Proc. Natl. Acad. Sci., India, Sect. B Biol. Sci.
https://doi.org/10.1007/s40011-024-01582-8

RESEARCH ARTICLE

DNA Barcoding for Paprika Authentication and Quality Assessment on the Moroccan Market

Aissatou Sakho[1] · Saloua Tazi[1] · Elmostafa El Fahime[2] · Souad Skalli[1]

Received: 29 December 2022 / Revised: 3 January 2023 / Accepted: 19 March 2024

Abstract Paprika powder (*C. annuum* L.) is widely used as a culinary and therapeutic spice. The present study was conducted to authenticate eight samples of paprika purchased in different markets in the Rabat region added with a reference sample collected in the Beni Mellal-Khenifra region using DNA barcoding. The applicability of the *rbcL*, *psbA-trnH*, *psbA-trnH_PA/TH* and *ITS* DNA regions as barcodes was tested for paprika authentication. For this purpose, DNA extraction, polymerase chain reaction (PCR), sequencing and phylogenetic trees were performed. PCR amplification for *rbcL*, *psbA-trnH*, *psbA-trnH_PA/TH* and *ITS* yielded amplicons of 200 bp, 750 bp, 700 bp and 200 bp size, respectively. ITS barcode sequences were of poor quality. The *rbcL* barcode was not discriminative, unlike the *psbA-trnH*, *psbA-trnH _PA* regions. Phylogenetic analysis showed that from the 8 samples of paprika powder, only the one coming from Bab el Had (B) was not authentic.

Significance Statement: The objective of this work is to determine and compare the authentication of paprika available on the Moroccan market by adopting the DNA barcode technique, allowing the evaluation of the quality of *Capsicum annuum* L. The manuscript provides proof, using DNA barcodes, that non-authentic paprika is present on the Moroccan market. This study is the first on the authentication of paprika from Morocco that is in high demand on the international market. We believe that these results could be of a great interest to readers interested in the topic of the authentication of medicinal plants.

✉ Aissatou Sakho
aissatou.sakho@um5r.ac.ma

[1] Faculty of Science, Mohammed V University in Rabat, 4 Avenue Ibn Battouta, RP, B.P. 1014, Rabat, Morocco

[2] Functional Genomic Platform, UATRS, Center for Scientific and Technical Research (CNRST), 10000 Rabat, Morocco

Keywords *Capsicum annuum* · Authentication · Quality control · *rbcL* · *psbA-trnH* · *ITS*

Introduction

Sweet bell pepper (*C. annuum* L.) is the second most consumed vegetable in the world [1]. Generally, sweet peppers are recognized as a potential dietary source of vitamins, phenolic compounds, carotenoids and flavonoids, all of whom possess known positive health effects [2].

Red peppers are most commonly consumed raw or powdered as a spice (paprika) [3]. In Morocco, red bell pepper is used to make paprika called "tahmira" or "niora." Niora is mainly grown in the Tadla region with 80% of the national production [4].

The global spice industry is perpetually threatened by fraudulent practices [5]. Paprika, a widely consumed spice, remains thus a primary target for adulteration [6]. To control food fraud, various technologies such as spectrophotometry and chromatography have been developed. Among these techniques, DNA barcoding has the biotechnological advantage to enable quality control and commercial monitoring of medicinal plants [7].

The objective of this work is to study the authenticity of several samples of paprika powder sold in different markets in the region of Rabat Morocco using DNA barcoding. This is an original work whose purpose is to enrich the Moroccan database on the genomics of plant species and to put at the service of the national system of pharmacovigilance of herbal medicines a robust and reliable molecular screening technique to detect the quality of the paprika powder.

Published online: 13 June 2024

Fig. 1 Different samples of red bell pepper powder purchased in Rabat market

Table 1: Purchases locations with codes assigned for the different samples of *C. annuum* L.

