Proteomic and peptidomic UHPLC-ESI MS/MS analysis of cocoa beans fermented using the Styrofoam-box method

Article

Accepted Version

Scollo, E., Neville, D. C. A., Oruna-Concha, M. J. ORCID: https://orcid.org/0000-0001-7916-1592, Trotin, M., Umaharan, P., Sukha, D., Kalloo, R. and Cramer, R. ORCID: https://orcid.org/0000-0002-8037-2511 (2020) Proteomic and peptidomic UHPLC-ESI MS/MS analysis of cocoa beans fermented using the Styrofoam-box method. Food Chemistry, 316. 126350. ISSN 0308-8146 doi: 10.1016/j.foodchem.2020.126350 Available at https://centaur.reading.ac.uk/88891/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

1 **Proteomic and peptidomic UHPLC-ESI MS/MS analysis of cocoa beans**

2 **fermented using the Styrofoam box method**

3

4 **Emanuele Scollo[1,2], David C. A. Neville[2], M. Jose Oruna-Concha[3], Martine**

5 **Trotin[2], Pathmanathan Umaharan[4], Darin Sukha[4], Rena Kalloo[4] and Rainer**

6 **Cramer[1*]**

7

8

9 [1] Department of Chemistry, University of Reading, Reading RG6 6AD, UK

10 [2] Mondelēz International, Reading Science Centre, Reading RG6 6LA, UK

11 [3] Department of Food and Nutritional Sciences, University of Reading, Reading RG6 6AP,

12 UK

13 [4] Cocoa Research Centre, University of West Indies, St Augustine, Trinidad and Tobago

14

15

16 *Address correspondence to:

17 Prof Rainer Cramer, Department of Chemistry, University of Reading, Whiteknights, Reading

18 RG6 6AD, UK.

19 Tel.: +44-118-378-4550; e-mail: r.k.cramer@rdg.ac.uk

20

23   **Abstract**

24   This work characterises the peptide and protein profiles of Theobroma cacao beans of the

25   genotype IMC 67 at different fermentation stages, using the Styrofoam-box fermentation

26   method and employing UHPLC-ESI MS/MS for the analysis of peptides and proteins

27   extracted from the beans. A total of 1058 endogenous peptides were identified and

28   quantified over four fermentation time points. The majority of these peptides were formed

29   after 2 and 4 days of fermentation, and originated predominantly from the proteolysis of

30   two storage proteins - vicilin and a 21 kDa albumin. The changes in the peptide profile

31   over fermentation were subsequently evaluated, and potential markers for assessing the

32   degree of fermentation were identified. In particular, changes of the relative abundance

33   of the major cocoa proteins detected can be proposed as potential markers for the

34   fermentation stage. Furthermore, PCA of both the peptidomic and proteomic data has

35   allowed differentiation of beans at different fermentation stages.

36

37   **1. Introduction**

38   Fermentation is an essential process for the development of cocoa flavour, and it is usually

39   undertaken locally, on the farms, shortly after harvesting. The cocoa pods are typically

40   opened with knives or machetes, the beans are removed and piled up into heaps or placed

41   in wooden boxes, with a total mass of up to 1.5 tonnes (Afoakwa, Paterson, Fowler, &

42   Ryan, 2008). Fermentation proceeds naturally, as the pulp surrounding the beans is

43   contaminated with microflora originating from the surrounding environment. Fermentation

44   can last 4-7 days, depending on the cocoa varieties and local practice. The cocoa mass is

45   turned every 2-3 days to allow aeration and favour the growth of aerobic microorganisms.

46   At the end of fermentation, the moisture content of the beans is reduced from an initial

47   value of around 20-30% to less than 8% by drying (Afoakwa, Paterson, Fowler, & Ryan,

48   2008). This drying process can be done naturally in the sun, with regular turning, or with

49   artificial dryers in closed rooms with temperatures not exceeding 60˚ C.

50   The pulp surrounding the beans is a rich medium for microbial growth and therefore an

51   ideal fermentation substrate. Yeasts are the main microorganisms present in the early

52   stage of fermentation, and their growth is favoured by the low pH of the pulp and low

53   oxygen level (Schwan & Wheals, 2004). Yeast cells generate ethanol and hydrolyse pectin

54   in the pulp walls with the release of juice and formation of void spaces into which air

55   percolates. Lactic acid bacteria subsequently replace the yeast cells for the next

56   fermentation stage, as they thrive in the presence of ethanol and anaerobic conditions.

57   These bacteria metabolise glucose, with the release of lactic acid, ethanol, acetic acid,

58   glycerol, mannitol, and CO_2 (Schwan & Wheals, 2004). As the oxygen level increases due

59 to disappearance of the pulp, and the temperature rises above 37º C, acetic acid bacteria
60 become the main microorganisms (Schwan & Wheals, 2004). These bacteria oxidise
61 ethanol to acetic acid, and further oxidise acetic acid to CO_2 and water. The exothermic
62 reactions of acetic acid bacteria cause the temperature of the fermenting mass to increase
63 to 50º C or more.

64 In addition, the penetration of acetic acid into the cotyledon activates endogenous
65 proteases that degrade the cocoa storage proteins (J. Voigt, Biehl, Heinrichs, Kamaruddin,
66 Marsoner, & Hugi, 1994). The released free amino acids and peptides can react with sugars
67 through Maillard reactions during roasting, forming volatile compounds that contribute to
68 the cocoa flavour. It is known that globulins are extensively degraded during fermentation,
69 while the degradation of albumins is less pronounced during this process, as these proteins
70 act as protease inhibitors (N. Kumari, Kofi, Grimbs, D'Souza, Kuhnert, Vrancken, et al.,
71 2016; Lerceteau, Rogers, Petiard, & Crouzillat, 1999; J. Voigt, Biehl, & Wazir, 1993).
72 However, there is some discrepancy as to the extent of albumin degradation during
73 fermentation reported in the literature (N. Kumari, et al., 2016; Lerceteau, Rogers,
74 Petiard, & Crouzillat, 1999; J. Voigt, Biehl, & Wazir, 1993).

75 Cocoa aroma can be produced from the roasting of peptide fractions, extracted from
76 fermented cocoa beans, in the presence of sugars and deodorised butter (J. Voigt, Biehl,
77 Heinrichs, Kamaruddin, Marsoner, & Hugi, 1994). This confirms that the peptide extracts
78 contain aroma precursors. MALDI-TOF MS and LC-ESI MS analysis of the fractions
79 producing cocoa aroma have revealed the presence of several peptides, whose amino acid
80 sequences were linked to cocoa globulins (J. Voigt, Janek, Textoris-Taube, Niewienda, &
81 Wostemeyer, 2016). It was also shown that the pH can significantly affect the type of
82 peptides released from cocoa vicilin, as proteolysis carried out at pH 5.2 produced longer
83 peptide sequences compared to those observed at pH 4.8 (Juergen Voigt, Textoris-Taube,
84 & Woestemeyer, 2018).

