Food Control 114 (2020) 107219

Contents lists available at ScienceDirect

Food Control

journal homepage: www.elsevier.com/locate/foodcont

Genetic authentication: Differentiation of fine and bulk cocoa (*Theobroma cacao* L.) by a new CRISPR/Cas9-based *in vitro* method

Alexandra Scharf[1], Christina Lang[1], Markus Fischer[*]

Hamburg School of Food Science, Institute of Food Chemistry, University of Hamburg, Hamburg, Germany

ARTICLE INFO

Keywords:
In vitro CRISPR/Cas9
Theobroma cacao L.
Variety differentiation
Food fraud
Arriba
CCN-51

ABSTRACT

Ecuador is one of the most important producers of fine cocoa. In addition to the fine cocoa type Nacional (Arriba), however, the cheaper but less aromatic bulk cocoa "Colección Castro Naranjal 51" (CCN-51) is also cultivated there. In a previous work, several characteristic SNPs could be identified to distinguish these two varieties. However, the detection of single base exchanges in order to differentiate raw material types or varieties is challenging, as some of the methods require the use of restriction endonucleases. As a prerequisite, the application is only successful if an appropriate recognition site is available. The use of the programmable Cas9 endonuclease known from the genome editing system CRISPR/Cas9 is a powerful alternative. For the detection of bulk cocoa in fine cocoa a single nucleotide polymorphism (SNP) was selected, which is located within a PAM region mandatory for the Cas9 endonuclease. Consequently, only the bulk cocoa is attacked by the nuclease. The result can be recorded using agarose or capillary gel electrophoresis (AGE and CGE). Both methods yielded comparable results. AGE can be used for a semi-quantitative estimation and the more sophisticated CGE for quantitation based on a calibration line. A reliable detection could be made up to an admixture of 10% CCN-51 which should represent a realistic scale for routine applications.

1. Introduction

The cocoa tree *Theobroma cacao* L. belongs to the genus Theobroma and the mallow family (Malvaceae) (Janick & Paull, 2008). It is a diploid tree fruit species with its origin in the rain forests of South America (Davie, 1935; Henderson, Joyce, Hall, Hurst, & McGovern, 2007). World production in 2017 amounted to 5.2 million tonnes (FAO). The leading cocoa producer is Ivory Coast with almost 40% of the world annual production (FAO). About 5–10% of the cultivated cocoa is fine cocoa (Argout et al., 2011; Rusconi & Conti, 2010). According to the International Cocoa Agreement 2010, fine or flavor cocoa is defined as "cocoa recognized for its unique flavor and colour and produced in countries designated in annex C of this agreement" (International Cocoa Organization [ICCO], 2010). One of the most important cultivation area of fine cocoa is Ecuador (ICCO, 2010, 2014). It is producing the fine cocoa type Nacional (Arriba), which is characterized by a distinctive intensive flavor (Afoakwa, Paterson, Fowler, & Ryan, 2008; Solorzano et al., 2013). Since the Ecuador's breeding program in the 1960s the clone "Colección Castro Naranjal 51" (CCN-51) has also been cultivated in Ecuador. It's a precocious and disease-resistant cultivar and is known for its high yields (Boza et al., 2014). In contrast to Arriba, however, it has a significantly weaker aroma with high levels of bitterness and astringency (Afoakwa et al., 2008; Boza et al., 2014; Kadow, Bohlmann, Phillips, & Lieberei, 2013). Due to the differences in quality and price, it is of great importance for the cocoa processing industry to determine adulterations of fine and consumer cocoa. Such blends can take place either unintentionally during harvesting and processing or intentionally in order to increase the profit (Herrmann, Haase, Blauhut, Barz, & Fischer, 2014). The adulteration of fine cocoa with consumer cocoa is a classic example for the need of methods to determine the authenticity of food. The genome is a suitable marker as it is stable against exogenous and anthropogenic influencing factors such as climate, soil conditions or fertilization (Fischer et al., 2014). There are already various methods available for distinguishing between fine and bulk cocoa. The analyses of microsatellite (simple sequence repeat, SSR) markers or the randomly amplified polymorphic DNA (RAPD) analyses are examples for non-targeted methods (Frances, Michelle, & Albertus, 2000; Herrmann et al., 2015; Lerceteau, Robert, Pétiard, & Crouzillat, 1997; N'Goran, Laurent, Risterucci, & Lanaud, 1994). The advantage of non-targeted methods is that a precise sequence information is not required. But especially the analysis of SSR markers is an expensive and elaborate method, which is rarely used in

[*] Corresponding author. Hamburg School of Food Science, Institute of Food ChemistryUniversity of Hamburg Grindelallee 117, 20146 Hamburg, Germany.
E-mail address: Markus.Fischer@chemie.uni-hamburg.de (M. Fischer).
[1] Equally contributing first authors.

https://doi.org/10.1016/j.foodcont.2020.107219
Received 20 January 2020; Received in revised form 3 March 2020; Accepted 4 March 2020
Available online 08 March 2020
0956-7135/

routine analysis. Furthermore, the identification and quantitation of mixtures is extremely complex and requires a high bioinformatic effort. In addition to these approaches, targeted methods can also be used to differentiate between various quality levels. Hermann et al. developed diverse methods for the differentiation of Arriba and CCN-51 on the basis of sequence differences within the chloroplast genome (Herrmann et al., 2014). On the one hand, a restriction fragment length polymorphism (RFLP) analysis based on a detected single nucleotide polymorphism (SNP) was established and on the other hand, the detection of a repeat sequence that is repetitive with varying frequency was established using capillary gel electrophoresis (CGE) and denaturing high-performance liquid chromatography (dHPLC) (Herrmann et al., 2014). One of the most cost-effective and easy to use method is RFLP analysis (Mata, Chanmalee, Punyasuk, & Thitamadee, 2020). A disadvantage of this method is that only SNPs within the recognition site of a specific endonuclease can be detected.

The *clustered regularly interspaced short palindromic repeats*–(CRISPR)-associated system (CRISPR/Cas system) was identified in the years 2005–2008 as adaptive immune system in archaea and bacteria. Only half a year later, it could be shown that the type II CRISPR/Cas9 system from *Streptococcus pyogenes* is suitable as a "tool" for the specific modification of the genome of different organisms (Jinek et al., 2012; Sternberg & Doudna, 2015). Three components are usually required for an application of the CRISPR/Cas9 system: The target specific crRNA, which is supported by the transactivating crRNA (tracrRNA) and the endonuclease Cas9, which is guided (programmed) by the crRNA/tracrRNA complex (guide RNA, gRNA) to the target sequence (protospacer sequence). Additionally, for binding and double-stranded DNA cutting a protospacer adjacent motif (PAM) sequence (in the case of the CRISPR/Cas9 system from *S. pyogenes* the triplet NGG, N = any nucleotide) next to the protospacer sequence is required (Marraffini & Sontheimer, 2010). To date, the CRISPR/Cas9 technology has been successfully used in human cell lines (Jinek et al., 2013), plants (Feng et al., 2013), mammals (Li et al., 2013), fish (Hwang et al., 2013), insects (Friedland et al., 2013), fungi (Waltz, 2016), yeasts (DiCarlo et al., 2013) and bacteria (Jiang, Bikard, Cox, Zhang, & Marraffini, 2013). Besides its application in living organisms, CRISPR/Cas9 has already been used as an *in vitro* screening method to predict possible off-target mutations in the human genome (Kim et al., 2015; Tsai et al., 2017).

