Pere *et al. BMC Plant Biology* (2024) 24:1178
https://doi.org/10.1186/s12870-024-05889-6

BMC Plant Biology

RESEARCH Open Access

Molecular identification, genetic diversity, and secondary structure predictions of *Physalis* species using ITS2 DNA barcoding

K. Pere[1], K. Mburu[2], E. K. Muge[1], J. M. Wagacha[3] and Evans N. Nyaboga[1*]

Abstract

Background The genus *Physalis* belongs to the *Solanaceae* family and has different species with important nutritional and medicinal values. Species within this genus have limited morphological differences, a characteristic that hinders accurate identification, safe utilization and genetic conservation of promising genotypes. In addition, to prevent the perceived loss of *Physalis* diversity due to habitat destruction, species delimitation needs attention. In this study, we used the sequence and structural information of the internal transcribed spacer 2 (ITS2) barcode to efficiently identify and discriminate *Physalis* species from a collection of 34 *Physalis* accessions.

Methodology *Physalis* plant samples were collected from eight Counties in Kenya based on the availability of the germplasm. The voucher specimens were identified using the botanical taxonomy method and were deposited in the University of Nairobi herbarium. A total of 34 *Physalis* accessions were identified and accessed for diversity based on the ITS2 barcode region. The sequence similarity of the ITS2 genes was analyzed through the Basic Local Alignment Search Tool (BLAST), the nearest Kimura-2-parameter (K2P) genetic distances were calculated and a phylogenetic tree was constructed using the Bayesian inference (BI) method in MrBayes 3.2.7a software. The differences in the ITS2 secondary structure between the species were analyzed.

Results The success rate of PCR amplification and sequencing was 75% and 67%, respectively. The analyzed ITS2 sequences displayed significant inter-specific divergences, clear DNA barcoding gaps and high species identification efficiency. Based on the constructed phylogenetic tree, three *Physalis* species (*Physalis peruviana*, *Physalis purpurea* and *Physalis cordata*) were identified and were clustered in a homogenized distribution. High genetic diversity (0.36923) and genetic distance (0.703) were observed between *Physalis peruviana* and *Physalis cordata*. The highest genetic nucleotide diversity (0.26324) and distance (0.46) within species was obtained for *Physalis peruviana*. The differences in the secondary structures generated from this study discriminated between the *Physalis* species.

Conclusions Our study demonstrated that ITS2 is a potential DNA barcode for effective identification and discrimination of *Physalis* species. The results of this study provide insights into the scientific basis of species identification, safe utilization, genetic conservation and future breeding strategies for this important nutritional and medicinal plant species.

*Correspondence:
Evans N. Nyaboga
nyaboga@uonbi.ac.ke

Full list of author information is available at the end of the article

Keywords DNA barcoding, ITS2, Phylogenetic, *Physalis*, Secondary structure predictions, Species identification and discrimination

Background

The genus *Physalis* is a well-known herbaceous plant that belongs to the family *Solanaceae*. *Physalis* is native to the Andes of South America which includes countries such as Colombia, Peru and Ecuador [1]. The genus has several species that grow in the wild with only few species cultivated like *P. peruviana* and *P. philadephica* as food crops and ornamentals [2]. *P. peruviana* was initially consumed largely by the Inca people but later introduced to Africa and India after the entry of Christopher Columbus into the Americas [3]. *Physalis* species including *P. peruviana*, *P. alkengi*, *P. purpurea*, *P. pubescens* and *P. philadelphica* are widely cultivated worldwide for their nutritional, medicinal and economic values [4, 5]. These plants are rich in both nutrients (provitamin A, vitamin C, vitamin B complex, phosphorus and fiber) and phytochemicals such as phytosterols, polyphenols, saponins, peruvioses, fisalins and withanolides [6, 7]. Awareness on economic importance of *Physalis* fruits increased over the years in countries like Colombia, Egypt, Kenya, Zimbabwe and South Africa where fruits are being exported [8].

Precise identification of *Physalis* species to ensure safe utilization especially for medicinal uses is paramount [9]. Conservation of *Physalis* genetic resources is vital to prevent loss/extinction of important species, because most natural habitats of plants including *Physalis* species are being destroyed as a result of urbanization [10]. The use of morphological identification of *Physalis* species is unreliable due to phenotypic similarities among the different species [11]. For instance, *Physalis minima* are morphologically confused for *P. angulata* or *P. pubescens* due to similar phenotypic characters [10]. Moreover, morphological characters are influenced by environmental and developmental factors like light intensity and quality of light which results in inaccurate identification of species [12]. Molecular characterization using DNA markers such as simple sequence repeats (SSR) and random amplified microsatellites (RAM) have been used to comprehensively show genetic variability among genotypes [1, 13, 14]. However, the use of these markers does not provide sufficient discriminating capacity for classifying the *Physalis* genotypes into different species. A simple and accurate method is necessary for foolproof identification and determination of genetic relationships between the cultivars of *Physalis* species.

DNA barcoding is a reliable tool for studying genetic relationships between plants for species identification and delineation [15, 16]. Chloroplast-plastid region DNA sequences (such as *matK*, *rbcL*, *psbA-trnH*, *ycf*) and a nuclear internal transcribed spacer (ITS) region with less interspecific barcode gaps has been used for plant barcodes [17]. The ITS2 is a DNA spacer localized between the ribosomal 5.8 S and 28 S, in the chromosomal or corresponding polycistronic transcript region most used for species discrimination studies [18]. DNA barcode, ITS2 has been proposed as an efficient barcode of medicinal plants [19]. It has been utilized in *Physalis* identification where it showed efficient species discrimination compared to chloroplast DNA barcodes due to its advantages such as small fragment length, good universality, high interspecific divergence and small intraspecific variation [20]. A study done on the identification of *Physalis* species using ITS2 in China showed that the barcode is effective in species identification [10]. ITS2 has also been used to identify *P. angulata* among *Solanaceae* plants [21].

To date, there have been no reports on the identification of *Physalis* species both in the wild and those cultivated in Kenya. Molecular characterization using simple sequence repeat (SSR) markers [13] assumed that only *Physalis peruviana* is present in Kenya. The aim of the present study is to identify and clarify the phylogenetic relationships of *Physalis* species in Kenya using the sequence and structural information of the ITS2 barcode gene. The ITS2 barcode-anchored species delimitation would be useful for genetic resource conservation augmenting future breeding programs.

Results

Amplification and sequencing success rate

The success rate of PCR amplification of ITS2 genes in samples studied was 75%. Sequencing success rate of the amplicons produced was for the ITS region was 67%. The lengths of the ITS2 sequences generated from *Physalis* accessions were in the range of 237–707 bp, with an average of 523 bp. The mean GC content was 61%, with a range of 55.1–66.9% (Supplementary Table 2). All the sequences generated from the amplification of the ITS2 barcode were successfully deposited into the GenBank database (Supplementary Table 3).