Samples	Code
Manal	M
Kamra	K
Bab el had	B
Temara	T
Salé	S
Marjane (weighted)	MP
Marjane (Javana)	MJ
Herbio	H
Reference sample	R

Material and Methods

Plant Material Collection

Eight powdered samples of paprika were procured from different local markets at Rabat region (Salé, Temara, Kamra, Bab el had, Marjane, Herbio cooperative, and Manal) and were used in this study (Fig. 1). To keep the traceability of the samples, paprika powders were put in bags on which the date and place of the purchase were written as well as the code assigned to them (Table 1). A reference sample was collected from a red bell pepper field in the Beni Mellal-Khenifra region and deposited at the herbarium of the Scientific Institute of Rabat to obtain a voucher code. The dried fruit from the *C. annuum* L. plant was used as a reference sample for comparison with the market samples of paprika.

DNA Extraction

Total genomic DNA was isolated from the dried fruit of the plant reference and paprika samples from the market using the MagPurix plant DNA extraction kit. The obtained genomic DNA was quantified and its quality was assessed with NanoDrop 8000 spectrophotometer and stored at − 20 °C.

PCR Amplification and Sequencing

Four barcoding loci, *psbA-trnH*, *psbA-trnH_PA/TH*, *rbcL* and *ITS,* were selected for PCR amplification using universal primers (Table 2). The reaction mix containing 100 ng of DNA, 5 µl of Buffer, 0.5 µl of Taq polymerase (BIOLINE, Memphis, USA), 0.5 µl of DNTP and 1 µl of each primer (forward/reverse) was prepared in a 1.5-ml microtube. Distilled water was added to each tube to fill them to a final volume of 25 µl. PCRs were performed in an automated thermal cycler Aeris™ Thermal Cycler Model G96 (BIOEVOPEAK, Jinan, CHN), under the optimized conditions for each locus (Table 2). The amplification first began with an initial denaturation phase at 95 °C for 2 min. Then 34 cycles of a three-step temperature profile are as follows: denaturation at 95 °C for 30 s, hybridization at a specific temperature for 30 s and elongation at 72 °C for 30 s. The PCR program was completed with a final elongation period of 30 s.

Table 2: Universal primers, their sequences and hybridization temperatures

Primer name	Sequence (5′–3′)	Hybridization temperature (°C)
rbcL R *rbcL* F	ATGTCACCACAAACAGAGACT TGTCCATGTACCAGTAGAAGA	62
psbA-trnH_pA/TH psbA-trnH _pA/TH	GTTATGCATGAACGTAATGCTC CGCGCATGGTGGATTCACAATCC	63
psbA-trnH F psbArnH R	CGCGCATGGTTGGATTCACAATCC GTTATGCATGAACGTAATGCTC	63
ITS F*ITS* R	TCCTCCGCTTATTGATATGC CCTTATCATTTAGAGGAAGGAG	57

To visualize the obtained PCR products, 1% agarose gel electrophoresis in 1X Tris Acetate EDTA (TAE) buffer with 100 bp size marker was performed (Parvathy et al., 2014). The mixture was heated for 1min 30 s in a microwave and then cooled under water; then, 2 µl of ethidium bromide (BET) was added. The gel was left to stand for 30 min until it solidified. For each sample, 8 µl of PCR product was mixed with 2 µl of "bromophenol blue" buffer and then deposited into the gel wells. Six microliters of 100 bp molecular marker (HyperLadder™) (BIOLINE, Memphis, USA) was deposited in the first well. DNA migration was done in the tank filled with TAE at 120 V for 35 min. The PCR products were visualized using the U-Violet light (BIO-RAD Gel Doc XR+) (Bio-Rad, California, USA). The PCR products were cleaned up using ExoSAP-IT Express kit; then, sequencing reaction was performed using BigDye Sequencing Kit. Finally, sequencing reaction cleanup was done using ethanol-EDTA method. All steps were done following standard protocol, and sequencing run was performed bi-directionally using Seq Studio Applied Biosystems genetic analyzer (Thermo Fisher, Waltham, USA).