The idea of the present study is the development of an alternative, Cas9-based *in vitro* method using the Cas9 endonuclease as programmable restriction endonuclease to broaden the spectrum of RFLP analyses for the detection of SNPs. This approach should be used to differentiate the cocoa varieties Arriba and CCN-51. We show that Cas9 nuclease can be used as a flexible and guidable endonuclease with its associated RNA sequences (crRNA/tracrRNA) as an analytical tool. In the study described below we will therefore not only speak of the Cas9 endonuclease but rather of the CRISPR/Cas9 system. With this approach, it is possible to use previously inaccessible SNPs for the development of a simple targeted method for routine application.

2. Materials and Methods

2.1. General methods and materials

All primers were purchased from IDT (Leuven, Belgium). Nucleotide sequences of PCR fragments were determined using an external sequencing service (GATC Biotech, Constance, Germany). Due to the high sensitivity of RNA to RNases and their ubiquitous occurrence, certain measures have been taken for work with RNA. For example, pipetting was carried out under a separate workbench, which was cleaned with RNase-ExitusPlus™ (AppliChem GmbH, Darmstadt, Germany) and irradiated with UV light for 30 min before and after use. All consumables were autoclaved before use and glass equipment was also heated. The ultrapure water used was additionally treated with diethyl pyrocarbonate (AppliChem GmbH, Darmstadt, Germany), which also inhibits enzymes like RNases. The required crRNA und tracrRNA were commercially acquired (see 2.7).

2.2. Plant material and sample preparation

For the study, cocoa beans of the varieties Arriba and CCN-51 of the harvest years 2015 and 2016 from Ecuador were used. A part of the Arriba cocoa beans was roasted for 30 min at 145 °C in an electric coffee roaster (VEVOR, Hamburg, Germany). A commercial chocolate (Naturata AG, Marbach, Germany) with a fine cocoa content of 70% of the variety Arriba (Ecuador) was purchased from a local supermarket in Schleswig-Holstein, Germany for further analysis. First, the cocoa beans were peeled. The prepared beans and the chocolate were grounded with liquid nitrogen to a fine powder and then sieved to obtain fine homogeneous products. All samples were stored at −20 °C.

2.3. DNA extraction

The cocoa DNA was isolated from the pestled and homogenized samples using the peqGOLD Plant DNA Mini Kit (Peqlab, VWR, Darmstadt, Germany) according to the manufacturer's instructions for fresh and frozen samples. The DNA concentration was quantified using the Quantus™ Fluorometer (Promega GmbH, Mannheim, Germany). The ratio of absorbance at 260 nm and 280 nm DNA was used to access the purity of the extracted DNA (UV spectrophotometer, DeNovix Inc., Wilmington, USA). All isolates were stored at −20 °C.

2.4. Selection of a specific Cas9 target site for in vitro cleavage of CCN-51 cocoa

Sequences of the chloroplast genomes of Arriba and CCN-51, which were generated within the IGF Project AiF 16796 N of the FEI supported via AiF, served as a basis of the analysis (Herrmann et al., 2014). These were examined for SNPs between the two varieties mentioned above. A suitable SNP had to comply the following requirements: (I) SNP within a PAM region, (II) SNP leads to a non-canonical PAM sequence in the Arriba variety, whereby the bulk cocoa is fragmented by the Cas9. Based on these requirements, two regions (A and B) both located on the large single copy (LSC) region of the chloroplast genomes, have been selected. The selected sequences used for this study are shown in Fig. S1 and Fig. S3.

2.5. PCR amplification for generation of the templates from Arriba and CCN-51 cocoa

DNA amplification for generation of the templates from Arriba and CCN-51 cocoa (Fig. S1 and Fig. S3) was carried out in a total volume of 100 µL, containing 1 x Q5 reaction buffer (contains 2 mM Mg^{2+}, NEB, Frankfurt am Main, Germany), 1 µM of each primer (Table 1), 75 µM of each deoxynucleotide triphosphate (Bioline GmbH, Luckenwalde, Germany), 2 ng DNA, 2 U Q5® High-Fidelity DNA Polymerase (NEB, Frankfurt am Main, Germany), DNase-/RNase-free water (treated with diethyl pyrocarbonate). PCR was performed under the following conditions: initial denaturation step at 98 °C for 30 s, 25 cycles with denaturation at 98 °C for 7 s, primer annealing for 20 s, elongation at 72 °C for 30 s, followed by a final elongation at 72 °C for 4 min. After the PCR has been completed, the samples were cooled to 4 °C until further use. The PCR products were purified using the Monarch® PCR & DNA Cleanup Kit (NEB, Frankfurt am Main, Germany) according to the manufacturer's instructions. The DNA concentration was measured using the Quantus™ Fluorometer (Promega GmbH, Mannheim, Germany). If the concentration was less than 25 ng/µL, the PCR was repeated using the above conditions. For quality control, all sequences were checked by Sanger sequencing.

Table 1
Primers used for the generation of the templates from Arriba and CCN-51 cocoa in this study. The primers Template-B.2_FW and Template-B.2_RV were used to generate the shorter DNA fragment from region B for the analysis of processed samples (roasted cocoa beans and chocolate).

Region	Primer	Primer sequence (5' → 3')	Annealing temperature [°C]
A	Template-A_FW	GTTCAGTCGGTACATAGTAATCTATC	60.0
	Template-A_RV	GAGAAGAGGTATTACTGGCATA	
B	Template-B.1_FW	CAATACTCGTAGAGAGAGAACCG	64.3
	Template-B.1_RV	AATGGATAGAGCTCTACGGCC	
	Template-B.2_FW	CTTTCCCTTGTGAGTAAGATACTAATC	62.2
	Template-B.2_RV	GACTGTATCGCACTATGTATCATTTG	

2.6. Overlap extension PCR for the generation of cocoa templates with altered PAM regions

In order to generate cocoa templates with non-canonical PAM regions (see Table S2), overlap extension PCR was performed using the flanking primers Template-A_FW and Template-A_RV (see Table 1) and the corresponding inner primers (overlap primers, see Table S1). Regarding the nucleotide sequence, the templates were copies of the respective region A of CCN-51 cocoa with a small deletion, to remove the canonical PAM-sequence (AGG) and an insertion immediately downstream of the CRISPR/Cas9 target sequence (protospacer region) consisting of one of the five non-canonical PAM regions and 12 further synthetic base pairs (see overlap sequence in Table S1). Both PCRs were performed as described under 2.5 with mentioned primers and the specific annealing temperatures. In the second PCR, 50 ng of each of the two parts were used. In addition, four cycles were added to the temperature program at the beginning, in which the annealing temperature was oriented to the overlapping sequence in order to promote hybridization of the two fragments and thus generate the long, modified fragments.

2.7. In vitro Cas9 digestion

For the analysis of the region A and the templates with non-canonical PAM regions the crRNA and tracrRNA from Dharmacon (Horizon Discovery Company, Cambridge, UK) and for the region B the CRISPR Cas9 crRNA and CRISPR Cas9 tracrRNA from IDT (Leuven, Belgium) were used (see Table 2). The tracrRNA from Dharmacon (74 nt) and the tracrRNA from IDT (67 nt) are shorter than the natural tracrRNA from *S. pyogenes* (89 nt). All the RNAs were specially modified for increased stability against RNases, for example by methylation of the ribose or adding phosphorothioate linkages in the phosphate backbone. The sequence of the tracrRNA from IDT as well as the modifications are not further specified by the manufacturers. Furthermore, the S.p. Cas9 Nuclease V3 (IDT, Leuven, Belgium) was used for all experiments. The experiments with the modified templates with non-canonical PAM regions were also performed with the S.p. HiFi Cas9 Nuclease V3 (IDT, Leuven, Belgium).