85 Analysis of free (endogenous) peptides by LC-ESI MS/MS of cocoa bean samples, at
86 different fermentation stages, have revealed that peptides generated from the degradation
87 of cocoa vicilin during fermentation are localised in different regions of the amino acid
88 sequence of this protein (N. Kumari, et al., 2016). The majority of the peptides observed
89 were released during early stages of fermentation, and proteolytic activity could be
90 observed up to 72 h from the start of fermentation (N. Kumari, et al., 2016).

91 Peptides formed by spontaneous fermentation and protein degradation, were extensively
92 characterised by Souza *et al.* (D'Souza, Grimbs, Grimbs, Behrends, & Corno, 2018)
93 employing LC-ESI MS/MS. Their results showed that during the early stage of fermentation
94 longer peptides were predominantly released, and subsequently degraded to shorter
95 peptides as the fermentation progressed. The identified peptides could be linked to the

96   action of both endo- and exo-peptidases, degrading mostly albumin and vicilin at both
97   protein termini. A similar methodology was used to assess differences in peptide profiles
98   between 25 samples from different geographic origins and various degrees of fermentation
99   (Neha Kumari, Grimbs, D'Souza, Verma, Corno, Kuhnert, et al., 2018). The authors stated
100 that the number of identified peptides was correlated to the fermentation stages, as poorly
101 or non-fermented beans showed a lower number of peptides compared to fully fermented
102 beans. The degree of fermentation was highly correlated to the main differences in the
103 peptide profiles of the samples analysed, while no significant differences were found when
104 taking into account the geographic origin only (Neha Kumari, et al., 2018).

105 Laboratory-based fermentation of cocoa beans, carried out in sterile glass bottles, allowed
106 the identification of 449 peptides by LC-MS, ranging from 4 to 23 amino acid residues in
107 length (Warren A., Bottcher, Asskamp, Bergounhou, Kumari, Ho, et al., 2019). Nine
108 peptides, derived from the proteolysis of cocoa vicilin and formed only during the late
109 fermentation stages, were significantly less abundant after roasting. According to the
110 authors, this loss suggests that these peptides may be responsible for the generation of
111 cocoa aroma (Warren A., et al., 2019).

112 The research presented here is based on the evaluation of the proteomic and peptidomic
113 profiles of fermented cocoa beans using the optimized small-scale Styrofoam box
114 fermentation method and UHPLC-ESI MS/MS for both bottom-up label-free proteomic
115 analysis and the analysis of endogenously fragmented peptides. The Styrofoam box
116 fermentation method was developed at the Cocoa Research Centre of the University of the
117 West Indies, and demonstrated that similar flavour profiles can be obtained when
118 compared to the traditional wooden box fermentation method. The advantage of
119 fermentation in Styrofoam boxes is the much smaller quantities of cocoa beans that are
120 required compared to the traditional wooden box fermentation. Another advantage is the
121 overall better control and reproducibility of the whole process and its environment. The
122 aim of this work was to evaluate whether markers for assessing the degree of cocoa bean
123 fermentation can be selected from beans fermented using the Styrofoam box method.

124

2. Materials and methods

2.1. Chemicals

127 Petroleum ether 40-60 and polyvinylpyrrolidone (PVPP) were purchased from Fisher
128 Scientific, Loughborough, UK. Urea, thiourea, dithiothreitol (DTT), water MS grade,
129 acetonitrile MS grade, methanol, formic acid MS grade, acetone and sodium ascorbate
130 were purchased from Sigma-Aldrich, Gillingham, UK. Trypsin was obtained from Promega,
131 Southampton, UK.

2.2. Plant materials and fermentation protocol

For all analyses, the *Theobroma cacao* genotype IMC 67 was used. Cocoa beans were harvested at the Cocoa Research Centre of the University of West Indies, St. Augustine, Trinidad. Approximately 3 hours after harvest, the pods were cut with a knife, and a total of ~3 kg of beans was placed into a Styrofoam box, which was covered with banana leaves so that fermentation could start naturally. The bean mass was turned every 2 days, and allowed to ferment for a total of 6 days. Approximately 50 g of beans were sampled at day 0, day 2, day 4, and day 6 (end of fermentation), all at the same time between 10:30 and 11:00 am local time. The temperature and pH of the beans were monitored during fermentation, and these parameters, together with a visual inspection of the beans, were used to assess the end of fermentation. Fermentation was stopped when the cocoa mass reached a pH of 3.58 and a temperature of 44° C. The aroma of the fermented mass at the end of fermentation was of a moderate acidic type.

Upon sampling, the beans were stored at -20° C with the testa removed and subsequently freeze-dried for 24 hours. The dried beans were stored at -20° C, and air-freighted in less than 96 hours to the University of Reading, UK. Upon receipt, the beans were stored at -20° C prior to analysis.

2.3. Defatting

Defatting of the beans was performed following a previously published procedure (Scollo, Neville, Oruna-Concha, Trotin, & Cramer, 2019) Fat was extracted from ground snap-frozen cocoa beans with petroleum ether (three times) by centrifugation and discarding the supernatant. For protein analysis approximately 160 mg of ground cocoa bean powder was used, while 200 mg of bean material was defatted for peptide analysis.

2.4. Removal of polyphenols

Removal of polyphenols was carried out following the protocol published by Scollo *et al.* (Scollo, Neville, Oruna-Concha, Trotin, & Cramer, 2019). Briefly, polyphenols were extracted from the defatted bean material three times using 3.5 ml of a wash solution (5 mM sodium ascorbate in 80% acetone / 20% water). A final extraction was then carried out with acetone only to remove residual water. The samples were subsequently dried under a stream of nitrogen to obtain acetone-dried powder (ACDP). This procedure was only carried out for protein extraction.

2.5. Protein extraction and Bradford assay

The ACDP was placed on a vertical shaker for 1 hour and proteins were extracted using 3.5 ml of a solubilisation solution (7 M urea, 2 M thiourea, and 20 mM DTT in water), following a previously published protocol (Scollo, Neville, Oruna-Concha, Trotin, & Cramer, 2019). The supernatant obtained after the extraction was stored at -80° C prior to analysis. The amount of total protein in each sample solution was measured with the

169 Bradford assay (Bradford, 1976), using bovine serum albumin (BSA) as reference
170 standard.

171 **2.6. Peptide extraction**
172 Approximately 25-30 mg of PVPP were added to the defatted cocoa beans, and 4 ml of a
173 0.5% trifluoroacetic acid (TFA) in water:methanol (80:20; v:v) solution were subsequently
174 added. The samples were vortexed for 1 minute and extracted for 1 hour at room
175 temperature in a vertical shaker at 700 rpm. The extracted peptide solution was
176 centrifuged at 3100 g for 10 minutes at 20° C. The supernatant was removed and stored
177 at -80° C prior to analysis.