Identification using BLASTn analysis

The ITS2 sequence of each sample was used to perform BLASTn analysis independently to retrieve top hits available in the database and filter them via pairwise identity. The BLASTn analysis results for the 34 *Physalis* accessions revealed that all sequences generated were of the targeted loci. BLASTn analysis of ITS2 sequences from this study identified the 34 (100%) *Physalis* accession as *Physalis* species. The highest similarity recorded for the 34 ITS2 sequences was 99.37% for *Physalis peruviana*

(AY665914.1) (Supplementary Table 3). Based on ITS DNA sequences, 23, 4, 3 and 2 accessions were found to be related to *Physalis purpurea*, *Physalis cordata*, *Physalis peruviana* and *Physalis aff. Philadelphica*, respectively. Additionally, one sample each was found to be related to *Physalis minimaculata* and *Physalis microcarpa* (Supplementary Table 3).

Multiple sequence alignment

Delimited ITS2 sequences were used for multiple sequence alignment (MSA) (Supplementary Figs. 1 and 2). The MSA of the curated 34 ITS2 *Physalis* sequences and their 7 reference sequences retrieved from BLASTn analysis had a sequence length of 707 bp. The MSA was compressed using ESPript 3 (http://espript.ibcp.fr) and is indicated in Supplementary Fig. 1 (https://espript.ibcp.fr/ESPript/temp/1101891838/0-0-1680467018-esp.pdf) [22]. This alignment had a high rate of nucleotide substitutions among and between *Physalis* species studied. The substitution mutations entailed both transition and transversion point mutations. The MSA of the 34 ITS2 sequences, prepared by MUSCLE, trimmed and viewed by Jalview had a sequence length of 533 bp. It was compressed using ESPript and is indicated in Supplementary Fig. 2 (https://espript.ibcp.fr/ESPript/temp/1035513530/0-0-1688384112-esp.pdf). The latter MSA also showed substitution mutations of the transition type between species. For example, at position 116 of the alignment, most *P. purpurea* have an adenine nucleotide while the *P. cordata* and *P. peruviana* have a guanine nucleotide. The substitution at position 116 of the MSA indicates a transition mutation between species. On the other hand, at positions 130 and 138 of the MSA, all *P. cordata* have an adenine nucleotide while *P. purpurea* and *P. peruviana* species have a guanine at this position. This is an indication of transition mutations between species.

Phylogenetic-based identification

The phylogenetic tree based on ITS2 sequences for the different *Physalis* species assigned each *Physalis* accessions to its related species (Fig. 1). The 3 *Physalis* species namely *P. cordata* (OQ5372012.1, OQ371998.1, OQ372001.1 and OQ371997.1), *P. peruviana* (OQ372016.1, OQ372008.1 and OQ372006.1) and *P. purpurea* (OQ371996.1, OQ371999.1, OQ372000.1, OQ372002.1 – OQ372005.1, OQ372007.1, OQ372009.1 - OQ2011.1, OQ372013.1 - OQ372015.1, OQ372017.1, OQ372018.1 - OQ372029) formed independent clades with >80% branch support (Supplementary Table 3, Fig. 1), indicating that *Physalis* species could be successfully discriminated using ITS2 sequences. The phylogenetic analysis indicated that the *Physalis* accessions showed species variation with different percentage indices (Fig. 1).

ITS2 RNA secondary structures predictions

Besides the use of divergence of primary sequences of ITS2, variations in ITS2 secondary structures were also used to identify *Physalis* species. The ITS2 secondary structure predictions based on minimum free energy (MFE) are shown in Fig. 2. The optimal secondary structure for *P. cordata* (OQ372001.1), *P. peruviana* (OQ372006.1) and *P. purpurea* (OQ371996.1) had minimum free energies of -204.90 kcal/mol, -312.90 kcal/mol and 266.90 kcal/mol with free energy of thermodynamic ensemble of -210.51 kcal/mol, -322.47 kcal/mol and – 275.94 kcal/mol and the frequency of the MFE structure in the ensemble of 0.01%, 0.00% and 0.00% respectively. The ensemble diversity was 136.00, 146.99 and 149.89 respectively.

The secondary structure of *P. cordata* (OQ372001.1 used as a representative) had 26 double helices, 26 loops and 2 single helices. The secondary structure of *P. peruviana* (OQ372006.1 used as a representative) had 39 double helices, 36 loops and 5 single helices. *Physalis* (OQ371996.1) was chosen as the representative accession for *P. purpurea* and the secondary structure had 43 double helices, 41 loops and 4 single helices. The secondary structure predictions showed variations among the 3 *Physalis* species. The predicted ITS2 secondary structures of the 3 *Physalis* species represented 3 different structures with a central ring and different helical orientations (Fig. 2). The studied *Physalis* species showed a unique secondary structure that differed with the reference structure in two respects, the length of helices and the number of loops on their helices (Fig. 2). The variation in helices length in the secondary structure of ITS2 was observed in different *Physalis* species (Fig. 2). The loop number, position, size and angle from the centroid were distinguishable in all the three *Physalis* species. Besides the differences in the number of stems and rings, the shape and distribution of stem-loops in the secondary structure of the 3 *Physalis* species were notably different (Fig. 2).

Genetic divergence analysis

DNA divergence between populations based on ITS2 sequences

The divergence between *Physalis* accessions in this study was determined by calculating the nucleotide diversity, average nucleotide substitutions per site between populations and number of net nucleotide substitutions per site between populations (Table 1). There were varying shared mutations among the *Physalis* accessions. The highest number of shared mutations at 11 was observed between *P. peruviana* and *P. cordata* populations. The lowest shared mutations at 3 were observed between *P. peruviana* and *P. purpurea*. The highest nucleotide diversity (0.36923) was obtained between *P. peruviana* and *P.*