The digestion was carried out in a total volume of 15 µL. In a first step, crRNA and tracrRNA were combined, incubated at 95 °C for 5 min and cooled down to room temperature for 10 min. This allows the RNAs to hybridize and form the duplex (gRNA). A total of 50 ng gRNA was used. The proportions of crRNA and tracrRNA vary depending on the length. The next step was the formation of the ribonucleoprotein (RNP) complex. 250 ng Cas9 nuclease was added and the solution was incubated at 37 °C for 10 min. Then 50–150 ng DNA template (qualification: 50 ng, quantitation: 150 ng), 1.5 µL 10x Cas9 reaction buffer and 3 µg BSA (both NEB, Frankfurt am Main, Germany) were added and filled up to 15 µL volume with DNase-/RNase-free water (treated with diethyl pyrocarbonate). The reaction mixture was incubated for 1 h at 37 °C, then heated to 80 °C for 5 min and finally cooled to 4 °C until the AGE was performed using the whole reaction volume (15 µL). For the analysis with CGE, the batches were purified using the Monarch® PCR & DNA Cleanup Kit (NEB, Frankfurt am Main, Germany). The elution was carried out with 6 µL DNase-/RNase-free water, whereby the eluate was applied to the column a second time to increase the yield.

2.8. Agarose gel electrophoresis (AGE)

Gels with a concentration of 1.5% agarose and TAE buffer (40 mM Tris acetate, 2 mM EDTA, pH 8) were used. Prior to electrophoresis, 2 µL Loading Dye (40 mM Tris acetate, 2 mM EDTA, 50% glycerol, 0.005% xylene cyanol, pH 8) was added to the samples. The separation was carried out with equipment from VWR (Darmstadt, Germany) and Carl Roth GmbH & Co. KG (Karlsruhe, Germany) at 120 V. Detection under UV light (Dark Hoods DH-40/50, Biostep® GmbH, Jahnsdorf, Germany) was performed after staining the DNA with Gel Red™ (diluted 1:3333 in water). A DNA ladder was used to determine the fragment length.

2.9. Quantitation using capillary gel electrophoresis (CGE)

For the quantitation based on a calibration line, seven mixtures of fine and bulk cocoa with the following proportions of CCN-51 were prepared: 2.5%, 5%, 10%, 20%, 40%, 60% and 80%. For this purpose, the ground and sieved samples were weight in corresponding proportions and the DNA was isolated as described under 2.3. Afterwards the template was generated by PCR (see 2.5) and the enzymatic digestion as described under 2.7 was performed with an amount of 150 ng DNA. After purification, fragments were recorded using a 2100 Bioanalyzer in combination with the Agilent DNA 7500 DNA Kit (Agilent Technologies, Inc., Santa Clara, USA). The measurement was carried out according to the manufacturer's instructions. The interpretation of the electropherograms was carried out with the software 2100 Expert (Agilent Technologies, Inc., Santa Clara, USA).

Table 2
Sequences and lengths of the crRNA and tracrRNA used in this study. The template specific parts of the crRNAs are underlined.

Region	Name	Sequence (5' → 3')	Length [nt]
A	crRNA	AUUUGACUCUUAGAAAGAGUUUUAGAGCUAUGCU GUUUUG	42
	tracrRNA	AACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAU CAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU	74
B	crRNA	CCCAAAAAAAUGGAUUUUGGUGUUUUAGAGCUAUGCU	36
	tracrRNA	(no specifications by the manufacturer)	67

3. Results and discussion

3.1. Selection of specific CRISPR/Cas9 target sites for in vitro cleavage of CCN-51 cocoa

The development of this method was based on our previously reported findings that the whole chloroplast genomes of Arriba and CCN-51 cocoa differ from each other by about 30–50 SNPs (Herrmann et al., 2014). These SNPs can be used to distinguish between the two cocoa varieties. The prerequisite for performing RFLP is the existence of a suitable recognition site of a restriction endonuclease. However, since not all SNPs create or destroy a recognition site, many SNPs cannot be detected with this method. In contrast, the use of an adjustable endonuclease would avoid this problem. Cas9, known from the Genome Editing System CRISPR/Cas9, should in principle be suitable for this. The targeting specificity of Cas9 is believed to be strictly controlled by the 20 nt guide sequence (crRNA) and the presence of a PAM region (5′-NGG-3′) immediately adjacent to the protospacer sequence (target sequence) in the genome. Nevertheless, several studies have shown that off-target effects can even occur in DNA sequences with three to five base pair mismatches to the PAM distal region of the crRNA (nucleotides 11–20) (Cho et al., 2014; Hsu et al., 2013; Mali et al., 2013; Xu, Duan, & Chen, 2017). The PAM proximal region (nucleotides 1–10), also called “seed region”, is generally considered to be more important for the Cas9 specificity than the rest of the crRNA sequence (see Fig. 1) (Jinek et al., 2012; Semenova et al., 2011; Wiedenheft et al., 2011). However, off-target effects are also possible here, especially if there is only a single base mismatch (Jiang et al., 2013). In this context, it was important to select a SNP that is directly located in the PAM region and can thus reliably avoid an off-target activity of the Cas9 nuclease. Since method development in the present case should focus on the quantitative detection of CCN-51 cocoa, a SNP which converts a canonical PAM region (NGG) in CCN- 51 to a non-canonical PAM region in Arriba cocoa (NHG or NGH, H = A, C or T) had to be selected. This should ensure that only PCR products of CCN-51 cocoa are recognized and cleaved by the Cas9 nuclease, while PCR fragments of Arriba cocoa remain uncleaved due to the lack of a canonical PAM region. A suitable target sequence was found that has the non-canonical PAM region (5′-AGA-3′) in Arriba cocoa and turns into the recognition sequence (5′-AGG-3′) due to a SNP in CCN-51 cocoa (see Supplementary, Fig. S1).

3.2. In vitro Cas9 digestion

PCR products with a size of 1031 bp (region A) were generated from DNA isolates of Ecuadorian cocoa beans of the varieties Arriba and CCN-51 and rechecked for the presence or absence of the corresponding SNP by Sanger sequencing (see 2.1–2.3 and 2.5 Materials and Methods). Using a target specific crRNA and tracrRNA from Dharmacon (Horizon Discovery Company, Cambridge, UK) *in vitro* digestion with the PCR products of Arriba and CCN-51 cocoa using Cas9 nuclease was performed (see Fig. 2, for detailed information see 2.7 Materials and Methods).

It is known, that the Cas9 nuclease cuts DNA 3 bp upstream of the PAM. The DNA template of CCN-51 (1031 bp) was accordingly digested into two smaller fragments of 626 bp and 405 bp (see Fig. 3A). Contrary to the assumption that the Cas9 nuclease only induces a double strand break if the correct PAM region is present next to the target sequence, the DNA template of Arriba cocoa was also completely cleaved into the two corresponding fragments. In addition to the canonical PAM region, several studies have already shown that the CRISPR Cas9 system from *S. pyogenes* can also use NAG as an alternative PAM region (Jiang et al., 2013). Also, for the non-canonical PAM (NGA) used in this study, an *in vivo* study with human cells found cleavage activity of the Cas9 nuclease (Zhang et al., 2014). On the other hand, in both cases the binding affinities were substantially lower compared to the canonical PAM region. However, this is not confirmed by our *in vitro* CRISPR/Cas9 application, in which the DNA templates of Arriba cocoa with the non-canonical PAM region NGA were cleaved at least as efficiently as the DNA templates of CCN-51 with the canonical PAM region. We assume that this is due to the higher number of copies of the target sequence in the *in vitro* approach and/or the corresponding high concentration of the Cas9 nuclease used compared to *in vivo* applications. However, only a slight improvement of this off-target effect could be achieved by reducing the Cas9 nuclease by about 30%. Again, for the determination of CCN-51 cocoa it is of vital importance that only PCR products of CCN-51 cocoa are recognized and cleaved by the Cas9 nuclease, while PCR fragments of Arriba cocoa remain intact. Therefore, the DNA template of CCN-51 cocoa was first modified to test further variants of non-canonical PAM regions with a single base exchange for Cas9 cleavage activity.