178 **2.7. Trypsin digestion**
179 Digestions of the protein extracts were carried out following a procedure published by
180 Scollo *et al.* (Scollo, Neville, Oruna-Concha, Trotin, & Cramer, 2019). The protein extracts
181 were spiked with BSA, the pH adjusted to approximately 8.0 with the addition of
182 ammonium bicarbonate, and incubated at 37° C after the addition of dithiothreitol (DTT)
183 to a final concentration of 10 mM. The solutions were subsequently alkylated by the
184 addition of an aqueous 200-mM iodoacetamide (IAA) solution to have a final concentration
185 of IAA of 20 mM, and incubated at room temperature for 30 minutes in the dark. The
186 samples were diluted with a 2-M urea solution to obtain a final urea concentration of 0.6-
187 0.7 M. The resulting solutions were digested with trypsin at a trypsin-to-protein ratio of
188 1:50 for approximately 16 hours at 37°C. The digestions were stopped by adding TFA.

189 **2.8. Desalting of tryptic digests and endogenous peptide extracts**
190 Based on a previously published methodology SOLAµ HRP 96-well plate 2 mg sorbent mass
191 SPE cartridges (Thermo Scientific, Waltham, MA USA) were used to desalt tryptic peptides
192 (Scollo, Neville, Oruna-Concha, Trotin, & Cramer, 2019).The resulting solutions were
193 stored at -80° C prior to UHPLC-ESI MS/MS analysis. A similar procedure was followed for
194 desalting endogenous peptide solutions. In this case an aliquot of 0.4 ml of each peptide
195 extract solution was diluted to a final volume of 2 ml with 0.5% aqueous TFA. The same
196 diluent was used for equilibration and washing of the SPE cartridges prior to and after
197 sample loading. Peptides retained on the cartridges were eluted with 2x 25 µl of 0.1% TFA
198 in acetonitrile:water 75:25 (v:v). The SPE eluates were dried in a SpeedVac and
199 reconstituted in 0.1 ml of 0.1% TFA prior to UHPLC-ESI MS/MS analysis.

200 **2.9. UHPLC-ESI MS/MS analysis of tryptic digests and endogenous peptides**
201 The UHPLC-ESI MS/MS analysis of the desalted tryptic digests and endogenous peptides
202 was carried out on a Dionex Ultimate 3000 (Thermo Scientific) UHPLC system coupled to
203 an Orbitrap Q Exactive (Thermo Scientific) mass spectrometer. The chromatographic
204 separation was performed on an Acquity Peptide CSH C18, 150 mm × 0.1 mm ID, 1.7 µm
205 particle size, analytical column (Waters, Elstree, UK), using a gradient with 0.1 % (v:v)

formic acid in water as mobile phase A and 0.1 % (v:v) formic acid in acetonitrile as mobile phase B. The injection volume was 15 μl for the proteomic analysis and 10 μl for the peptidomic analysis. Further details of the chromatographic conditions and MS parameters for this analysis can be found in a previously published article (Scollo, Neville, Oruna-Concha, Trotin, & Cramer, 2019).

2.10. Data analysis

All raw MS/MS spectra files were processed using Proteome Discoverer software (Version 2.1). The obtained peak lists were loaded on the Mascot server (Version 2.6.1), and searches were carried out against the Cacao Matina 1-6 Genome v1.1 *Theobroma cacao* database (http://www.cacaogenomedb.org/Tcacao_genome_v1.1#tripal_analysis-downloads-box; downloaded on 31st May 2015; 59,577 sequences; 23,720,084 residues). The amino acid sequence of BSA was added to the *Theobroma cacao* database.

Searches were performed using parameter settings as previously published (Scollo, Neville, Oruna-Concha, Trotin, & Cramer, 2018) with trypsin selected as enzyme for proteomic analysis while no enzyme was selected for peptidomic analysis, setting the false discovery rate (FDR) to 1% for all searches. For proteomic analysis only, searches were also carried out against a custom-made contaminants database, containing mainly trypsin and keratin-related proteins (70 sequences; 31,845 residues). Label-free protein quantitation was carried out using the average of the top 3 unique peptides for each protein. Protein intensities were normalised against BSA. Only proteins with a minimum of two unique peptides were quantified. For statistical analyses, JMP Pro 13.0 (SAS, Marlow, UK) and XLSTAT 2108.5 (Addinsoft, Paris, France) software were used.

3. Results

Cocoa beans from the genotype IMC 67 were fermented in a Styrofoam box, and aliquots of samples taken at day 0, 2, 4, and 6 at the same day time. To evaluate how the peptide profiles changed during fermentation, UHPLC-ESI MS/MS analysis was performed on bean aliquots from the four different fermentation time points. Four preparative replicates were prepared for each fermentation time point, and each preparative replicate was analysed by one UHPLC-ESI MS/MS run. Only peptides detected in at least three preparative replicates of a fermentation time point were selected for comparative peptide analysis. With this requirement, a total of 1058 peptides obtained from fermentation were identified and quantified (see Supplementary Table 1) across all fermentation time points. The mean of the intensities for the preparative replicates of the same fermentation time point was calculated for each quantified peptide. The fold differences between the beans at different

241 fermentation time points are reported as the ratio of the highest mean versus the lowest
242 mean for each quantified peptide.

243 The highest number of peptides were detected in the beans sampled at 2 and 4 days of
244 fermentation, with 710 and 626 entries, respectively. A significantly lower number of
245 peptides was found in the beans at day 6 (286 entries), while 315 peptides were detected
246 in the day 0 beans. The graphical distribution of the detected peptides for all the
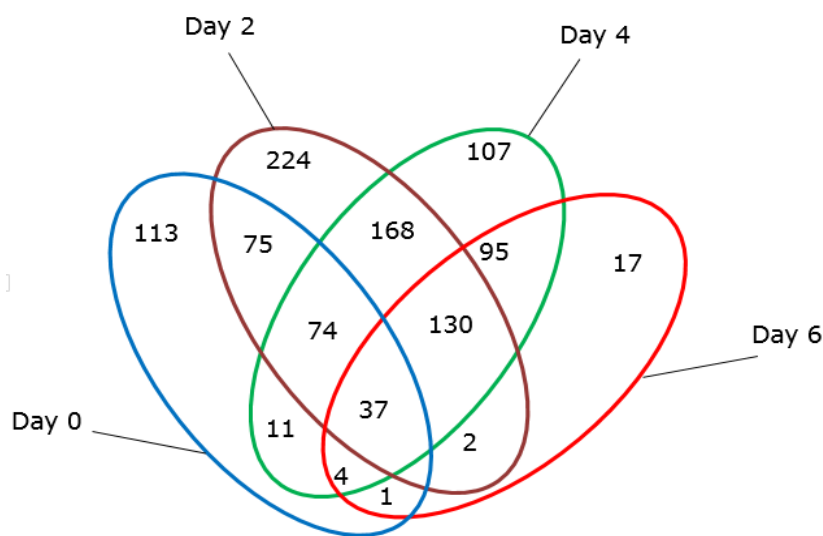247 fermentation time points is shown as a Venn diagram in Figure 1.