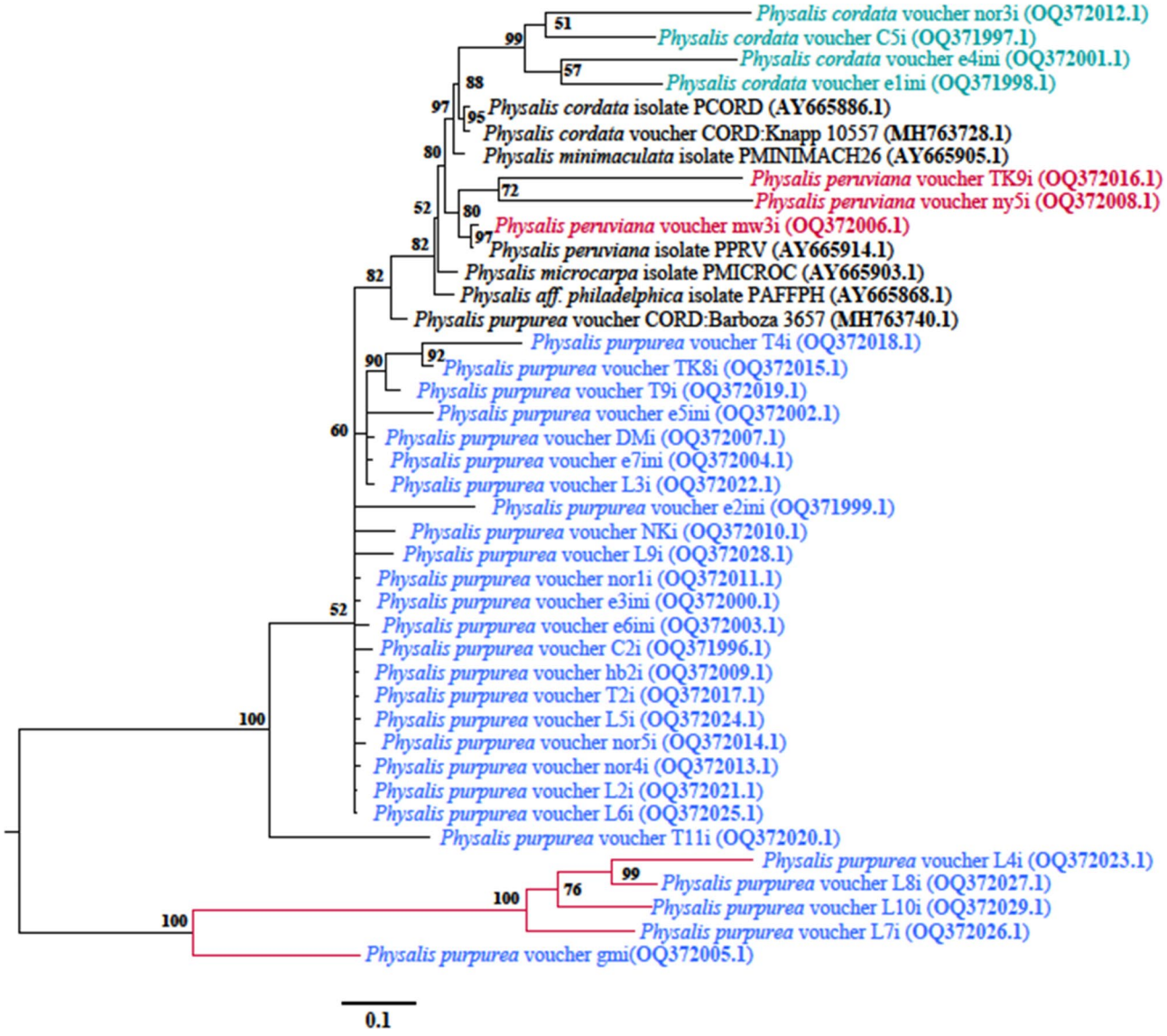

Fig. 1 Phylogenetic tree derived from Bayesian inference analysis of the ITS2 gene of 34 *Physalis* sequences. Plants from this study are presented in blue, red and green colors. Black color represents reference sequences for different *Physalis* species retrieved from GenBank after BLASTn analysis, red represents *P. peruviana*, blue represents *P. purpurea* and green represents *P. cordata*. The numbers above branches represent the percentage posterior probability statistic for the MrBayes phylogenetic tree

cordata while the lowest nucleotide diversity (0.15062) was recorded between *P. peruviana* and *P. purpurea*. There were no fixed differences between *P. peruviana* and *P. purpurea* populations and between *P. purpurea* and *P. cordata*. Two fixed differences were recorded between *P. peruviana* and *P. cordata* populations, which had the highest nucleotide diversity.

DNA divergence within populations based on ITS2 sequences

DNA divergence within each population of the identified *Physalis* species was determined based on ITS2 sequences (Table 2). The highest nucleotide diversity was recorded within the *P. peruviana* population at 0.26324 despite this population having a moderate number of polymorphic segregating sites at 80 and a moderate number of nucleotide substitutions at 89 (Table 2). The lowest nucleotide diversity was recorded for *P. purpurea* population at 0.15883 and this corresponded to the lowest number of polymorphic segregating sites at 37 and the lowest number of substitutions at 37. Moderate nucleotide diversity was recorded for *P. cordata* population at 0.17167 and this population had the highest number of polymorphic segregating sites (89) and number of nucleotide substitutions (102).

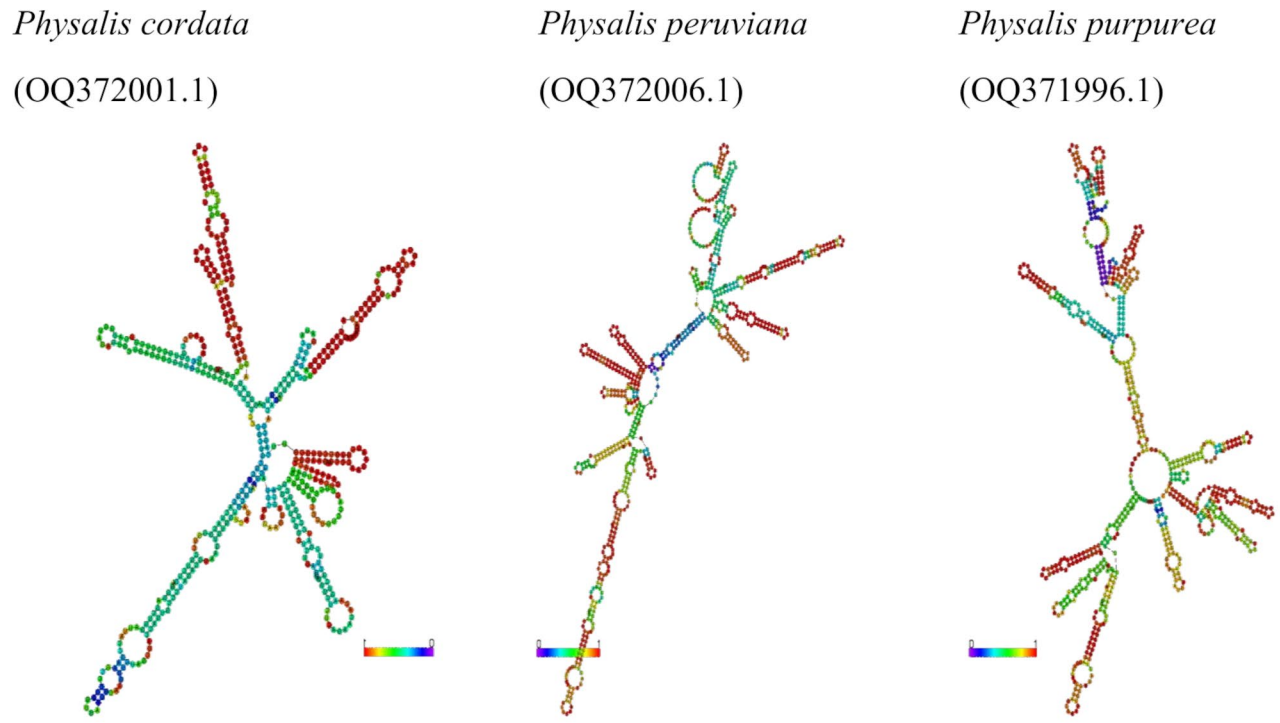

Fig. 2 The predicted secondary structures of the internal transcribed spacer (ITS2) region for the three *Physalis* species (refer each structure to the corresponding species)

Table 1 DNA divergence between (interspecific) *Physalis* species populations based on ITS2 sequences

Population	Physalis peruviana (P1)	Physalis cordata (P2)	Physalis peruviana (P1)	Physalis purpurea (P2)	Physalis purpurea (P1)	Physalis cordata (P2)
Polymorphic sites in each population	37	33	7	16	17	8
Total number of polymorphic sites	49		17		18	
Average number of nucleotide differences	24.000		3.464		3.796	
Nucleotide diversity Pi (t)	0.36923		0.15062		0.16503	
Number of fixed differences	2		0		0	
Mutations polymorphic in population 1 (P1) but monomorphic in population 2 (P2)	32		4		17	
Mutations polymorphic in P2 but monomorphic in P1	28		18		3	
Shared mutations	11		3		6	
Average number of nucleotide differences between populations	25.750		4.284		4.185	
Average nucleotide substitution per site between populations (Dxy)	0.39615		0.18626		0.18196	
Number of net nucleotide substitutions per site between populations (Da)	0.04359		0.01383		0.00449	

Genetic distance between and within *Physalis* species based on ITS2 sequences

The overall average genetic distance among all *Physalis* accessions studied was determined as 0.51±0.04. The highest genetic distance (0.703) was obtained between *P. cordata* and *P. peruviana* and the lowest genetic distance (0.050) was between *P. purpurea* and *P. peruviana* (Table 3).