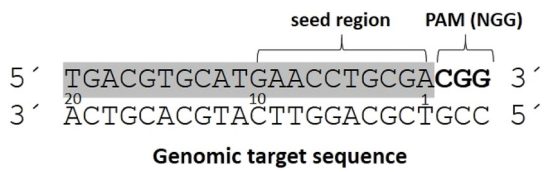

Fig. 1. Schematic figure of a genomic target sequence. Seed region and PAM region are indicated. Target sequence (protospacer sequence) is shadowed grey.

3.3. Cleavage analysis of cocoa templates with altered PAM regions

To create the DNA templates of CCN-51 cocoa with non-canonical PAM regions, overlap extension PCR with appropriate primers was performed (see 2.6 Materials and Methods). The extension of overlapping gene segments by PCR is a versatile technique for creating base substitutions, insertions and/or deletions in a DNA molecule. Regarding the nucleotide sequence, the templates were copies of the respective chloroplast DNA region A of CCN-51 cocoa with a non-canonical PAM region (NGT, NGC, NTG or NCG) and 12 further synthetic base pairs, which were used for overlapping. It has already been shown that non-canonical PAM regions which contain the purine base adenine instead of the purine base guanine (NGA or NAG) can be used as an alternative PAM region by the Cas9 nuclease. Therefore, non-canonical PAM regions with a base exchange to a pyrimidine base (T or C) were selected for further cleavage analysis. For control purposes, an additional DNA template was prepared which does not contain any PAM-specific base (TCT) immediately adjacent to the target sequence. With the DNA templates of CCN-51 cocoa modified in this way, Cas9 *in vitro* digestions were again carried out (see Fig. 3B). In addition, all digestions were performed with the HiFi Cas9 nuclease (IDT Technologies), which is expected to have similar targeted potency to wild type Cas9, but significantly reduced off-target activity. The DNA template that had no PAM-specific bases (TCT) was not cleaved by the Cas9 wild type nuclease. Furthermore, the DNA template with the non-canonical PAM region AGT remained completely intact. DNA templates, which had a non-canonical PAM region resulting from an exchange with the base cytosine (ACG and AGC), were partially cleaved, albeit to a lower extent than the template with the non-canonical PAM ATG. Using the HiFi Cas9 nuclease almost identical results were achieved. Similar to the wild type Cas9 nuclease the DNA templates with the non-canonical PAM regions ACG, AGC and ATG were partially cleaved (see Supplementary Fig. S2). Based on these results, a SNP should now be selected which converts the canonical PAM region in CCN-51 into the non-canonical PAM region NGT in Arriba cocoa and thus may successfully prevent the cleavage of the DNA template of Arriba cocoa by the Cas9 nuclease.

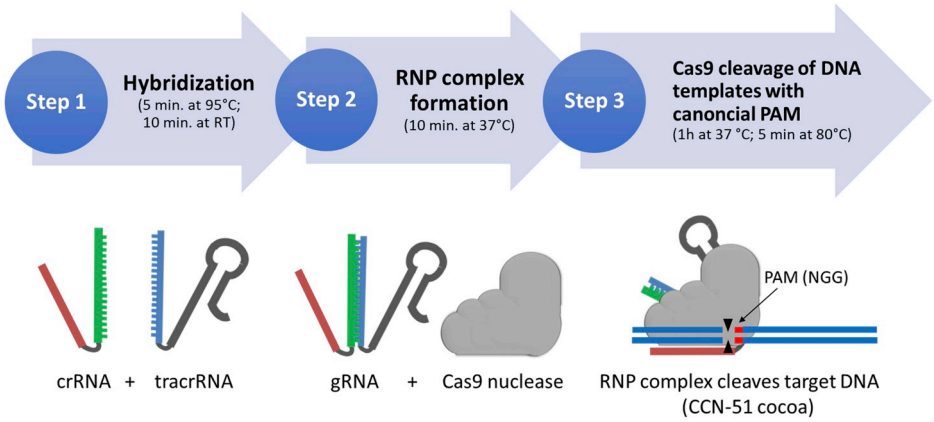

Fig. 2. Scheme of an *in vitro* Cas9 digestion. The procedure is carried out in three steps and takes only 1.5 h.