Data Analysis

The resulting sequences of the barcode regions were assembled using DNA Baser Sequence Assembler version 5.15. After automatic edit, sequences were organized in FASTA file format. Contaminated and low-quality sequences were removed from the data set. The assembled sequences were used as queries in the BLAST (Basic Local Alignment Search Tool) analysis to find related nucleotide sequences in the NCBI sequence database (National Center for Biotechnology Information). The sequences from the GenBank that had maximum query coverage, highest homology and highest score were downloaded as FASTA files and were included in the analyses. A multi-alignment was done using MEGA-11 to assess the identity degree and conservation of commercial paprika barcode sequences with those of the reference material. Then, two barcode regions were concatenated to produce combinations of data sets.

Phylogenetic Analysis

Phylogenetic analyses were performed with MEGA-11 based on the barcode regions of samples that showed success during sequencing. The neighbor-joining (NJ) method was used, and 500 bootstrap replicates were applied to measure the robustness of trees. Phylogenetic trees of single and concatenated barcode regions were constructed.

Results and Discussion

Plant Material Collection

The botanical confirmation at the scientific institute of Rabat certified that the reference plant coming from Beni Mellal-Khenifra was *C. annuum* L. and the voucher code RAB113638 was obtained (Fig. 2).

DNA Extraction

Quantitative and qualitative analyses showed absorbance ratios mostly between 1.8 and confirming that the isolated DNA is pure, free of contamination and undegraded. DNA concentrations obtained ranged from 46.37 to 211.16 ng/µl.

PCR Amplification and Sequencing

PCR products migration on 1% Agarose gel showed 100% amplification efficiency of all the four. PCR amplification

Fig. 2 Reference material from the region of Beni Mellal-Khenifra

of *rbcL*, *psbA-trnH_PA/TH*, *psbA-trnH* and *ITS* loci yielded single and specific band sizes of 200, 700, 750 and 200 bp, respectively (Fig. 3).

After sequencing, 56 sequences (forward and reverse) of *ITS* (13), *rbcL* (13), *psbA-trnH* (14) and *psbA-trnH_PA/TH* (16) regions were obtained and visualized by DNA Baser Assembler software.

Data Analysis

Sequences of most samples were clear and non-overlapping except for those generated from *ITS* barcode which were of poor quality. Sequences generated by *rbcL* were short (< 200 bp) compared to both *psbA-trnH* and *psbA-trnH_PA/TH* genes (> 700 bp). *ITS* sequences could not be assembled due to low quality. In addition, *rbcL* sequences from K and T, as well as *psbA-trnH* sequence from MP and the reference, did not yield a consensus sequence due to their poor quality. Therefore, these data were removed from the analysis.

BLASTn tool showed that most of the sequences have a high percentage of similarity (> 99%) to *C. annuum* L. sequence available on NCBI for *psbA-trnH* and *psbA-trnH_pA/TH* loci except those from Bab el Had (91.68%). Blast identification of this sample only resulted in *Capsicum sp.* identification. The reference sample (R) showed the highest similarity with 99.82 and 99.83%, respectively, for *psbA-trnH* and *psbA-trnH _PA/TH*; this proves that our reference sample is indeed *C. annuum* L.

Multi-alignment of the sequences showed a set of identical and conserved positions between the different sequences (market and reference samples) for the two genes *psbA-trnH* and *psbA-trnH_pA/TH* showing their high degree of similarity. Only sample B showed some differences from the reference material.

Phylogenetic Analysis

Phylogenetic trees designed from the barcode regions of the *psbA-trnH_pA/TH* gene and the *psbA-trnH* gene (Figs. 4 and 5) showed similar results and allowed us to classify the targeted samples into 2 divergent groups (clade A and B). In addition, another outgroup species (*Oryza sativa*) was added to root the phylogeny and allow meaningful comparison.

Clade A is a monophyletic clade isolated from clade B by a bootstrap of 60% and 61% for *psbA-trnH_pA/TH* and *psbA-trnH,* respectively. It contains H, R, S, K, M, T, MJ, MP.

Clade B only contains sample B.