The average intra-specific distance within each *Physalis* population was determined based on ITS2 sequences. The highest mean intraspecific distance (0.46±0.05) was recorded within the *P. peruviana* with no significant difference with the intra-specific distance of *P. purpurea* (0.43±0.05). The lowest mean intra-specific distance was within the *P. cordata* population at 0.28±0.03.

Nucleotide polymorphism and neutrality tests

Eight polymorphic sites were identified for the ITS2 gene sequences of the 34 *Physalis* accessions used in this study (Table 4). The nucleotide diversity (Pi) of the gene

Table 2 DNA divergence within (intraspecific) *Physalis* species populations based on ITS2 sequences

Physalis species	Physalis peruviana	Physalis cordata	Physalis purpurea
Total number of sequences	3	4	27
Number of polymorphic (segregating) sites (S)	80	89	37
Nucleotide diversity Pi (Total)	0.26324	0.17167	0.15883
Nucleotide diversity Pi (JC-Total)	0.32423	0.19494	0.17848
Theta (Total)	0.27726	0.18545	0.26665
Total number of substitutions	89	102	37

Table 3 Mean genetic distance between (interspecific) *Physalis* species based on ITS2 sequences

Groups	Physalis purpurea	Physalis peruviana	Physalis cordata
Physalis purpurea		0.050	0.057
Physalis peruviana	0.571		0.071
Physalis cordata	0.633	0.703	

Table 4 DNA polymorphism of *Physalis* accessions based on ITS2 marker

Polymorphic sites/ Segregation sites (S)	8	Position in the gene	Variants
Singleton	1	188	4
Parsimony informative sites	7	169	2
		171	2
		172	2
		183	2
		186	2
		189	2
		170	3
Nucleotide diversity (Pi)	0.14810		
Average number of nucleotide differences (k)	1.777		
Sequence length (base pairs)	533		
Number of sequences	34		

sequences was 0.14810 (Table 4). The eight polymorphic sites had one singleton and seven parsimony informative site mutations (Table 4).

Tajima's neutrality test was performed on all 34 ITS2 sequences of *Physalis* accessions to establish selection of the species based on the Tajima D value and the nucleotide diversity. The number of segregating sites (S) and nucleotide diversity (π) was 464 and 0.155388, respectively. The Tajima D value obtained was −1.034267.

Genetic differences and barcoding gap analysis

Automatic Barcode Gap Discovery (ABGD) results generated by K80 Kimura measure of distance based on ITS2 marker for *Physalis* accessions for the determination of the presence of a barcode gap and grouping of species into operational taxonomic groups (Figs. 3, 4 and 5). The pairwise distances were ranked by increasing distance values from 0.02 to 1.36 and barcode gap not detected (Fig. 3). The highest (with a distance value of 0.02) and lowest (with a distance value of 1.36) bars on the histogram represented intraspecific and interspecific divergence, respectively (Fig. 3). The sudden sharp increase in ordered values was not identified (Fig. 4) and therefore no barcode gap.

The 34 *Physalis* accessions were grouped into fifteen groups (*p*-value: 0.001-0.1). The ABGD analysis derived a total of fifteen operational taxonomic units with a prior intraspecific divergence of $P_{max} = 0.1$ based on the ITS2 sequences. A barcode gap was observed at prior intraspecific divergence (p) thresholds of 0.0077 (0.7%) and 0.0159 (1.59%) supporting 15 groupings of the *Physalis* accessions based on ITS2 sequences (Fig. 5).

Discussion

Currently, a growing demand for *Physalis* species worldwide due their nutritional and medicinal values [5]. Physalis species are widely diverse as there are more than 85 species of *Physalis* distributed throughout the world [23]. Different *Physalis* species have different applications [5] and therefore accurately identifying the *Physalis* plants using molecular genetics will enhance the quick and precise identification of species for utilization, genetic resource conservation and development of genetic breeding programs [10]. Morphological identification is not reliable due to high phenotypic similarities among *Physalis* species [24]. It is therefore important to use molecular identification tools such as DNA barcoding in the identification of *Physalis* [25]. This study focused on the use of the DNA ITS2 region as a recognition tag to identify and discriminate *Physalis* species in Kenya.

The results from sequence characteristics, genetic distance, phylogenetic relationships and secondary structure analyses showed the remarkable potential of ITS sequences for distinguishing *Physalis* species. The potential to differentiate at the species level and ease of amplification make ITS a favorable locus for the barcoding of plants [26, 27]. The nuclear ITS region is useful for DNA barcode by several researchers due to its ability to determine high inter-specific genetic divergence among land plants [28, 29]. The success rate of amplification and sequencing of the ITS2 region in this study is comparable to other reported studies whereby the amplification of the ITS2 gene gives a higher success rate than the sequencing [30, 31]. Contrary to the present study, Feng [10] reported 100% PCR amplification and sequencing success rate of the ITS region in *Physalis* species. The amplification and sequencing success rate of ITS2 has been shown in other studies to be much lower compared to other barcoding genes such as *rbcL*, *matK* and *psbA-trnH* [30, 32].

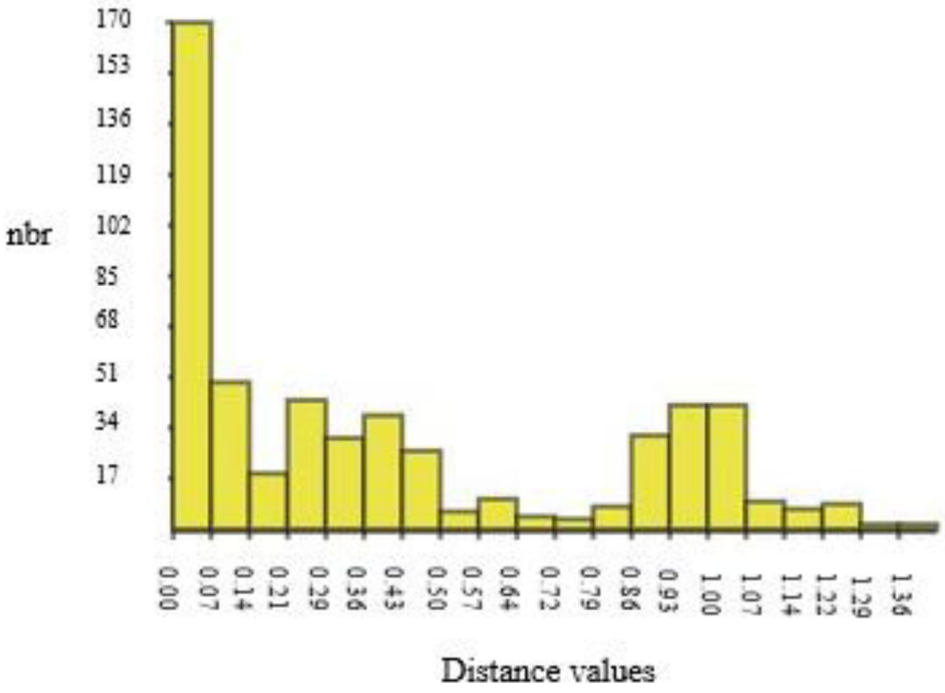

Fig. 3 A histogram indicating the hypothetical distribution of pairwise differences of ITS2 gene sequences for 34 *Physalis* accessions. Low divergence is presumably intraspecific divergence whereas higher divergence indicates interspecific divergence. nbr on the y-axis of the histogram represents a number of pairwise comparisons