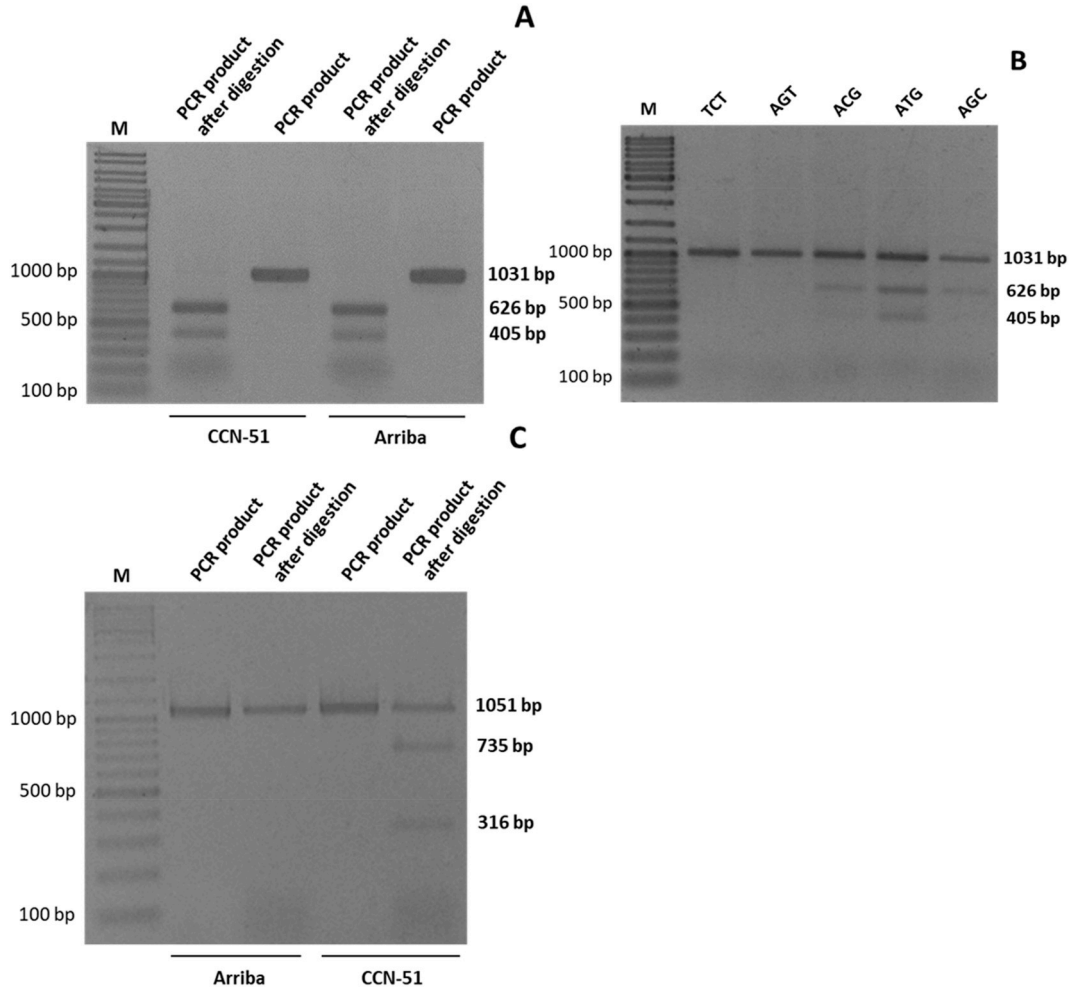

Fig. 3. Different *in vitro* Cas9 digestions. The agarose gel (A) shows the PCR products from the purified genomic DNA and the respective digestions of CCN-51 and Arriba cocoa of region A. Both PCR products were cleaved into two DNA fragments of 626 bp and 405 bp using the *in vitro* CRISPR/Cas9 system. Cleavage analyses of cocoa templates with non-canonical PAM regions using the wild type Cas9 nuclease is shown in Figure B. The agarose gel shows the Cas9 digestions of the DNA templates from CCN-51 cocoa (region A) with altered PAM regions (TCT, AGT, ACG, ATG, AGC). The DNA template that had no PAM-specific bases (TCT) and the template with the non-canonical PAM region AGT remained intact, while the DNA templates with the other non-canonical PAM regions (ACG, ATG and AGC) were partially cleaved.

The agarose gel (C) shows the PCR products from the purified genomic DNA and the respective Cas9 digestions of CCN-51 and Arriba cocoa of region B. A part of the DNA template of CCN-51 cocoa was cleaved into the two DNA fragments of 735 bp and 316 bp. The DNA template of Arriba cocoa with the non-canonical PAM region TGT remained completely intact. DNA marker (M).

3.4. In vitro CRISPR/Cas9 approach with the non-canonical PAM region NGT in Arriba cocoa

From three potentially suitable SNPs, one was selected that has the non-canonical PAM region TGT in Arriba cocoa and, due to a SNP in CCN-51 cocoa the recognition sequence TGG (region B, for sequences see Supplementary, Fig. S3). PCR products with a size of 1051 bp were generated from DNA isolates of Ecuadorian cocoa beans of the varieties Arriba and CCN-51. The sequence range of the DNA templates was selected in such a way that two fragments with a length of 735 bp and 316 bp were obtained when the Cas9 nuclease cleave the DNA templates 3 bp upstream of the PAM region. After the correct sequences of the DNA templates were confirmed by Sanger sequencing, *in vitro* Cas9 digestion was performed under the same conditions as described before. While the DNA template of CCN-51 cocoa with the *in vitro* CRISPR/Cas9 system were cleaved into the two DNA fragments of 735 bp and 316 bp, the DNA templates of Arriba cocoa with the non-canonical PAM region TGT remained completely intact (see Fig. 3C). However, in contrast to the first experiment, only a part of the DNA templates of CCN-51 cocoa was cleaved. A clear DNA band at 1051 bp was still visible. The efficiency of cleavage by Cas9 nuclease is depending on the respective target sequence, which has also been observed by other authors in different *in vivo* as well as *in vitro* studies (Doench et al., 2014; Mehravar, Shirazi, Mehrazar, & Nazari, 2019). In the next step, it should be checked in some optimization approaches whether the DNA template of CCN-51 cocoa can be completely cleaved into its two fragments. Therefore, different concentrations of the main components (Cas9 nuclease and crRNA as well as tracrRNA) and extended incubation times (step 3 of the process) were tested.

3.5. Optimization of the Cas9 in vitro digestion

Initially, it was investigated whether the Cas9 nuclease and the crRNA/tracrRNA complex concentrations could be optimized. For a digestion (15 µL) routinely 250 ng Cas9 nuclease and 1.34 pmol each of tracrRNA and crRNA were used (see 2.5 Materials and Methods). In addition to this regular composition, digestions were prepared, each with a 20% reduced or increased concentration of Cas9 nuclease and/or crRNA as well as tracrRNA (see Supplementary, Fig. S4). Despite the concentrations of crRNA and tracrRNA should be equimolar, a deviating ratio was also tested. With the digestions modified in this way, neither a decrease in the amount of uncleaved DNA template nor an increase could be detected, which indicates that the tested components are already present in an optimal concentration.

Another parameter that must be optimized is the incubation time. It is possible that the Cas9, comparable to other endonucleases, may require more time to completely cut the template. The incubation time was 1 h for each of the previous reactions. In the case of the CCN-51 sample, the incubation time was extended to 2 h, 4 h, 8 h, 10 h, 12 h, 18 h, 24 h, 36 h and 48 h. In order to estimate the efficiency of extending the incubation time, the intensity of the bands on the agarose gel, which can be assigned to the uncut template, were compared. In the course of 1–24 h, a steady improvement was observed. A significant decrease in the intensity of the uppermost band (uncut template) compared to 1 h incubation only becomes apparent after 8 h (see Supplementary, Figure S5 A). From 24 to 48 h no further improvement was visible. However, complete fragmentation of the template of CCN-51 was not possible (see Supplementary, Figure S5 B). For further analysis it was important that the Arriba cocoa was not fragmented by the Cas9 even if the incubation time is extended. Therefore, Arriba cocoa was incubated with the nuclease for up to 24 h. No fragmentation of the template could be detected (see Supplementary, Figure S5 C). This confirmed the previous results that the chosen template with the SNP (T/G) on the third position of the PAM region is stable against the Cas9 over a longer time period. Since significant improvements of the band intensity occurs after 8 h, for the following analysis the time saving variant with only 1 h incubation was chosen.

For a quantitation, however, the prerequisite is that the proportion on uncut template is always the same. Therefore, the efficiency of the Cas9 nuclease was examined. Three identical batches with 100% CCN-51 DNA were cut using Cas9 and subsequently analyzed with CGE. For comparison, the percentage of uncut fragment in the total area was determined (see Supplementary, Fig. S6). The total area is composed of the areas of the uncut template and the two fragments. The proportion of uncut template is 42.8 ± 2.9%. Thus, a constant reproducibility of the enzymatic digestion can be assumed, and a quantitation can be developed.