The third phylogenetic tree (Fig. 6) is the result of the concatenation and combination of the two genes *psbA-trnH* and *psbA-trnH_PA/TH*. The results obtained are

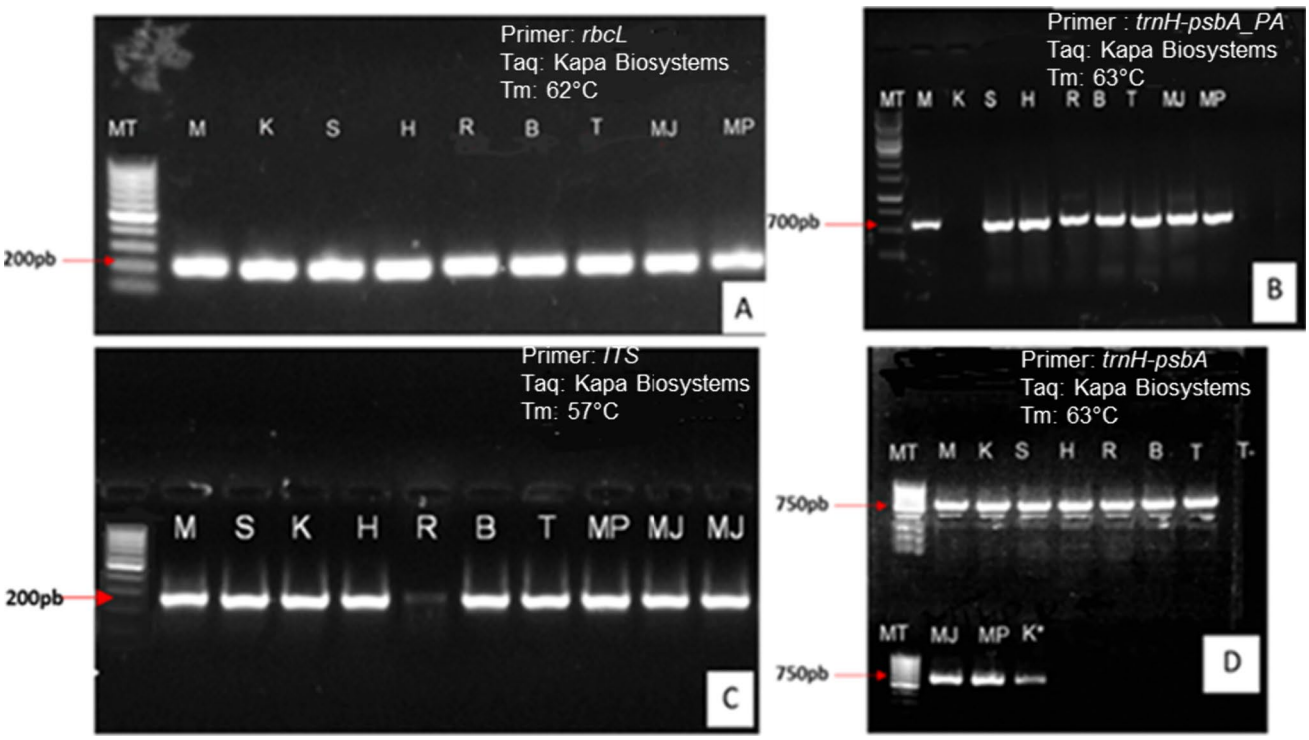

Fig. 3 Amplification of **A** *rbcL*, **B** *psbA-trnH _PA/TH*, **C** ITS and **D** psbA-*trnH*. MT: marker size of 100 bp and 1000 bp. M, K, S, H, R, B, T, MP, MJ: codes assigned for samples from different regions. T-: negative control