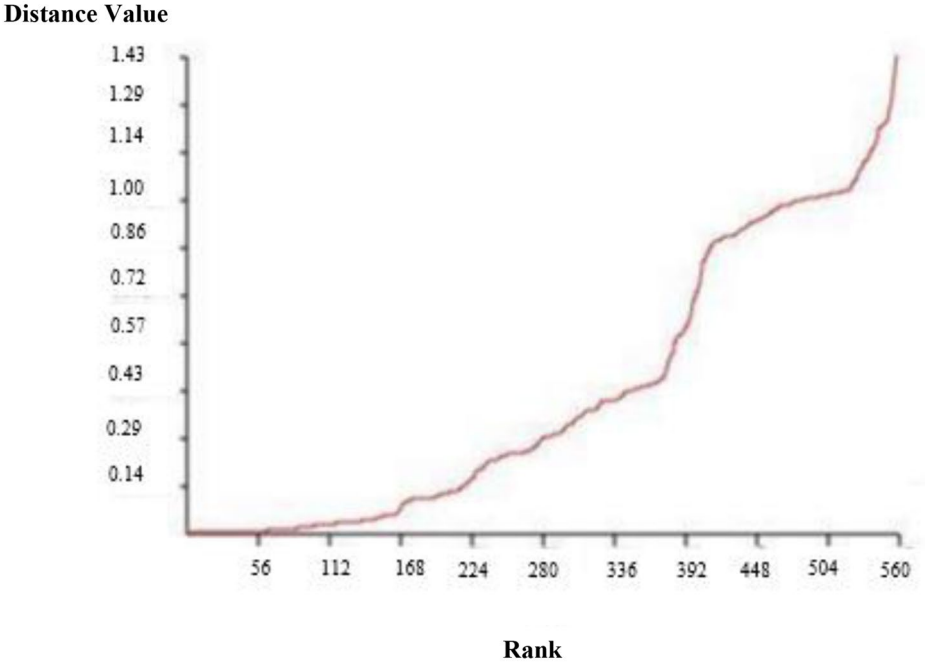

Fig. 4 Ranked pairwise differences of ITS2 gene sequences from 34 *Physalis* accessions. The distribution of pairwise differences in Fig. 3 can be assessed by representing the same as ranked ordered values

The ITS2 phylogenetic tree based on Bayesian inference revealed that the *Physalis* accessions were closely clustered together with the sequences of *P. cordata*, *P. peruviana* and *P. purpurea* chosen from the BLASTn search. The phylogenetic tree indicated that different samples/accessions from the same species can be grouped together. The resultant phylogenetic tree identified and discriminated the *Physalis* accessions into three

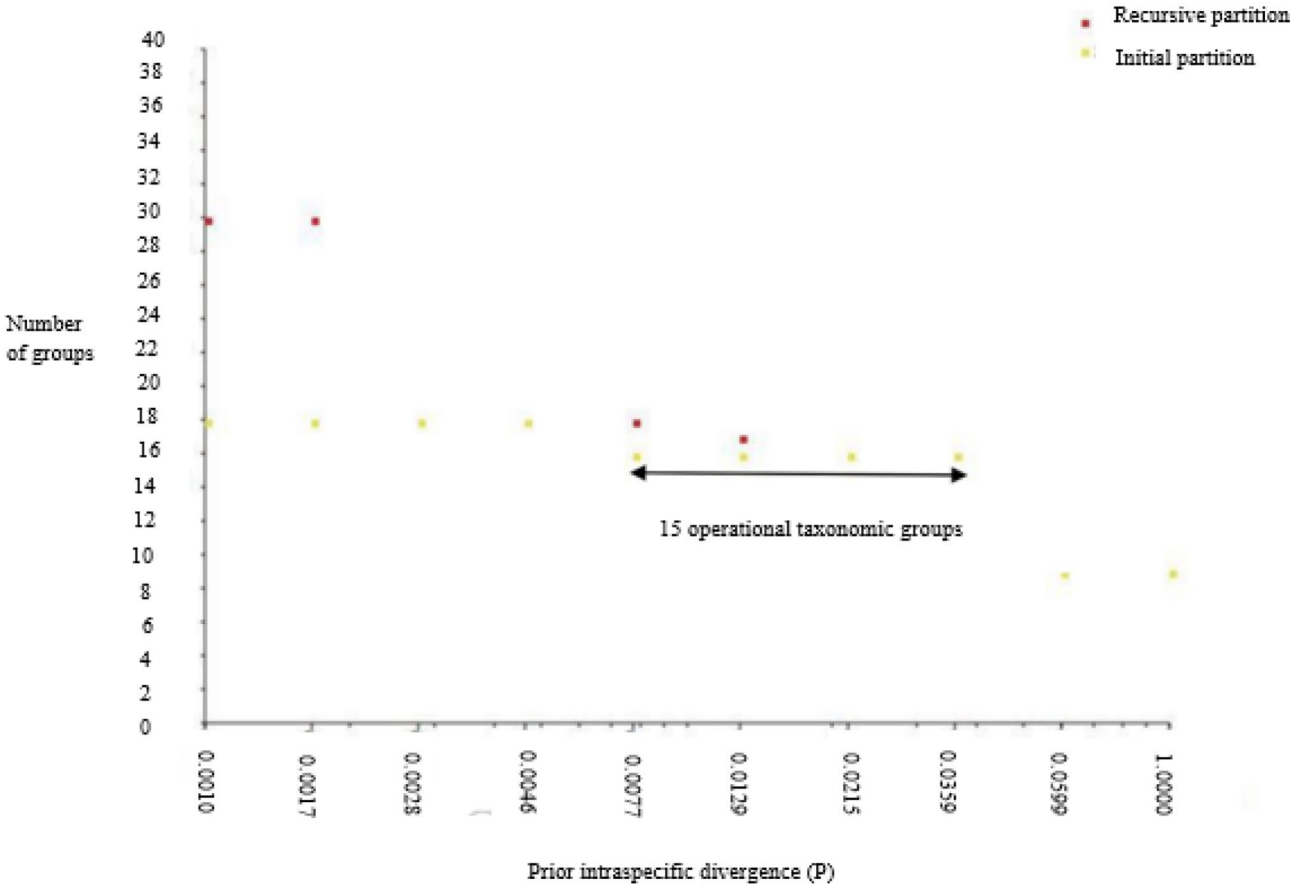

Fig. 5 Automatic partition of 34 *Physalis* accessions defined by each prior intraspecific divergence (P) values of ITS2 gene between 0.001 and 0.1. Results indicate two cases: initial and recursive partition