248
249 **Figure 1.** Venn diagram detailing the number of peptides identified for each time point.
250

251 Almost half of the detected peptides derived from vicilin (502 entries), while a total of 144
252 peptides were associated with a 21 kDa albumin. A significant number of peptides could
253 also be associated with the following proteins: lipoxygenase (60 entries), RmLC cupin (40
254 entries), glyceraldehyde-3-phosphate dehydrogenase (29 entries) and peroxygenase (28
255 entries). The complete list of proteins associated with each peptide is provided in
256 Supplementary Table 2. The list of peptides originating from vicilin and the 21 kDa albumin
257 are provided in Supplementary Tables 3 and 4, respectively.

258 The vast majority of the peptides detected in the day 2, 4 or 6 beans were either not
259 present or detected at a lower level in the day 0 beans. The number of peptides detected
260 at a higher level (compared to the other fermentation points) at each fermentation time
261 point, and the normalised intensities of the corresponding proteins measured by label-free
262 bottom-up quantitative proteomic analysis are shown in Table 1.

263

Table 1. Number of endogenous peptides detected at a higher level (compared to the other fermentation points) at each fermentation time point with the normalised intensities of the corresponding proteins.

Accession	Description	Fermentation time point				Total number of peptides
		Day 0	Day 2	Day 4	Day 6	
Thecc1EG020665t1	Vicilin-A	8	359	101	34	502
		248.9	181.6	64.8	57.6	
Thecc1EG012658t1	21 kDa seed albumin	3	43	83	15	144
		153.1	241.3	151.7	69.8	
Thecc1EG026543t1	Lipoxygenase 1	28	1	22	9	60
		6.7	11.7	18.1	12.9	
Thecc1EG041085t1	RmlC-like cupins superfamily protein	2	29	9	0	40
		5.6	3.7	0.5	0.2	
Thecc1EG017080t2	Glyceraldehyde-3-phosphate dehydrogenase C2	1	21	7	0	29
		4.1	4.6	3.7	3.0	
Thecc1EG020975t1	Peroxygenase 2	6	2	17	3	28
		1.8	2.9	1.6	0.7	
Thecc1EG004867t1	Acidic endochitinase	1	15	3	0	19
		2.9	4.5	2.2	0.6	
Thecc1EG021809t1	HSP20-like chaperones superfamily protein	5	2	12	0	19
		0.4	0.5	0.3	0.2	
Thecc1EG026589t1	Eukaryotic aspartyl protease family protein	0	6	8	1	15
		4.2	6.3	6.5	6.3	
Thecc1EG042481t1	Late embryogenesis abundant protein B19.1A	4	2	4	1	11
		0.9	1.8	0.6	0.3	
Thecc1EG016949t1	Oleosin family protein	8	0	3	0	11
		1.4	2.9	3.5	3.4	
Thecc1EG030267t1	Saposin-like aspartyl protease family protein	0	5	3	2	10
		4.1	6.5	8.5	8.5	
	Others	83	26	45	16	170
	Total number of peptides at a higher level at the given fermentation point	150	509	315	84	1058

Proteins with less than 10 peptides have been labelled as "Others". The grey rows show the number of endogenous peptides detected at a higher level at each fermentation time point, while the rows below the number of peptides show the normalised intensities of the proteins from the bottom-up label-free quantitation analysis. Only the number of endogenous peptides are reported for proteins labelled as "Others".

As vicilin and the 21 kDa albumin were the proteins contributing the vast majority of peptides, the sequence coverage and cleavage sites for these two proteins were assessed, see Supplementary Figures 1-2. A total of 246 and 108 cleavage sites were localised in the sequences of vicilin and 21 kDa albumin, respectively. The peptides, generated from vicilin, covered 68% of the sequence of this protein, while a sequence coverage of 76% was observed for the 21 kDa albumin.

To assess how peptide length and intensity vary during fermentation, the sum of the MS ion signal intensities for all peptides with the same length (number of amino acid residues) for each fermentation time point is plotted in Table 2. In addition, the number of peptides with the same length (number of amino acid residues) for each fermentation time point is shown in Table 3.

Table 2. Sums of MS ion signal intensities for peptides with the same length (7-42 amino acid residues) detected at each fermentation time point.

		Sum of peptide ion signal intensity (arb. units)			
	7	1.7E+06	3.4E+04	2.5E+05	1.5E+05
	8	3.5E+06	1.2E+07	1.9E+07	2.7E+07
	9	3.0E+06	6.8E+07	6.3E+07	4.9E+07
	10	9.0E+06	2.7E+07	2.4E+07	2.4E+07
	11	5.0E+06	1.1E+08	5.7E+07	1.9E+07
	12	1.0E+07	2.0E+08	8.3E+07	1.9E+07
	13	1.6E+07	1.8E+08	1.2E+08	4.0E+07
	14	3.0E+07	1.7E+08	7.3E+07	1.5E+07
	15	1.2E+07	9.9E+07	7.9E+07	2.8E+07
	16	1.0E+07	7.0E+07	1.1E+08	1.1E+07
	17	5.7E+07	1.9E+08	7.7E+07	2.3E+07
	18	7.4E+06	6.8E+07	4.3E+07	2.0E+07
Peptide length (number of amino acid residues)	19	6.9E+07	1.5E+08	6.6E+07	1.3E+07
	20	3.6E+06	3.9E+07	1.7E+07	5.7E+06
	21	2.1E+06	3.2E+07	2.4E+07	1.0E+07
	22	2.0E+07	1.2E+08	5.6E+07	1.2E+07
	23	1.7E+07	5.6E+07	1.5E+07	2.6E+06
	24	7.8E+06	3.7E+07	1.7E+07	4.9E+06
	25	2.6E+05	2.3E+07	5.5E+06	1.1E+06
	26	2.3E+05	3.8E+07	2.4E+07	7.5E+06
	27	8.8E+05	7.8E+06	2.9E+06	0.0E+00
	28	0.0E+00	4.9E+06	2.2E+06	0.0E+00
	29	9.6E+04	2.4E+07	3.3E+06	0.0E+00
	30	0.0E+00	7.6E+05	2.1E+06	2.4E+05
	31	0.0E+00	2.1E+06	2.2E+05	0.0E+00
	32	0.0E+00	1.5E+06	0.0E+00	0.0E+00
	33	3.7E+05	7.9E+05	0.0E+00	0.0E+00
	34	0.0E+00	0.0E+00	0.0E+00	0.0E+00
	35	5.9E+04	0.0E+00	0.0E+00	0.0E+00
	36	0.0E+00	0.0E+00	0.0E+00	0.0E+00
	37	0.0E+00	0.0E+00	0.0E+00	0.0E+00
	38	4.4E+05	4.6E+06	0.0E+00	0.0E+00
	39	2.0E+05	2.8E+05	0.0E+00	0.0E+00
	40	2.0E+05	5.6E+05	0.0E+00	0.0E+00
	41	0.0E+00	0.0E+00	0.0E+00	0.0E+00
	42	7.4E+05	0.0E+00	0.0E+00	0.0E+00
		Day 0	Day 2	Day 4	Day 6

Color scale legend:
- 2.0E+08
- 1.0E+07
- 8.0e+05
- 0

Table 3. Number of peptides with a length of 7-42 amino acid residues detected at each fermentation point.