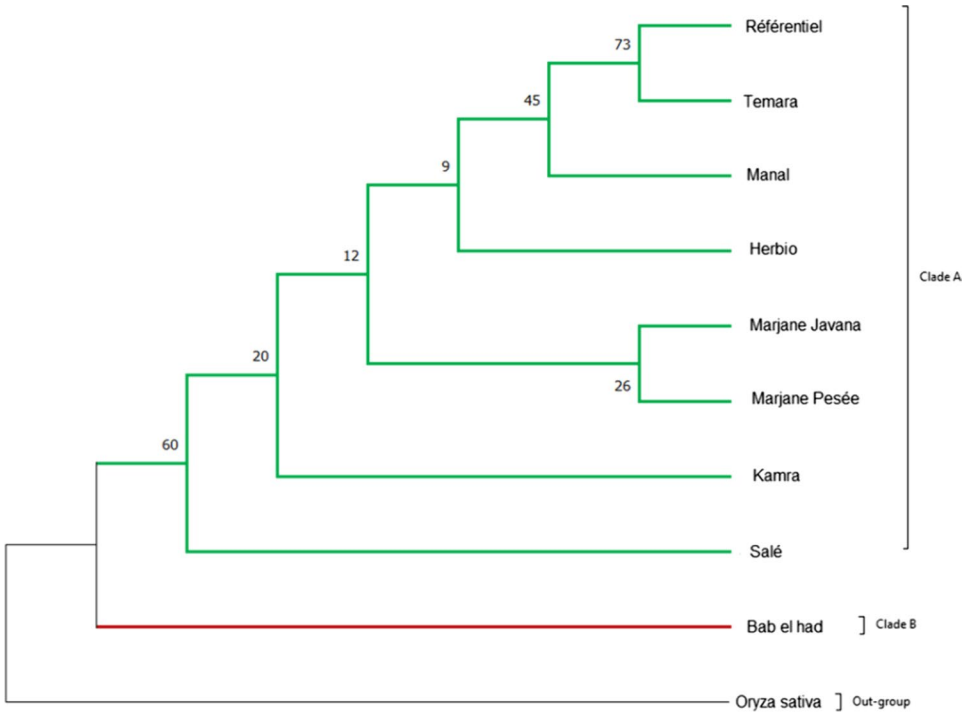

Fig. 4 The N-J tree constructed from *psbA-trnH_pA/TH* sequences using the Neighbor-Joining method, with statistical significance based on the bootstrap of 500 replicates.

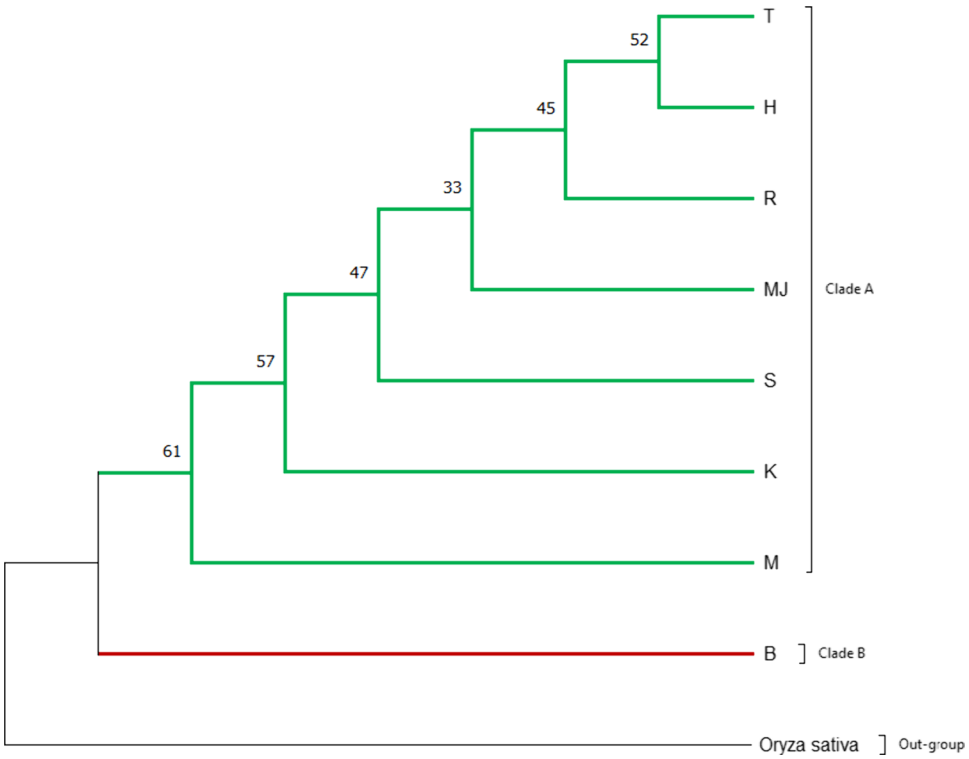

Fig. 5 Neighbor-Joining tree constructed from *psbA-trnH* sequences with statistical significance based on the bootstrap of 500 replicates.

similar to those of the first two trees. The samples are separated into two distinct clades, clade A which groups the reference with H, T, MJ, S, K and M samples. The B sample forms an isolated clade.