species namely *P. cordata*, *P. peruviana* and *P. purpurea*. As in previous studies [10, 24] the dendrogram constructed with ITS2 data indicated that the genus *Physalis* is paraphyletic. Studies have shown that plant species identification based on BLASTn and phylogenetic tree analysis are reliable if the species under study have a reference dataset in the GenBank [33]. This is the first report indicating the identification of *Physalis purpurea* and *Physalis cordata* from *Physalis* accessions in Kenya. This is an indication that the ITS2 region can accurately discriminate *Physalis* species and identify new species that had not been previously reported in a particular region. ITS2 was able to effectively distinguish *Physalis* species and facilitate the identification of three *Physalis* species among the accessions studied. This supports previous studies that the ITS2 barcode has a higher discriminatory ability among species than other barcodes such as *rbcL* which tend to be highly conserved in different plant species [34]. In other studies, ITS2 has been identified as a promising DNA barcode due to its fast substitution rate and variability that provides satisfactory resolving power for closely related species [31].

Although ITS2 nucleotide sequences are substituted at a fast rate, their secondary structures are largely conserved and can also be utilized in the identification of species [10, 31]. The ITS2 sequence usually functions as a secondary structure in cells. In addition, the secondary structure contains genetic information that can be used for classification and identification of plant species [35]. In this study, we explored the differences in ITS2 secondary structures between three species of the genus *Physalis*. The ITS2 secondary structure inherently gave a visual distinctiveness between the three *Physalis* species. In our study, the prediction of the secondary structures in the three *Physalis* species revealed diverse secondary structures with distinguishable loop numbers, positions and elevations from the centroid. The ITS2 secondary structures revealed the uniqueness of the generated DNA barcode sequences. These unique genetic structures at the conserved nuclear region of *Physalis* species can be used to develop species-specific markers for the identification of *Physalis* accessions. The prediction of secondary structures differentiation indicated variation among RNA molecules in all the species when using ITS2.

The genetic diversity of *Physalis* accessions studied largely that concurred with the phylogenetic and genetic distance analysis. Both genetic diversity and genetic distance indicated the relatedness of species and in this study, *P. purpurea* and *P. peruviana* seem to be closely related while *P. peruviana* and *P. cordata* seem to be more diverse. Genetic diversity within species is assessed to determine the diversity of a group of organisms within a species. The high nucleotide diversity within *P. peruviana* is an indication that genetic change is much higher in this species than other species identified in this study. This information corresponds with a previous study that showed high genetic divergence within *P. peruviana* populations and even when compared to other related taxa [36]. The ABDG analysis was able to show the presence of a barcoding gap and group *Physalis* accessions into operational taxonomic units. The histogram of pairwise distance of *Physalis* accessions based on ITS2 sequences showed that intraspecific divergence was much higher than interspecific divergence an indication of the lack of a barcoding gap. Intraspecific divergence is important as it creates variation among a species and allows for better conservation of the species in different environments.

The nucleotide polymorphism showed that the 34 *Physalis* accessions had eight polymorphic sites, with one singleton and seven parsimony informative sites. Higher polymorphic sites are consistent with high genetic diversity, an indication that the gene analyzed was effective in the discrimination of *Physalis* species. The Tajima D value for the 34 *Physalis* accessions in this study gave a negative value of -1.034267, an indication that the population had a negative selection and higher low frequency mutations that can help in the differentiation of the *Physalis* species [37]. This study shows that ITS2 is an ideal candidate barcode for accurate identification, discrimination and estimation of genetic diversity of *Physalis* species.

The problem of low species-level identification rates in plants by DNA barcoding is aggravated by the fact that reference databases are not comprehensive. A wider database could determine whether the high success in distinguishing taxa in the genus Physalis is due to the low number of deposited samples in these databases or to the high efficiency of ITS2. Some studies have concluded that DNA barcoding sequences do not usually have sufficient phylogenetic signals to resolve genetic relationships [38]. In addition, multiple copies of ITS2 may suggest that the sequences obtained through PCR are not stable and representative and this might result in misleading phylogenetic inferences [39]. Furthermore, single-locus barcodes do not offer enough variants. However, after comparison with previous studies, the results from this study have demonstrated that ITS2 is a useful DNA barcode that could be used to identify *Physalis* species, and build relatively reliable molecular phylogenies for the genus *Physalis*.

Conclusion

In the present study, ITS2 was found to possess a sufficient variable region between the different species and accessions for genetic divergence with high discriminatory ability. DNA barcode ITS2 was highly efficient in the identification and discrimination of *Physalis* species. ITS2 was found to possess a sufficiently variable region between the different accessions and species for the determination of genetic divergence with high discriminatory ability. The *Physalis* accessions were identified and discriminated into three species namely *Physalis peruviana*, *Physalis purpurea* and *Physalis cordata*. There is need to carry out sampling from a broader geographic range and more accessions to capture the genetic diversity of *Physalis* species. It would be beneficial to compare the effectiveness of ITS2 region with other barcoding genes such as rbcL, matK, and psbA-trnH to provide insights into the relative advantages and limitations of using ITS2 for species identification. Our study revealed significant variations in ITS2 secondary structure predictions that enhanced discrimination among the three identified *Physalis* species. This study provides insight into the scope of the development of species-specific primers for the discrimination of the three *Physalis* species. The secondary structures of the ITS2 region were found to be conserved yet distinct among *Physalis* species. Future work should investigate the functional implications of these structures about species identification and classification, as well as their role in the biology of plants.

Materials and methods

Plant material and sample collection

Leaves of *Physalis* plants were randomly collected from eight Counties in Kenya namely Nyamira, Kiambu, Nakuru, Kajiado, Nyeri, Homa Bay, Elgeyo-Marakwet and Kericho (Fig. 6; Table 5). The samples were collected from April to June, 2019. The eight counties were selected based on the presence of *Physalis* germplasm. The collected *Physalis* plant samples were identified by the taxonomist Mr. Patrick Mutiso and the voucher specimens were preserved in the University of Nairobi herbarium in the Department of Biology (Codes of Voucher Specimens: KP/UON2019/001- KP/UON2019/068) (Supplementary Table 1). Geographical coordinates of locations and the number of *Physalis* plants sampled are presented in Table 5. Leaves were sampled from three plants of 68 *Physalis* accessions (Supplementary Table 1) based on availability as all plants were collected from the wild except for the samples from Elgeyo-Marakwet County collected from a *Physalis* plant farmer. The collected leaf