Peptide length (number of amino acid residues)	Peptide number			
	Day 0	Day 2	Day 4	Day 6
7	1	1	2	1
8	5	7	11	10
9	10	24	30	13
10	14	32	30	17
11	18	55	64	28
12	26	70	63	25
13	30	67	72	39
14	32	63	69	38
15	23	54	53	25
16	22	53	46	19
17	15	44	34	15
18	30	50	39	14
19	18	36	24	10
20	16	31	24	8
21	13	23	13	6
22	11	20	13	5
23	11	16	9	3
24	7	17	9	3
25	3	12	6	4
26	1	7	5	2
27	2	9	5	0
28	0	3	2	0
29	1	4	1	0
30	0	2	1	1
31	0	2	1	0
32	0	2	0	0
33	1	2	0	0
34	0	0	0	0
35	1	0	0	0
36	0	0	0	0
37	0	0	0	0
38	1	2	0	0
39	1	1	0	0
40	1	1	0	0
41	0	0	0	0
42	1	0	0	0

Color scale: 72 – 45 – 20 – 0

297   The majority of the identified peptides showed an increase in intensity at fermentation
298   time point day 2 and a significant reduction at the end of fermentation, with the exception
299   of peptides with 8 residues which were more abundant at the end of fermentation. The
300   highest MS ion signal intensities were observed for peptides with either 12-14, 17, or 19
301   residues at fermentation point day 2 (see Table 2) while the highest number of peptides
302   was observed for peptides with a number of residues ranging from 11 to 15 at fermentation
303   time points day 2 and 4 (see Table 3).

304   To understand whether peptides could be used as markers for the progress of
305   fermentation, the most abundant peptides at each fermentation point (excluding day 0
306   beans) were selected. These peptides are listed in Supplementary Tables 5-7. Only
307   peptides that met the following filters were included: highest abundance amongst the four
308   fermentation time points with an RSD<20 for the peptide intensities among the
309   preparative replicates of the same fermentation time point, a fold difference of >4 for the
310   ratio between the highest and lowest mean intensity for each peptide across the different
311   time points, and absence in the day 0 beans. With these filters, 192 potential markers
312   were found for the fermentation time point at day 2, while 46 and 21 potential markers
313   were obtained for the fermentation time points at day 4 and day 6, respectively.

314   To evaluate whether the data from the peptidomic analysis can be employed to graphically
315   differentiate the four fermentation time points, PCA analysis was performed by loading the
316   intensities of all detected peptides as variables and the fermentation time points as
317   observations. In this case the data from all replicates were used. A clear separation of all
318   fermentation time points can be observed on the PCA score plot based on the first two
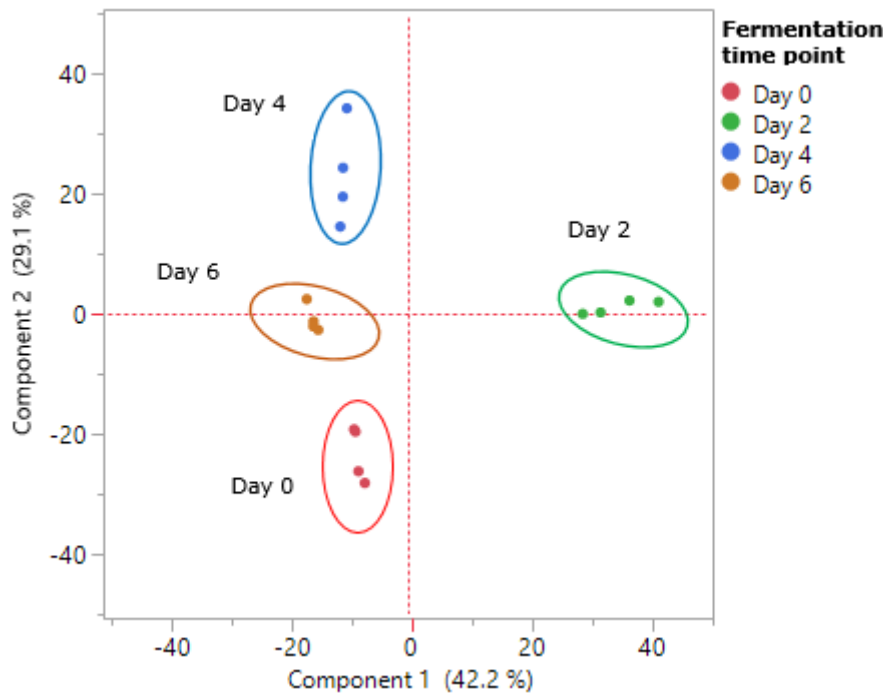319   PCA components (see Figure 2).

320

Figure 2. PCA score plot of all (preparative) replicates using the peptide data listed in Supplementary Table 1. Preparative replicates of the same fermentation time point are displayed with the same colour.

324

325 To understand the proteomic changes of cocoa beans in relation to the fermentation time,

326 label-free quantitative bottom-up proteomic analysis was carried out on the same samples

327 used to analyse the peptidome. The same number of preparative replicates were analysed,

328 and proteins which were detected and quantified in at least three preparative replicates of

329 a fermentation time point were evaluated. A total of 350 proteins were quantified when

330 combining the entries of all fermentation time points. Protein intensities were normalised

331 against BSA, which had been added as an internal standard to the samples prior to

332 digestion, and the fold difference was calculated as previously described for the peptidome

333 analysis. A list of all quantified proteins is provided in Supplementary Table 8.

334 The highest number of proteins were found in the beans of fermentation day 2 with 326

335 entries and day 0 with 325 entries. A lower number of proteins were detected in the beans

336 at fermentation day 4 (268 entries) and day 6 (245 entries). The majority of the proteins

337 (219 entries) were present at every fermentation time point. From the proteins detected

338 at all fermentation time points, a total of 91 entries showed a fold difference of >2 between

339 at least one time point pair. A considerable number of proteins (131 entries) were detected

340 in one or more but not all fermentation time points.