Phylogenetic tree construction from *ITS* and *rbcL* was impossible because the sequences could not be assembled due to their poor quality.

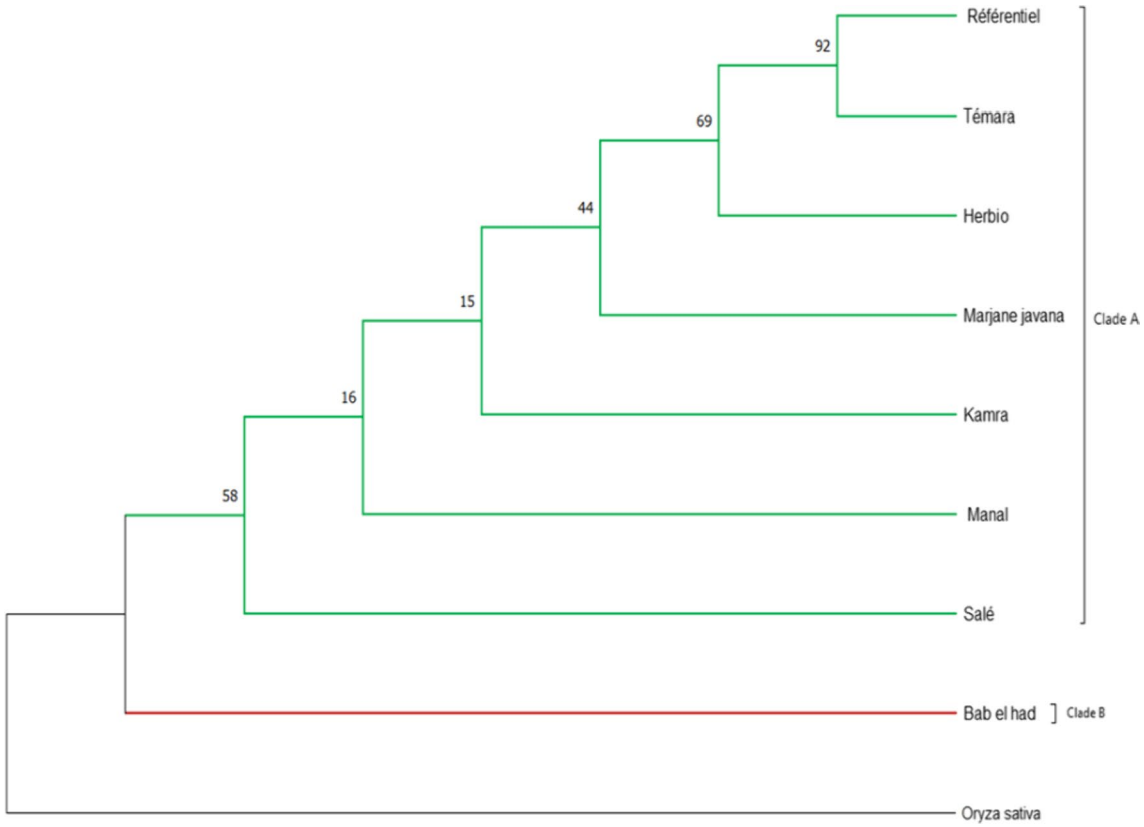

Fig. 6 Phylogenetic tree constructed from concatenation of *trnH-psbA_PA* and *trnH-psbA* sequences using the Neighbor-Joining method, with statistical significance based on the bootstrap of 500 replicates.

Thus, considering all the results obtained and according to the two trees generated by the *psbA-trnH_pA/TH* and *psbA-trnH* genes, only B sample is not authentic, since it forms a separate clade than that of the reference.