3.6. Quantitation of mixtures of fine and bulk cocoa

The next step was the development of a method for quantifying mixtures of fine and bulk cocoa using capillary gel electrophoresis for separation of the DNA fragments. Since there is a linear relationship between the CCN-51 content and the areas of the fragments, a calibration line can be created using linear regression. The quantitation was done using the 2100 Bioanalyzer (Agilent Technologies Inc., USA) and calibration with standard mixtures of known concentration. For this purpose, seven mixtures of fine and bulk cocoa with proportions of 2.5–80% CCN-51 were prepared and analyzed. The mixtures were used for both qualitative and quantitative approaches using AGE and CGE respectively. For qualification, the whole reaction volume was used for the AGE. A gel image of the mixtures treated with Cas9 is shown in Fig. 4A. For quantitation, the electropherograms were evaluated using 2100 Expert Software (Agilent Technologies Inc., USA). Calibration was carried out by plotting the proportion of bulk cocoa against the ratio of the area of a fragment to the total area (sum of the areas of the uncut template and the two fragments). This step of normalization is necessary to compensate concentration differences after purification. The

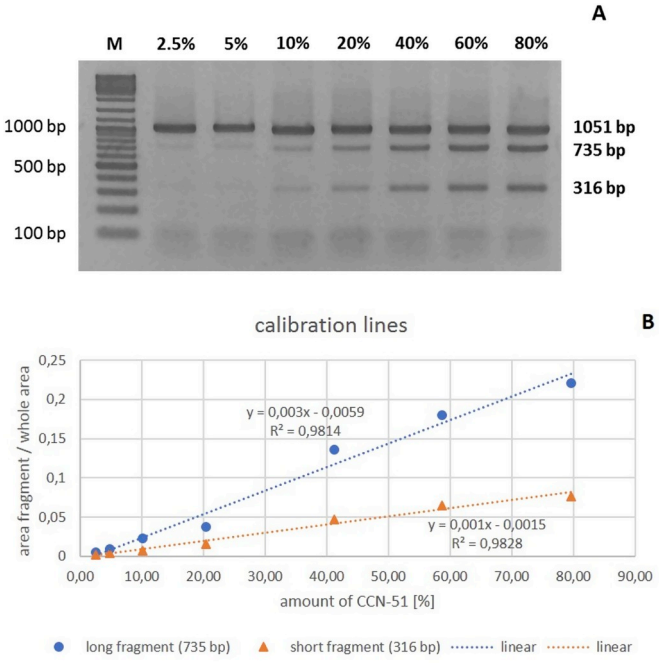

Fig. 4. Analysis of different mixtures of fine and bulk cocoa using the Cas9 nuclease. The agarose gel (A) shows the results for seven mixtures with 2.5–80% CCN-51 after digestion with Cas9 nuclease and a template amount of 150 ng. Both fragments (lengths: 735 bp, 316 bp) are visible up to an admixture of 10% CCN-51. DNA marker (M). (B) shows the calibration lines after analysis of those mixtures with CGE resulted by plotting the proportion of bulk cocoa against the ratio of the area of a fragment to the total area (sum of the areas of the uncut template and the two fragments). The calibration lines for both fragments are shown.

two resulting calibration lines obtained by linear regression using the long and short fragments are shown in Fig. 4B. Both have a coefficient of determination of $R^2 = 0.98$, confirming an adequate linear relationship between the share of bulk cocoa and the resulting areas. These plots show additionally that purification and measurement did not discriminate against any of the fragments. For further analysis, the long fragment (735 bp) was used, as it has higher fluorescence due to its length which allows a more sensitive detection in the range of small concentrations. Since first measurements showed that 50 ng template is too low as a starting material for the measurement with CGE and in order to improve the detection limit, the amount of DNA used was increased from 50 ng to 150 ng per reaction. Both the electropherograms and the agarose gel show that both fragments (fragment I: 735 bp, fragment II: 316 bp) can be detected in addition to the uncut template (1051 bp) at a proportion of 10% CCN-51. As already mentioned, the band/area of fragment I is more intensive/larger due to its length. Both methods can be used to detect a 10% CCN-51 content mixed with fine cocoa. At 2.5% and 5%, however, only fragment I is weakly recognizable next to the template. Altogether it can be said that AGE and CGE provide comparable results and for semi-quantitative estimation the AGE can be used, whereas for a more exact quantitation the CGE with corresponding calibration line is necessary.

3.7. Application to processed products

A decisive step in the development of methods for authenticity control is the applicability to processed commodities. A major issue in the DNA analysis of processed products is the fragmentation of the DNA in the course of processing. In summary, various steps during production are the cause of DNA degradation: (i) First, fermentation carried out immediately after harvesting to remove the pulp and form essential aroma precursors. (ii) In the next step the beans are dried before being transported to the processing industry. Already these steps can lead to an incipient degradation of the DNA due to heat generation. (iii) An even more relevant step leading also to DNA fragmentation is the roasting process of the cocoa beans. Due to the high temperatures of > 100 °C, the DNA can be degraded (Di Bernardo, Galderisi, Cipollaro, & Cascino, 2005; Mano et al., 2017; Vijayakumar, Martin, Gowda, & Prakash, 2009). (iv) Later, the steps of rolling, fine rolling and finally conching in chocolate production must also be taken into account, which additionally causes strong mechanical exposure on the sample and can also contribute to fragmentation. The results presented so far are based on fermented and dried but unroasted cocoa beans. Therefore, the next essential step in the development was to transfer the analysis method to roasted cocoa beans and chocolate. The DNA isolates showed a different and sometimes very high degree of fragmentation (average length < 500 bp) (see Supplementary, Fig. S7). PCR was used to test whether the 1051 bp template could still be amplified. Despite optimization of the PCR parameters, amplification was not possible due to the identified fragmentation. For the analysis of processed cocoa products therefore a new pair of primers (Template-B.2_FW/RV, for primer sequences see Table 1) was designed covering a region shortened to a length of 430 bp (see Supplementary, Fig. S3). This fragment could be used for all processed products in this study (see Supplementary, Fig. S8). A major difference is the mandatory addition of BSA for amplification of processed samples compared to unprocessed cocoa beans. The improvement of amplification by addition of BSA has been described in the literature (Abu Al-Soud & Rådström, 2000; Forbes & Hicks, 1996; Kreader, 1996). This effect is caused by the binding affinity of BSA and can be seen, for example, in PCR-inhibiting substances from the group of polyphenols, which are also contained in cocoa (Tengel, Schüssler, Setzke, Balles, & Sprenger-Haussels, 2001). These react with or bind to proteins and thus have an inhibitory effect on amplification (Loomis, 1974). Thus, BSA also shows an enzyme-stabilizing effect, which can also be transferred to the Taq DNA polymerase (Jordan, Zugay, Darke, & Kuo, 1992). From this, it can be concluded that more PCR-inhibiting substances are present in the isolates which cannot be removed during isolation. Another aspect that must be considered for processing products is the Maillard reaction. This is a non-enzymatic reaction in which carbonyl groups of reducing sugars react with primary amines. Reactive α-dicarbonyl compounds are formed, which can crosslink with macromolecules such as proteins or nucleic acids (Papoulis, al-Abed, & Bucala, 1995; Vasan et al., 1996). These compounds can inhibit both DNA isolation and amplification (Di Bernardo et al., 2005; Poinar et al., 1998). Finally, the Cas9 digestion with the short templates was successfully carried out. No further optimizations were necessary. All reactions were performed with 50 ng DNA for qualitative detection as well as 150 ng for quantitation. Both batches were successfully cut using Cas9 nuclease (see Fig. 5).

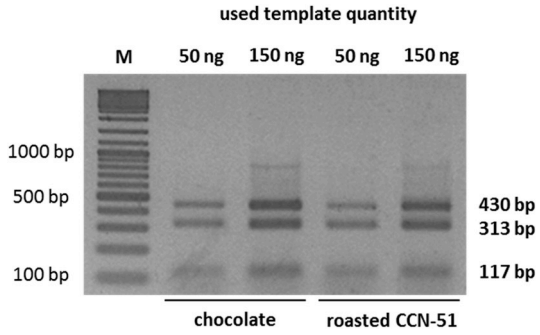

Fig. 5. Analysis of the digestion of the short template of processed products using the Cas9 nuclease. The agarose gel shows the results for chocolate and roasted CCN-51 beans with two different template concentrations each. At both template concentrations used for qualitative and quantitative detection, both fragments can detect despite processing. DNA marker (M).