341 The total amount of proteins, based on the Bradford assay, showed a consistent decline

342 over fermentation, starting at 10.0% (w:w) in the samples of the beans from day 0,

13

343　dropping to 8.2% after 2 days of fermentation, and 5.2% and 4.9% after 4 and 6 days of
344　fermentation, respectively.

345　The normalised intensities of the proteins of all preparative replicates were loaded as
346　variables in a PCA score plot, while the different fermentation time points were loaded as
347　observations. This was to determine whether this classification would allow differentiating
348　the fermentation time points based on the proteomic profile. As per the peptidome
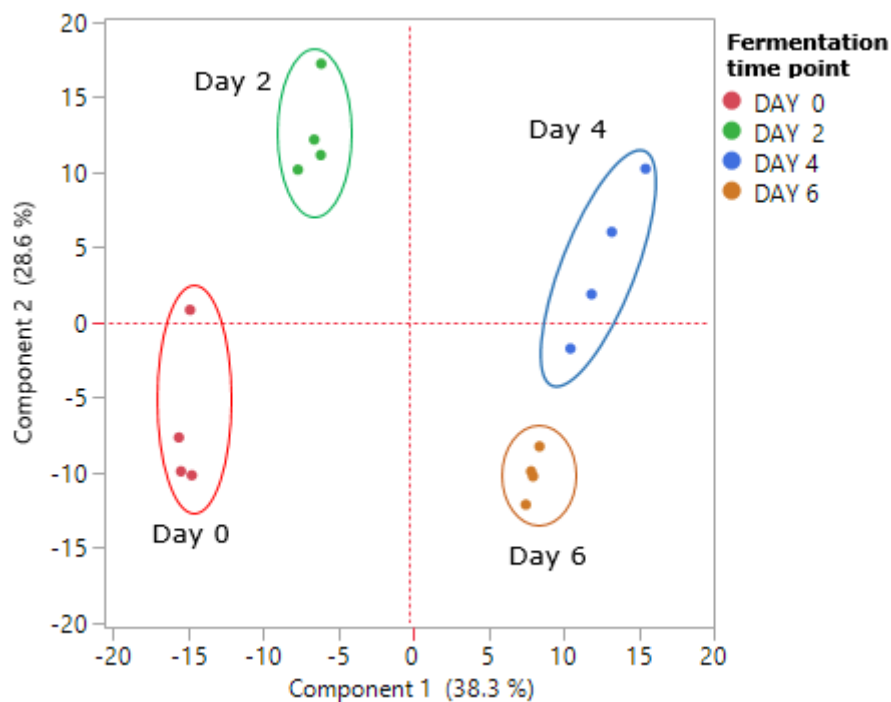349　analysis, a clear separation of all fermentation time points can be observed, see Figure 3.

350

351　**Figure 3.** PCA score plot of all preparative replicates using the protein data listed in
352　Supplementary Table 8. Preparative replicates of the same fermentation time point are
353　displayed with the same colour.

354　Data supporting the results of this work are available in the PRIDE (Proteomics
355　Identifications Database) partner repository at the European Bioinformatics Institute,
356　PXD014434 (http://www.ebi.ac.uk/pride/).

357

358　**4. Discussion**

359　The methodology employed in this study allowed the confident identification and
360　quantitation of >1000 endogenous peptides, when combining the peptidomes of cocoa
361　beans at different stages of fermentation, making it one of the largest sets of endogenous
362　peptides ever reported for cocoa beans. Similar analyses had been previously undertaken
363　(D'Souza, Grimbs, Grimbs, Behrends, & Corno, 2018; Neha Kumari, et al., 2018),
364　identifying up to 900 peptides.

14

365 The detected peptidome varied substantially with fermentation time, with the number of
366 identified peptides reaching a peak after 2 days of fermentation. A considerable drop in
367 the number of identified peptides was observed in the beans at the final stage of
368 fermentation. Only 37 peptides, were detected at all fermentation time points. The highest
369 degree of overlap was found for the fermentation time points day 2 and day 4, with 168
370 common peptides. A total of 224 peptides were solely detected in the fermentation time
371 point day 2.

372 It has been reported that the release of peptides from cocoa proteins is at its highest
373 during the early stage of fermentation, and that the proteolytic activity is considerably
374 reduced after three days (Lerceteau, Rogers, Petiard, & Crouzillat, 1999). This may explain
375 why a higher number of peptides was detected in beans collected after 2 days of
376 fermentation. However, a slightly higher number of peptides were detected in the beans
377 at fermentation day 0 compared to the beans at the latest fermentation time point
378 investigated (315 vs 286). These results are in contrast to the findings of Kumari *et al.*
379 (Neha Kumari, et al., 2018) who suggested that well-fermented beans should have more
380 than 300 peptides while beans with less than 100 peptides are to be considered under-
381 fermented. However, in this study different cocoa varieties and fermentation procedures
382 were employed. Also, cocoa pods were not immediately opened after harvest and as
383 described in section 2.2, an aliquot of cocoa beans (day 0) was stored at -20° C overnight
384 and freeze-dried for approximately 16 hours. Therefore, proteolysis by endogenous
385 proteases might have occurred during this period, which could have led to the release of
386 peptides even in beans that were not prepared for fermentation. However, endogenous
387 proteases in cocoa beans are mainly activated during fermentation due to the generation
388 of organic acids which lower the pH of the cocoa mass to levels which are optimal for these
389 enzymes (Afoakwa, Paterson, Fowler, & Ryan, 2008; Schwan & Wheals, 2004).

390 Overall, the majority of the detected peptides originated from vicilin (502 entries), followed
391 by 21 kDa albumin (144 entries), lipoxygenase (60 entries), RmLC cupin (40 entries),
392 glyceraldehyde-3-phosphate dehydrogenase (29 entries) and a peroxygenase (28
393 entries). Vicilin and albumin are the most abundant cocoa bean proteins, and therefore it
394 is reasonable to expect that their proteolytic products are well represented after
395 fermentation. More than three times as many peptides originated from vicilin compared to
396 21 kDa albumin, reflecting its similarly greater number of amino acids. A high number of
397 peptides generated from the degradation of these two proteins and a significant number
398 of peptides related to lipoxygenase have already been reported by other authors (D'Souza,
399 Grimbs, Grimbs, Behrends, & Corno, 2018).

400 The vast majority of vicilin peptides were present at a higher abundance at fermentation
401 time point day 2, while almost 60% of the peptides derived from the 21 kDa albumin were

more abundant at fermentation time point day 4, see Table 1. This indicates that these two proteins are degraded at a different rate during fermentation, confirming previously published results (Caligiani, Marseglia, Prandi, Palla, & Sforza, 2016). Only a very small proportion of peptides derived from these two proteins were found at a higher level in the fermentation day 0 beans. Interestingly, no further degradation products of these peptides were found at the later stages of fermentation. In general, for these two proteins there is no clear observable trend of (smaller) subsequent degradation products being more abundant at later time points, indicating in many cases rapid peptide ladder degradation once a new protein fragment is produced.