DNA barcoding is an advanced marker-based molecular technology that is efficient and fast for the evaluation and identification of target plant species [8] in various forms, including powder [9].

To authenticate paprika powder sold in the Rabat region, four barcode regions *rbcL*, *psbA-trnH*, *psbA-trnH_PA/TH* and *ITS* were tested in this study. Raclariu-Manolică et al [10] reported that paprika manufacturing process involves excessive heating during the drying of peppers which leads to degradation of the DNA into smaller fragments [11]. PCR amplification of *rbcL*, *psbA-trnH*, *psbA-trnH_PA/TH* and *ITS* barcode markers was successful with all samples. The *psbA-trnH* and *rbcL* DNA barcoding loci are reported to have high amplification success with universal primers [12]. The sequencing success rate was high for *psbA-trnH* and *psbA-trnH_PA*/TH regions. These two loci provided accurate species identifications at over 98% identity with *C. annuum* L. for most of the samples except B sample. For *rbcL*, the sequences obtained were of short size (< 200 bp) and sequences alignment revealed no differences. *ITS* sequences were of very poor quality. Sequencing and amplifying difficulties of this barcode are the main reasons that have so far prevented it from being a core component of the plant barcode [13]. Shiragaki et al [14] also reported that the preliminary attempt to use direct nucleotide sequencing for *ITS* amplicons often failed.

Phylogenetic analysis using *psbA-trnH* and *psbA-trnH_PA/TH* sequences showed good discrimination. The samples were split into 2 groups, B sample formed a separated clade from others samples. The concatenation of *psbA-trnH* and *psbA-trnH_PA* sequences result still showed the same result. All these results indicate that out of the 8 samples of paprika powder analyzed, only the B sample is not authentic.

Both *psbA-trnH* and *psbA-trnH_PA/TH* genes might be a better choice for paprika barcoding. The efficiency of *psbA-trnH* has been reported in several works. Rosario et al [15] used the DNA barcoding tool to identify solanaceous species including *C. annuum* L. Phylogenetic analysis of the *psbA-trnH* gene showed better species-level discrimination compared to *maturaseK (matK)*. Similar results were reported by Cristina et al [16] in their study to discriminate between species of *Solanum* subgenus *Leptostemonum*. The *psbA-trnH* barcode region clearly shows a high degree of intra- and inter-specific evolutionary relationship among all

medicinally important plant families (Khan Shinwari et al. 2018). Indeed, it has high insertion/deletion rates. This makes it a very suitable plant barcode for species discrimination [17].

Conclusion

Paprika (*C. annuum* L.) is a high-volume, high-value spice. This spice is prized for its scent, color, and medicinal effects. Paprika has medicinal benefits, but only if it is authentic. This study proved that DNA barcoding can authenticate powdered spice samples effectively and reliably. This work is merely the first step toward authenticating *C. annuum* L. by DNA barcoding. This tool cannot identify herbal raw materials independently. This work has to be developed to consolidate the results, uncover paprika frauds on a national scale and give authorities a reliable tool to enhance molecular identification of *C. annuum* L. In addition, Morocco needs robust legal frameworks to promote the purity of herbal goods.

Author Contributions All authors contributed equally to the study conception and design. Material preparation, data collection and analysis were performed by AS, ST, EELF, SS. The first draft of the manuscript was written by AS, ST, EELF, SS, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Declarations

Conflict of interest The authors have no conflicts of interest that are directly relevant to the content of this study.