4. Conclusion

In this study, we successfully developed a fast and easy CRISPR/Cas9 based *in vitro* method for the quantitative determination of the bulk cocoa CCN-51 in fine flavored cocoa Arriba. We were able to show for the first time that the programmable Cas9 endonuclease could be an efficient tool for detection of SNPs which are not accessible by Type II endonucleases as part of "classical" RFLP analysis. By using the Cas9 programmable endonuclease, it was possible to detect an admixture of at least 10% CCN-51 cocoa in fine cocoa type Nacional (Arriba). The sensitivity of the method is therefore adequate to reliably detect an intentional adulteration with low-priced consumer cocoa. Furthermore, it could be shown, that this method, with some modifications, is also suitable for the differentiation of roasted cocoa beans and derived products such as chocolate. For semi-quantitative analysis, also a conventional AGE can be used instead of CGE.

In order to adapt the method presented here to other targets, three simple steps would be usually necessary: (i) Selection of a SNP that forms a recognition site (5′-NGG-3′) in variety A and alters it in variety B by a base exchange at position three into a thymine base (5′-NGT-3′). (ii) The target specific crRNA is defined by the 20 nucleotides immediately upstream of the PAM region and can be ordered from a supplier together with the Cas9 specific tracrRNA. (iii) Generation of PCR templates of both varieties. The length of the template and the position of the SNP should be chosen in such a way that two fragments of different size can be obtained after Cas9 digestion. After method implementation and preparation of a calibration series, a fast performance is possible.

The method presented here, which is based on the CRISPR/Cas9 system from *S. pyogenes*, has some limitations resulting from the restricted selection of a SNP within a defined non-canonical PAM region. However, due to the constant increasing number of newly discovered CRISPR systems in bacteria and archaea (for example CRISPR/Cpf1

with the PAM region (5′-TTTV-3′) as well as synthetically generated modifications of already known CRISPR nucleases, a large number of different recognition sequences becomes available. Hence, previously challenging SNPs, such as those found in AT-rich regions of the plastid genome, become accessible by the analytical use of the CRISPR/Cas systems.

CRediT authorship contribution statement

Alexandra Scharf: Conceptualization, Investigation, Methodology, Formal analysis, Writing - original draft, Visualization. **Christina Lang:** Conceptualization, Investigation, Methodology, Formal analysis, Writing - original draft, Visualization. **Markus Fischer:** Conceptualization, Resources, Writing - review & editing, Supervision.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgements

We would like to thank Dr. Boris Illarionov and Christine Felbinger, who supported us in the preliminary work and in the course of the project. We would also like to thank Nicolas Cain for sample material.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodcont.2020.107219.

References

Abu Al-Soud, W., & Rådström, P. (2000). Effects of amplification facilitators on diagnostic PCR in the presence of blood, feces, and meat. *Journal of Clinical Microbiology, 38*(12), 4463–4470.

Afoakwa, E. O., Paterson, A., Fowler, M., & Ryan, A. (2008). Flavor formation and character in cocoa and chocolate: A critical review. *Critical Reviews in Food Science and Nutrition, 48*(9), 840–857. https://doi.org/10.1080/10408390701719272.

Argout, X., Salse, J., Aury, J.-M., Guiltinan, M. J., Droc, G., Gouzy, J., et al. (2011). The genome of Theobroma cacao. *Nature Genetics, 43*(2), 101–108. https://doi.org/10.1038/ng.736.

Boza, E. J., Motamayor, J. C., Amores, F. M., Cedeño-Amador, S., Tondo, C. L., Livingstone, D. S., et al. (2014). Genetic characterization of the cacao cultivar CCN 51: Its impact and significance on global cacao improvement and production. *Journal of the American Society for Horticultural Science, 139*(2), 219–229. https://doi.org/10.21273/JASHS.139.2.219.

Cho, S. W., Kim, S., Kim, Y., Kweon, J., Kim, H. S., Bae, S., et al. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Research, 24*(1), 132–141. https://doi.org/10.1101/gr.162339.113.

Davie, J. H. (1935). Chromosome studies in the malvaceae and certain related families. *II. Genetica, 17*(5–6), 487–498. https://doi.org/10.1007/BF01508190.

Di Bernardo, G., Galderisi, U., Cipollaro, M., & Cascino, A. (2005). Methods to improve the yield and quality of DNA from dried and processed figs. *Biotechnology Progress, 21*(2), 546–549. https://doi.org/10.1021/bp049710p.

DiCarlo, J. E., Norville, J. E., Mali, P., Rios, X., Aach, J., & Church, G. M. (2013). Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic Acids Research, 41*(7), 4336–4343. https://doi.org/10.1093/nar/gkt135.

Doench, J. G., Hartenian, E., Graham, D. B., Tothova, Z., Hegde, M., Smith, I., et al. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nature Biotechnology, 32*(12), 1262–1267. https://doi.org/10.1038/nbt.3026.

Feng, Z., Zhang, B., Ding, W., Liu, X., Yang, D.-L., Wei, P., et al. (2013). Efficient genome editing in plants using a CRISPR/Cas system. *Cell Research, 23*(10), 1229–1232. https://doi.org/10.1038/cr.2013.114.

Fischer, M., Creydt, M., Felbinger, C., Fischer, C., Klockmann, S., Werner, P., et al. (2014). Food profiling - Strategien zur Überprüfung der Authentizität von Rohstoffen. *Journal of Consumer Protection and Food Safety,* (9), 400–404.

Forbes, B. A., & Hicks, K. E. (1996). Substances interfering with direct detection of Mycobacterium tuberculosis in clinical specimens by PCR: Effects of bovine serum albumin. *Journal of Clinical Microbiology, 34*(9), 2125–2128.

Frances, B., Michelle, E., & Albertus, E. (2000). *Proceedings of the international workshop on new Technologies and cocoa breeding: 16th-17th october 2000* (Kota Kinabalu, Sabah, Malaysia. London).

Friedland, A. E., Tzur, Y. B., Esvelt, K. M., Colaiácovo, M. P., Church, G. M., & Calarco, J. A. (2013). Heritable genome editing in C. Elegans via a CRISPR-Cas9 system. *Nature Methods, 10*(8), 741–743. https://doi.org/10.1038/nmeth.2532.

Henderson, J. S., Joyce, R. A., Hall, G. R., Hurst, W. J., & McGovern, P. E. (2007). Chemical and archaeological evidence for the earliest cacao beverages. *Proceedings of the national academy of sciences of the United States of America: 104*, (pp. 18937–18940). . https://doi.org/10.1073/pnas.0708815104 48.

Herrmann, L., Felbinger, C., Haase, I., Rudolph, B., Biermann, B., & Fischer, M. (2015). Food fingerprinting: Characterization of the ecuadorean type CCN-51 of theobroma cacao L. Using microsatellite markers. *Journal of Agricultural and Food Chemistry, 63*(18), 4539–4544. https://doi.org/10.1021/acs.jafc.5b01462.

Herrmann, L., Haase, I., Blauhut, M., Barz, N., & Fischer, M. (2014). DNA-based differentiation of the Ecuadorian cocoa types CCN-51 and Arriba based on sequence differences in the chloroplast genome. *Journal of Agricultural and Food Chemistry, 62*(50), 12118–12127. https://doi.org/10.1021/jf504258w.

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology, 31*(9), 827–832. https://doi.org/10.1038/nbt.2647.

Hwang, W. Y., Fu, Y., Reyon, D., Maeder, M. L., Tsai, S. Q., Sander, J. D., et al. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature Biotechnology, 31*(3), 227–229. https://doi.org/10.1038/nbt.2501.