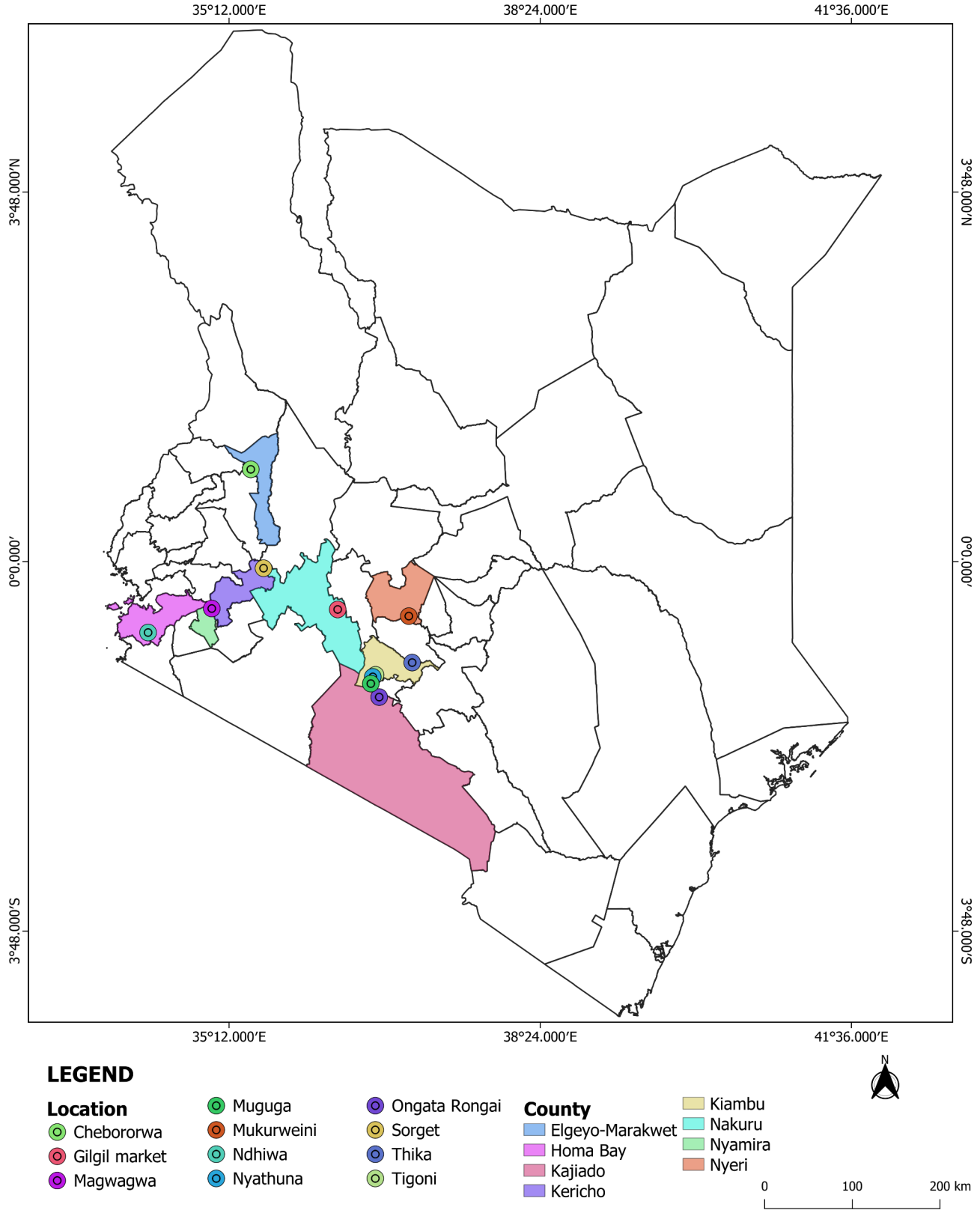

Fig. 6 Locations from the eight counties in Kenya where *Physalis* plants were sampled

Table 5 Geographical coordinates and number of *Physalis* samples collected from each of the selected eight counties in Kenya

No.	Main Location (County)	Specific Location	Latitude	Longitudes	Number of leaf samples collected
1.	Kericho	Londiani, Sorget	0.0684° S	35.5548° E	10
2.	Elgeyo-Marakwet	Chebororwa	0.9487° N	35.4234° E	13
3.	Homa Bay	Ndhiwa	0.7299° S	34.3671° E	3
4.	Nyeri	Mukurweini	0.5609° S	37.0488° E	5
5.	Kajiado	Ongata Rongai	1.3939° S	36.7442° E	5
6.	Nakuru	Gilgil market	0.4923° S	36.3173° E	1
7.	Kiambu	Tigoni	1.1651° S	36.7065° E	17
		Thika	1.0388° S	37.0834° E	9
		Nyathuna	1.1859° S	36.6782° E	1
		Muguga	1.2551° S	36.6580° E	1
8.	Nyamira	Magwagwa	0.4830° S	35.0222° E	3

samples were wrapped with aluminum foil, kept in an ice box and transferred to the Molecular Biology Laboratory in the Department of Biochemistry, University of Nairobi (UoN). The samples were kept at -80 °C before genomic DNA extraction.

Genomic DNA extraction

Frozen leaf samples were first thawed before proceeding with extraction using the Cetyltrimethyl ammonium bromide (CTAB) method as described by Pere [40]. RibonucleaseA (RNAse, 0.6 mg/ml) was added to the DNA samples followed by incubation at 37 °C in a water bath for 30 min to digest any contaminating RNA. The integrity and quality of the extracted genomic DNA were verified using 0.8% (w/v) agarose gel electrophoresis and viewed under a gel documentation system with a UV transilluminator (BioRad, USA). Genomic DNA was stored at -20 °C until required as a DNA template for polymerase chain reaction (PCR) amplification.

PCR amplification and sequencing

Polymerase chain reaction (PCR) amplification for internal transcribed spacer (ITS) of nuclear ribosomal DNA was performed using ITS2 primer. The primer sequences were ITS2-F: 5' – CCTTATCATTTAGAGGAAGGAG –3' and ITS2-R: 5' - TCCTCCGCTTATTGATATGC –3' [41]. PCR amplification was done using a One Taq® Hot start 2× master mix with standard buffer (New England Biolabs, USA) following the manufacturer's instructions. Amplification was conducted in Veriti, 96 well Thermal Cycler (Thermo Fischer Scientific, USA). The annealing temperature for the primer was optimized based on six different temperature regimes namely 50 °C, 51 °C, 52 °C, 54 °C, 56 °C and 58 °C. The optimized PCR profile used for the amplification of the DNA samples comprised of initial denaturation at 94 °C for 5 min, followed by 30 cycles of 94 °C for 30 s, annealing at 58 °C for 45 s, extension at 72 °C for 1 min, and final extension at 72 °C for 7 min.

The PCR products were verified in 1.5% (w/v) agarose gel electrophoresis in 1× TAE buffer. The gels were visualized under a UV transilluminator (BioRad, USA). Amplicons were cleaned using a gel clean-up kit (Applied Biosystems, Thermo Fischer Scientific, USA) and sent for Sanger's sequencing at the University of Nairobi Center of Excellence in HIV Medicine (CoEHM) using a 3730s DNA analyzer (Thermo Fischer, USA). Sequencing of the amplicons was performed in both the forward and reverse directions using the same primers used for their amplification.