References

1. Guo WL, Chen RG, Du XH et al (2014) Reduced tolerance to abiotic stress in transgenic Arabidopsis overexpressing a *Capsicum annuum* multiprotein bridging factor 1. BMC Plant Biol 14(1):1–13. https://doi.org/10.1186/1471-2229-14-138/FIGURES/12
2. Guilherme R, Aires A, Rodrigues N, Peres AM, Pereira JA (2020) Phenolics and antioxidant activity of green and red sweet peppers from organic and conventional agriculture: a comparative study. Agriculture 10(12):652. https://doi.org/10.3390/AGRICULTURE10120652
3. Baenas N, Belović M, Ilic N, Moreno DA, García-Viguera C (2019) Industrial use of pepper (*Capsicum annum* L.) derived products: technological benefits and biological advantages. Food Chem 274:872–885. https://doi.org/10.1016/J.FOODCHEM.2018.09.047
4. Zaki N, Hakmaoui A, Ouatmane A, Fernandez-Trujillo JP (2013) Quality characteristics of Moroccan sweet paprika (*Capsicum annuum* L.) at different sampling times. Food Sci Technol 33(3):577–585. https://doi.org/10.1590/S0101-20612013005000072
5. Galvin-King P, Haughey SA, Elliott CT (2018) Herb and spice fraud; the drivers, challenges and detection. Food Control 88:85–97. https://doi.org/10.1016/J.FOODCONT.2017.12.031
6. Oliveira MM, Cruz-Tirado JP, Roque JV, Teófilo RF, Barbin DF (2020) Portable near-infrared spectroscopy for rapid authentication of adulterated paprika powder. J Food Compos Anal 87:103403. https://doi.org/10.1016/J.JFCA.2019.103403
7. Nehal N, Choudhary B, Nagpure A, Gupta RK (2021) DNA barcoding: a modern age tool for detection of adulteration in food. Crit Rev Biotechnol 41(5):767–791. https://doi.org/10.1080/07388551.2021.1874279
8. Khan Shinwari Z, Jan SA, Talha Khalil A et al (2018) Identification And Phylogenetic analysis of selected medicinal plant species from Pakistan: DNA barcoding approach. Pak J Bot 50(2):553–560
9. Jiao J, Huang W, Bai Z, Liu F, Ma C, Liang Z (2018) DNA barcoding for the efficient and accurate identification of medicinal polygonati rhizoma in China. PLoS One 13(7):e0201015. https://doi.org/10.1371/JOURNAL.PONE.0201015
10. Raclariu-Manolică AC, Anmarkrud JA, Kierczak M et al (2021) DNA metabarcoding for quality control of basil, oregano, and paprika. Front Plant Sci 12:665618. https://doi.org/10.3389/FPLS.2021.665618/FULL
11. Parveen I, Gafner S, Techen N, Murch SJ, Khan IA (2016) DNA Barcoding for the identification of botanicals in herbal medicine and dietary supplements: strengths and limitations. Planta Med 82(14):1225–1235. https://doi.org/10.1055/S-0042-111208
12. Parvathy VA, Swetha VP, Sheeja TE, Leela NK, Chempakam B, Sasikumar B (2014) DNA barcoding to detect chilli adulteration in traded black pepper powder. Food Biotechnol 28(1):25–40. https://doi.org/10.1080/08905436.2013.870078
13. Hollingsworth PM, Graham SW, Little DP (2011) Choosing and using a plant DNA barcode. PLoS One 6(5):e19254. https://doi.org/10.1371/JOURNAL.PONE.0019254
14. Shiragaki K, Yokoi S, Tezuka T (2020) Phylogenetic analysis and molecular diversity of capsicum based on rDNA-ITS region. Horticulture 6(4):87. https://doi.org/10.3390/HORTICULTURAE6040087
15. Rosario LH, Padilla JOR, Martínez DR et al (2019) DNA barcoding of the Solanaceae family in Puerto Rico including endangered and endemic species. J Am Soc Hortic Sci 144(5):363–374. https://doi.org/10.21273/JASHS04735-19
16. Cristina I, Sanchez C, Giraldo C, Gonzalez MA (2022) Molecular identification of the economically important *Solanum* subgenus *Leptostemonum* (Solanaceae) using DNA barcodes cryptic diversity in amphibians and reptiles of Colombia view project phytoauthent-molecular authentication of complex herbal food supplements for safety and efficacy view project. Artic J Plant Biochem Biotechnol 31:938–952. https://doi.org/10.1007/s13562-022-00773-6
17. Li X, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen S (2014) Plant DNA barcoding: from gene to genome. Biol Rev Camb Philos Soc 90(1):157–166. https://doi.org/10.1111/BRV.12104