International Cocoa Organization (2010). *International cocoa agreement* (London, U.K).

International Cocoa Organization (2014). *ICCO annual report 2014/2015*(London, U.K).

Janick, J., & Paull, R. E. (Eds.). (2008). *The encyclopedia of fruit & and nuts*Wallingford: CABI. Retrieved from http://www.loc.gov/catdir/enhancements/fy0724/2006027763-d.html.

Jiang, W., Bikard, D., Cox, D., Zhang, F., & Marraffini, L. A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology, 31*(3), 233–239. https://doi.org/10.1038/nbt.2508.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science, 337*(6096), 816–821. https://doi.org/10.1126/science.1225829.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., & Doudna, J. (2013). RNA-programmed genome editing in human cells. *ELife, 2*, e00471. https://doi.org/10.7554/eLife.00471.

Jordan, S. P., Zugay, J., Darke, P. L., & Kuo, L. C. (1992). Activity and dimerization of human immunodeficiency virus protease as a function of solvent composition and enzyme concentration. *Journal of Biological Chemistry, 267*(28), 20028–20032.

Kadow, D., Bohlmann, J., Phillips, W., & Lieberei, R. (2013). Identification of main fine flavour components in two genotypes of the cocoa tree (Theobroma cacao L.). *Journal of Applied Botany and Food Quality, 86*(1).

Kim, D., Bae, S., Park, J., Kim, E., Kim, S., Yu, H. R., et al. (2015). Digenome-seq: Genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nature Methods, 12*(3), 237–243. https://doi.org/10.1038/nmeth.3284.

Kreader, C. A. (1996). Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32 protein. *Applied and Environmental Microbiology, 62*(3), 1102–1106.

Lerceteau, E., Robert, T., Pétiard, V., & Crouzillat, D. (1997). Evaluation of the extent of genetic variability among Theobroma cacao accessions using RAPD and RFLP markers. *Theoretical and Applied Genetics, 95*(1–2), 10–19. https://doi.org/10.1007/s001220050527.

Li, D., Qiu, Z., Shao, Y., Chen, Y., Guan, Y., Liu, M., et al. (2013). Heritable gene targeting in the mouse and rat using a CRISPR-Cas system. *Nature Biotechnology, 31*(8), 681–683. https://doi.org/10.1038/nbt.2661.

Loomis, W. D. (1974). [54] Overcoming problems of phenolics and quinones in the isolation of plant enzymes and organelles. In S. Fleischer (Vol. Ed.), *Methods in enzymology: 31*, (pp. 528–544). New York, NY: Acad. Press. https://doi.org/10.1016/0076-6879(74)31057-9 Biomembranes (Vol. 31.

Mali, P., Aach, J., Stranges, P. B., Esvelt, K. M., Moosburner, M., Kosuri, S., et al. (2013). Cas9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology, 31*(9), 833–838. https://doi.org/10.1038/nbt.2675.

Marraffini, L. A., & Sontheimer, E. J. (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nature Reviews Genetics, 11*(3), 181–190. https://doi.org/10.1038/nrg2749.

Mata, W., Chanmalee, T., Punyasuk, N., & Thitamadee, S. (2020). Simple PCR-RFLP detection method for genus- and species-authentication of four types of tuna used in canned tuna industry. *Food Control, 108*, 106842. https://doi.org/10.1016/j.foodcont.2019.106842.

Mehravar, M., Shirazi, A., Mehrazar, M. M., & Nazari, M. (2019). In vitro pre-validation of gene editing by CRISPR/Cas9 ribonucleoprotein. *Avicenna Journal of Medical Biotechnology, 11*(3), 259–263.

N'Goran, J. A. K., Laurent, V., Risterucci, A. M., & Lanaud, C. (1994). Comparative genetic diversity studies of Theobroma cacao L using RFLP and RAPD markers. *Heredity, 73*(6), 589–597. https://doi.org/10.1038/hdy.1994.166.

Papoulis, A., al-Abed, Y., & Bucala, R. (1995). Identification of N2-(1-carboxyethyl)guanine (CEG) as a guanine advanced glycosylation end product. *Biochemistry, 34*(2), 648–655. https://doi.org/10.1021/bi00002a032.

Poinar, H. N., Hofreiter, M., Spaulding, W. G., Martin, P. S., Stankiewicz, B. A., Bland, H., et al. (1998). Molecular coproscopy: Dung and diet of the extinct ground sloth Nothrotheriops shastensis. *Science, 281*(5375), 402–406. https://doi.org/10.1126/science.281.5375.402.

Rusconi, M., & Conti, A. (2010). Theobroma cacao L., the food of the gods: A scientific approach beyond myths and claims. *Pharmacological Research, 61*(1), 5–13. https://doi.org/10.1016/j.phrs.2009.08.008.

Semenova, E., Jore, M. M., Datsenko, K. A., Semenova, A., Westra, E. R., Wanner, B., et al. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proceedings of the national academy of sciences of the United States of America: 108*, (pp. 10098–10103). . https://doi.org/10.

1073/pnas.1104144108 25.
Solorzano, R. G. L., Fouet, O., Lemainque, A., Pavek, S., Boccara, M., Argout, X., et al. (2013). Correction: Insight into the wild origin, migration and domestication history of the fine flavour nacional theobroma cacao L. Variety from Ecuador. *PloS One, 8*(2).
Sternberg, S. H., & Doudna, J. A. (2015). Expanding the biologist's toolkit with CRISPR-cas9. *Molecular Cell, 58*(4), 568–574. https://doi.org/10.1016/j.molcel.2015.02.032.
Tengel, C., Schüssler, P., Setzke, E., Balles, J., & Sprenger-Haussels, M. (2001). PCR-based detection of genetically modified soybean and maize in raw and highly processed foodstuffs. *Biotechniques, 31*(2), 426–429. https://doi.org/10.2144/01312pf01.
Tsai, S. Q., Nguyen, N. T., Malagon-Lopez, J., Topkar, V. V., Aryee, M. J., & Joung, J. K. (2017). Circle-seq: A highly sensitive in vitro screen for genome-wide CRISPR-cas9 nuclease off-targets. *Nature Methods, 14*(6), 607–614. https://doi.org/10.1038/nmeth.4278.
Vasan, S., Zhang, X., Kapurniotu, A., Bernhagen, J., Teichberg, S., Basgen, J., et al. (1996). An agent cleaving glucose-derived protein crosslinks in vitro and in vivo. *Nature, 382*(6588), 275–278. https://doi.org/10.1038/382275a0.
Waltz, E. (2016). Gene-edited CRISPR mushroom escapes US regulation. *Nature, 532*(7599), 293. https://doi.org/10.1038/nature.2016.19754.
Wiedenheft, B., van Duijn, E., Bultema, J. B., Bultema, J., Waghmare, S. P., Waghmare, S., et al. (2011). RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proceedings of the national academy of sciences of the United States of America: 108*, (pp. 10092–10097). . https://doi.org/10.1073/pnas.1102716108 25.
Xu, X., Duan, D., & Chen, S.-J. (2017). CRISPR-Cas9 cleavage efficiency correlates strongly with target-sgRNA folding stability: From physical mechanism to off-target assessment. *Scientific Reports, 7*(1), 143. https://doi.org/10.1038/s41598-017-00180-1.
Zhang, Y., Ge, X., Yang, F., Zhang, L., Zheng, J., Tan, X., et al. (2014). Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells. *Scientific Reports, 4*, 5405. https://doi.org/10.1038/srep05405.