The detected endogenous peptides from the 21 kDa albumin cover 73% of the sequence of this protein and were localised in specific zones spread throughout the sequence except for the initial 24 residues at the N-terminal, see Supplementary Figure 1. It has been reported that the initial 26 residues at the N-terminal of this protein constitute a signal peptide which is not present in the mature protein (Spencer & Hodge, 1991). However, the results of this work show that part of this peptide can be detected. Furthermore, peptides originating from cleavage sites within the initial 26 residues have been previously reported (D'Souza, Grimbs, Grimbs, Behrends, & Corno, 2018), indicating that the signal peptide may be still present in the mature protein. No peptides were detected for regions of the sequence localised at amino acid residues 109-116, 169-179, 214-221. The data allowed localising cleavage sites for aspartyl proteases and exopeptidases at both the N- and C-terminus. A peptide with the sequence SNADSKDDVVRVSTDVNIEF at position 89-108, possibly generated from the cleavage of an internal peptide bond by aspartyl protease activity, was further degraded into 9 smaller peptides by subsequent release of the C-terminal residue, likely due to the action of a carboxypeptidase, see Supplementary Table 4. N-terminal peptide cleavages were also observed such as by the degradation of the peptide TVWRLDNYDNSAGKW (position 120-134) into 6 smaller peptides as shown in Supplementary Table 4.

A similar sequence coverage was observed for the peptides released from vicilin (see Supplementary Figure 2). No peptides localised in the N-terminal region between 1-131 of the cocoa vicilin were identified, indicating that this N-terminus was not present at the time of fermentation or was not degraded during fermentation, as reported in previous studies (N. Kumari, et al., 2016; Marseglia, Sforza, Faccini, Bencivenni, Palla, & Caligiani, 2014; J. Voigt, Janek, Textoris-Taube, Niewienda, & Wostemeyer, 2016). Since no cocoa vicilin peptide has been previously identified from this region, it has been suggested that the annotation of vicilin at the N-terminus may not be correct (N. Kumari, et al., 2016). Examples of putative carboxypeptidase activity are found by the degradation of the peptide RSEEEEGQQRNNPYYFPKRRSFQTR at position 131-155 into 11 smaller peptides by

439 subsequent amino acid release at the C-terminus, see Supplementary Table 3.
440 Fragmentation of the peptide RSEEEEGQQRNNPYYFPKR into 11 smaller peptides generated
441 by subsequent cleavages of the N-terminal residue was a strong indication of
442 aminopeptidase activity, see Supplementary Table 3.

443 One of the aims of this project was to evaluate whether peptides can be used as markers
444 to assess the degree of cocoa bean fermentation. Thus, the peptides more abundant in
445 the fermentation time points day 2, 4 and 6 are listed in Supplementary Tables 5, 6 and
446 7, respectively. However, in order to select strong marker candidates, only peptides that
447 were not present in the day 0 beans, showing an RSD of <20 and fold difference of >4,
448 were selected. The highest number of potential markers can be found for beans fermented
449 for 2 days, while only 21 entries meet these criteria for fully fermented beans. Markers
450 selected for the fermentation time point day 6 should be the most abundant peptides
451 detected in beans that have undergone full fermentation. A higher abundance of markers
452 selected for fermentation time points day 2 and day 4 could indicate that fermentation has
453 not been achieved in full.

454 As there are quite a few potential markers showing differential abundance dependent on
455 the fermentation time but no obvious peptide grouping that would display a fermentation
456 time-dependent trend, PCA was deployed to evaluate its potential for distinguishing the
457 four investigated fermentation time points (see Figure 2). Fermentation time points day
458 0, day 4 and day 6 were separated mainly by the first component, while fermentation time
459 point day 2 was differentiated mainly based on the second component. Therefore, PCA as
460 applied in this work is a useful tool to differentiate cocoa beans based on their fermentation
461 stage.

462 Label-free quantitative bottom-up proteomic analysis was also employed in this study and
463 based on the intensities of the most abundant unique tryptic digests for each protein. For
464 the two main proteins, vicilin and 21 kDa albumin, a sharp decline of their normalised
465 protein intensities obtained from this analysis was observed for the beans at the final
466 fermentation stage compared to the day 0 beans (see Table 1). In agreement with the
467 above-mentioned general lack of a trend amongst the entire set of endogenous peptides,
468 these two major proteins followed a different degradation path during fermentation.
469 Though the normalised intensity of vicilin showed a consistent decline as the fermentation
470 progressed, the intensity for the 21 kDa seed albumin reached its highest level after 2
471 days of fermentation and started to decline after 4 days of fermentation to reach its lowest
472 level in the fully fermented beans. Almost half of the endogenous peptides derived from
473 lipoxygenase were present at a higher level in the day 0 beans, while the normalised
474 intensity for this protein was higher in the later fermentation time points, suggesting that
475 proteolytic activity on this protein during fermentation was only moderate. A similar trend

476 was observed for an oleosin protein. Apart from these and two aspartyl proteases, the
477 normalised intensities of the proteins listed in Table 1 decreased over fermentation.

478 The total protein amount determined by the Bradford assay was highest in the day 0
479 beans, halved after 96 hours of fermentation and stabilized until the end of fermentation,
480 indicating that protein degradation was significantly reduced after 96 hours of
481 fermentation. The Bradford assay is based on the binding of the dye Coomassie Blue G250
482 to the arginyl and lysyl residues of proteins (Compton & Jones, 1985; Congdon, Muth, &
483 Splittgerber, 1993). Therefore, a variation in response can be observed for different
484 proteins. More importantly, the dye does not bind efficiently to peptides smaller than 3,000
485 Da, which could explain why the detected protein amount is reduced in the fermented
486 beans.

487 Vicilin was the most abundant protein in the day 0 beans, followed by a 21 kDa albumin
488 and 2S albumin. The intensities of these proteins varied over fermentation and a slightly
489 higher abundance of the 21 kDa albumin with respect to vicilin was observed for the beans
490 of the fermentation time point day 6. The ratio of the intensity of the 2S albumin versus
491 both the 21 kDa albumin and vicilin showed a steady increase with fermentation time, see
492 Supplementary Figure 3.

493 Unlike vicilin and the 21 kDa albumin, which released a high number of peptides during
494 fermentation, the 2S albumin was not degraded during this process. This fact could be one
495 of the reasons why the relative intensity of this protein increased over fermentation time.
496 A lower amount of the highly abundant vicilin and 21 kDa albumin proteins may have also
497 favoured the detection of the tryptic 2S albumin peptides, which would result in a higher
498 intensity for this protein as well. The ratios of 2S albumin versus the 21 kDa albumin and
499 vicilin could potentially be used to evaluate the degree of fermentation of cocoa beans.

