	0.36*
Mg	0.90*	0.92*	0.93*	0.93*	0.87*
Mn	0.77*	0.87*	0.80*	0.72*	0.74*
Mo	0.25*	0.56*	-0.30	0.18	0.25*
P	0.73*	0.81*	0.51	0.76*	0.74*
S	0.76*	0.82*	0.58*	0.86*	0.78*
Sr	0.71*	0.91*	0.53	0.48*	0.67*
V	0.12	0.06	0.22	0.50*	0.08
Zn	0.85*	0.93*	0.79*	0.84*	0.86*

519 * significant correlation ($P \leq 0.05$)

520

Table 3: Average concentrations ($\pm$ standard deviation) of 18 elements measured in chocolate bars from different origins. Superscript letters (a, b and c) indicate significant differences in the elemental concentrations between regions, levels which are not connected by the same letter are significantly different (Tukey test, $P \leq 0.05$, NS: no significant effect). Other elements were below LOQ in over 15% of the samples and are not shown.

	LOQ	Africa (n = 33)	Asia Pacific (n = 14)	Central America (n = 22)	South America (n = 69)
mg kg ⁻¹					
B	0.45	$8.0 \pm 3.0^{\text{NS}}$	7.1 ± 2.6	8.7 ± 2.4	7.9 ± 2.2
Ba	2.0	$8.5 \pm 5.2^{\text{a}}$	$3.9 \pm 2.7^{\text{b}}$	$3.7 \pm 2.1^{\text{b}}$	$4.9 \pm 3.3^{\text{b}}$
Ca	26	$1030 \pm 560^{\text{NS}}$	1130 ± 640	810 ± 400	830 ± 460
Cd	0.01	$0.20 \pm 0.11^{\text{bc}}$	$0.12 \pm 0.07^{\text{c}}$	$0.33 \pm 0.30^{\text{ab}}$	$0.58 \pm 0.52^{\text{a}}$
Co	0.01	$0.32 \pm 0.14^{\text{NS}}$	0.39 ± 0.32	0.31 ± 0.13	0.35 ± 0.18
Cr	0.04	$0.65 \pm 0.75^{\text{NS}}$	0.44 ± 0.38	0.52 ± 0.36	0.58 ± 1.1
Cs	0.001	$0.016 \pm 0.013^{\text{NS}}$	0.022 ± 0.017	0.020 ± 0.013	0.021 ± 0.031
Cu	0.07	$11.8 \pm 4.0^{\text{NS}}$	11.4 ± 4.2	13.2 ± 3.8	13.0 ± 4.2
K	8.9	$7100 \pm 1990^{\text{NS}}$	6360 ± 1170	7030 ± 1950	6330 ± 1550
Mg	1.6	$1960 \pm 650^{\text{NS}}$	1640 ± 475	1840 ± 499	1870 ± 511
Mn	0.03	$17.4 \pm 7.6^{\text{NS}}$	16.3 ± 11.9	14.1 ± 4.4	13.6 ± 5.9
Mo	0.01	$0.17 \pm 0.07^{\text{b}}$	$0.10 \pm 0.05^{\text{c}}$	$0.10 \pm 0.04^{\text{c}}$	$0.24 \pm 0.1^{\text{a}}$
Ni	0.16	$3.7 \pm 1.7^{\text{NS}}$	3.3 ± 2.2	4.6 ± 2.8	3.2 ± 1.8
P	0.76	$2920 \pm 595^{\text{NS}}$	2670 ± 296	2600 ± 435	2870 ± 654
S	58	$960 \pm 170^{\text{NS}}$	920 ± 90	930 ± 140	900 ± 190
Sr	0.04	$7.5 \pm 2.9^{\text{a}}$	$5.0 \pm 2.1^{\text{b}}$	$5.6 \pm 2.2^{\text{ab}}$	$6.3 \pm 3.5^{\text{ab}}$
V	0.007	$0.064 \pm 0.066^{\text{NS}}$	0.052 ± 0.051	0.056 ± 0.034	0.041 ± 0.036
Zn	1.2	$27 \pm 7^{\text{NS}}$	23 ± 6	26 ± 6	29 ± 8