500 Finally, the PCA score plot using the normalised protein intensities shows that the four
501 fermentation time points were more spread out compared to the PCA score plot created
502 by using the peptidomic data, thus potentially being more powerful for fermentation stage
503 analysis.

504

505 **5. Conclusion**

506 The data presented in this work allowed the identification and quantification of over 1000
507 endogenous peptides from cocoa beans at different fermentation stages, which makes it
508 one of the largest sets of endogenous cocoa bean peptides reported to date. The number
509 of identified cocoa bean peptides peaked at fermentation day 2 compared to day 0, 4, and
510 6. Most peptides originated from the abundant seed proteins vicilin and 21 kDa albumin.
511 This data set enabled the evaluation of the peptidomic profile changes over fermentation,

512   and allowed the selection of potential markers that could be used to assess the degree of
513   fermentation of cocoa beans. The abundance of the main cocoa proteins changed
514   significantly over fermentation but followed different trends, providing an alternative to
515   fermentation stage analysis of cocoa beans. PCA analysis of both the peptidomic and
516   proteomic data allowed a clear separation of the beans at different fermentation stages.
517   Thus, this methodology could be employed as a tool for creating larger databases of
518   fermentation-relevant proteins and peptides and for an easy differentiation of beans at
519   different fermentation stages.

520

521   **6. Acknowledgements**

523 **7. References**

524 Afoakwa, E. O., Paterson, A., Fowler, M., & Ryan, A. (2008). Flavor formation and
525   character in cocoa and chocolate: A critical review. *Critical Reviews in Food Science*
526   *and Nutrition, 48*(9), 840-857.

527 Bradford, M. M. (1976). A Rapid and Sensitive Method for the Quantitation of Microgram
528   Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Analytical*
529   *Biochemistry, 72*(1-2), 248-254.

530 Caligiani, A., Marseglia, A., Prandi, B., Palla, G., & Sforza, S. (2016). Influence of
531   fermentation level and geographical origin on cocoa bean oligopeptide pattern. *Food*
532   *Chemistry, 211*, 431-439.

533 Compton, S. J., & Jones, C. G. (1985). MECHANISM OF DYE RESPONSE AND
534   INTERFERENCE IN THE BRADFORD PROTEIN ASSAY. *Analytical Biochemistry, 151*(2),
535   369-374.

536 Congdon, R. W., Muth, G. W., & Splittgerber, A. G. (1993). THE BINDING INTERACTION
537   OF COOMASSIE BLUE WITH PROTEINS. *Analytical Biochemistry, 213*(2), 407-413.

538 D'Souza, R. N., Grimbs, A., Grimbs, S., Behrends, B., & Corno, M. (2018). Degradation of
539   cocoa proteins into oligopeptides during spontaneous fermentation of cocoa beans.
540   *Food Research International, 109*, 516-516.

541 Kumari, N., Grimbs, A., D'Souza, R. N., Verma, S. K., Corno, M., Kuhnert, N., & Ullrich,
542   M. S. (2018). Origin and varietal based proteomic and peptidomic fingerprinting of
543   Theobroma cacao in non-fermented and fermented cocoa beans. *Food research*
544   *international (Ottawa, Ont.), 111*, 137-147.

545 Kumari, N., Kofi, K. J., Grimbs, S., D'Souza, R. N., Kuhnert, N., Vrancken, G., & Ullrich,
546   M. S. (2016). Biochemical fate of vicilin storage protein during fermentation and
547   drying of cocoa beans. *Food Research International, 90*, 53-65.

548 Lerceteau, E., Rogers, J., Petiard, V., & Crouzillat, D. (1999). Evolution of cacao bean
549   proteins during fermentation: a study by two-dimensional electrophoresis. *Journal of*
550   *the Science of Food and Agriculture, 79*(4), 619-625.

551 Marseglia, A., Sforza, S., Faccini, A., Bencivenni, M., Palla, G., & Caligiani, A. (2014).
552   Extraction, identification and semi-quantification of oligopeptides in cocoa beans. *Food*
553   *Research International, 63*, 382-389.

554 Schwan, R. F., & Wheals, A. E. (2004). The microbiology of cocoa fermentation and its
555   role in chocolate quality. *Critical Reviews in Food Science and Nutrition, 44*(4), 205-
556   221.

557 Scollo, E., Neville, D., Oruna-Concha, M. J., Trotin, M., & Cramer, R. (2018).
558   Characterization of the Proteome of Theobroma cacao Beans by Nano-UHPLC-ESI
559   MS/MS. *Proteomics, 18*(3-4).

560   Scollo, E., Neville, D. C. A., Oruna-Concha, M. J., Trotin, M., & Cramer, R. (2019). UHPLC-
561   MS/MS analysis of cocoa bean proteomes from four different genotypes. *Food*
562   *Chemistry, 303*, 8.

563   Spencer, M. E., & Hodge, R. (1991). Cloning and sequencing of the cDNA-encoding the
564   major albumin of *Theobroma cacao*-Identification of the protein as a member of the
565   Kunitz protease inhibitor family. *Planta, 183*(4), 528-535.

566   Voigt, J., Biehl, B., Heinrichs, H., Kamaruddin, S., Marsoner, G. G., & Hugi, A. (1994). In-
567   vitro formation of cocoa-specific aroma precursors: aroma-related peptides generated
568   from cocoa-seed proteins by co-operation of an aspartic endoprotease and a
569   carboxypeptidase. *Food Chemistry, 49*(2), 173-180.

570   Voigt, J., Biehl, B., & Wazir, S. K. S. (1993). The major seed proteins of *Theobroma cacao*
571   *L. Food Chemistry, 47*(2), 145-151.

572   Voigt, J., Janek, K., Textoris-Taube, K., Niewienda, A., & Wostemeyer, J. (2016). Partial
573   purification and characterisation of the peptide precursors of the cocoa-specific aroma
574   components. *Food Chemistry, 192*, 706-713.

575   Voigt, J., Textoris-Taube, K., & Woestemeyer, J. (2018). pH-Dependency of the proteolytic
576   formation of cocoa- and nutty-specific aroma precursors. *Food Chemistry, 255*, 209-
577   215.

578   Warren A., J., Bottcher, N. L., Asskamp, M., Bergounhou, A., Kumari, N., Ho, P.-W.,
579   D'Souza, R. N., Nevoigt, E., & Ullrich, M. S. (2019). Forcing fermentation: Profiling
580   proteins, peptides and polyphenols in lab-scale cocoa bean fermentation. *Food*
581   *Chemistry, 278*, 786-794.

582