Sequence similarity search and multiple sequence alignment

Sequences were manually curated using BioEdit version 7.0.5.3 software [42]. The quality of each sequence was checked by examining the peaks correspond to each nucleotide. A consensus sequence was obtained after trimming the poor-quality DNA sequence ends. The BLASTn algorithm was used to assess the discriminatory efficacy of ITS2 sequences for *Physalis* species. Consensus ITS2 sequences were identified using the BLAST program (https://blast.ncbi.nlm.nih.gov/Blast.cgi) in the NCBI database, applying default parameters. The sequence with the highest similarity was chosen and used as a reference sequence for the *Physalis* species accession in the multiple sequence alignment (MSA) by MUltiple sequence Comparison by Log-Expectation (MUSCLE) version 3.8 [43]. Two MSA were prepared for phylogenetic analysis and genetic diversity studies of *Physalis* based on the ITS2 marker. The first MSA was performed using the ITS2 sequences of 34 *Physalis* accessions from this study and 7 *Physalis* ITS2 reference sequences from the NCBI database while the second was prepared using the ITS2 sequences of the 34 *Physalis* accessions only. The first MSA was used for phylogenetic analysis of *Physalis* accessions based on ITS2 marker while the second was utilized for the determination of genetic diversity, distance, polymorphism, neutrality, and barcoding gap analysis of *Physalis* accessions based on ITS2 marker. The MSA was compressed using ESPript 3 (http://espript.ibcp.fr).

Phylogenetic analysis

A Phylogenetic tree was prepared using the Bayesian inference method by MrBayes software version 3.2.7a (https://nbisweden.github.io/MrBayes/). Statistical analysis of the tree was done using the posterior distribution

of model parameters which was estimated using the Markov Chain Monte Carlo (MCMC) [44–46]. The MCMC sampling was performed over 18,000,000 generations at a sampling frequency of 1000 and the first 25% (relburnin=yes burninfrac=0.25) of samples were discarded when estimating the posterior probabilities of phylogenetic trees. After 18,000,000 generations, the analysis was stopped when the average standard deviation of split frequencies was less than 0.01 and tree parameters summarized. The resulting MrBayes consensus trees constructed were visualized and modified by FigTree software version 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

ITS2 secondary structure predictions

The RNA secondary structure predictions of the identified *Physalis* species were performed using the nucleotide sequences based on three species identified from the MrBayes phylogenetic tree, using the rRNA database of *RNAfold* WebServer v2.4.18 (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) [47–49]. The secondary structures of the representative *Physalis* accessions used in the current study were generated. RNA secondary structures were predicated based on the minimum free energy prediction method.

DNA divergence between and within *Physalis* accession populations

DNA divergence between and within *Physalis* species identified from the MrBayes phylogenetic tree based on ITS2 sequences was determined using the DnaSP software version 6.12.03 [50]. The multiple sequence alignment (MSA) prepared using the 34 ITS2 *Physalis* sequences only was first uploaded and trimmed on Jalview software version 2.11.2.0 to attain uniform lengths utilized for the DNA polymorphism analysis [51, 52]. The edited MSA was then fed into the DnaSP software to assess divergence within and between *Physalis* species accessions.

DNA divergence between the three *Physalis* species identified was based on the analysis of nucleotide diversity (Pi), average nucleotide substitution per site between populations (Dxy) and number of nucleotide substitutions per site between populations (Da) as outlined by Jukes and Cantor algorithm on DnaSP. DNA divergence within each of the three *Physalis* species identified was based on the analysis of several polymorphic segregating sites (S), nucleotide diversity and the total number of substitutions determined as outlined by the Jukes and Cantor algorithm on DnaSP.

Determination of genetic distance between and within *Physalis* accessions

The interspecific, intraspecific and overall mean genetic distance of 34 *Physalis* accessions was calculated based on the ITS2 sequences using the Kimura 2 parameter (K2P) with gamma distribution and a gamma parameter of 0.27 on the MEGA version 11 software [53]. The multiple sequence alignment (MSA) used in the genetic diversity analysis was uploaded into the MEGA 11.0 software and grouping of ITS2 *Physalis* sequences into the 3 identified species was performed. Genetic distance was then calculated between and within grouped species based on the K2P model.

Determination of DNA polymorphism

Sequence polymorphism of the ITS2 sequences was determined for all 34 *Physalis* accessions using DNA Sequence Polymorphism (DnaSP) software. The multiple sequence alignment (MSA) was uploaded into the DnaSP software and a permutation approach and pairwise nucleotide differences used to assess DNA polymorphism of ITS2 gene sequence. DNA polymorphism was assessed based on polymorphic segregating sites, singleton and parsimony informative sites, nucleotide diversity and average number of nucleotide differences.

Tajima's neutrality test

Tajima's neutrality test was performed using MEGA 11.0 software [54]. This test helped to determine the frequency of mutations and selection among the 34 *Physalis* accessions studied as described by Tajima [55]. All ambiguous positions were removed for each sequence pair (pairwise deletion option).

DNA barcoding gap analysis

To delimit *Physalis* species based on their intraspecific divergence within a population and group species into operational taxonomic groups (OTUs) the Automatic Barcode Gap Discovery (ABGD) method described by Puillandre [56] was used. The ITS2 multiple sequence alignment (MSA) for the 34 *Physalis* sequences only was uploaded to the ABGD website (https://bioinfo.mnhn.fr/abi/public/abgd/abgdweb.html) and distance analysis was performed based on K80 Kimura measure of distance. The default value for relative gap width (X) was set at 1.5. Prior limits to intraspecific diversity (*P* values) of intraspecific divergence were set at prior minimum (P_{min}) and prior maximum (P_{max}) divergence of intraspecific diversity at 0.001 and 0.1 respectively. Default settings were maintained for all other parameters.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-024-05889-6.

Supplementary Material 1

Acknowledgements
I would like to thank the Department of Biochemistry and Center for Biotechnology and Bioinformatics (CEBIB), University of Nairobi for allowing us to use their research facilities for this work. The results reported in this study are part of PhD work for the first author.

Author contributions
K.P., K.M., E.K.M., J.M.W. and E.N.N. conceptualized and designed the study. K.P. performed the experiments, analyzed and interpreted data of the experiments. E.N.N. assisted in data analysis and interpretation. K.M., E.K.M., J.M.W. and E.N.N. supervised the research work. K.P. drafted the manuscript, which was revised by K.M., E.K.M., J.M.W. and E.N.N. All authors reviewed and approved the final manuscript.

Funding
This research received no external funding.

Data availability
The sequencing data are available at NCBI GenBank, accession numbers: OQ371996.1 - OQ372029.1.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
The authors agreed to the publication the article in the journal.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Biochemistry, Faculty of Science and Technology, University of Nairobi, P.O. Box 30197, Nairobi, Kenya
[2]Department of Life Sciences, South Eastern Kenya University, Kitui, Kenya
[3]Department of Biology, University of Nairobi, Nairobi, Kenya

Received: 21 July 2023 / Accepted: 28 November 2024
Published online: 18 December 2024

References

1. 1.Simbaqueba J, Sanchez P, Sanchez E, Nunez Zarantes VM, Chacon MI, Barrero LS, Mariño-Ramírez L. Development and characterization of microsatellite markers for the Cape gooseberry. Physalis peruviana. PLoS ONE. 2011;6(10):e26719.
2. Khan W, Bakht J. Development of efficient and optimized protocol for rapid micropropagation of *Physalis ixocarpa*, a medicinal herb. Pakistan J Weed Sci Res. 2015;21(3).
3. Popenoe H, King S, Leon J, Kalinowski L. Goldenberry (cape gooseberry). Lost Crops of the Incas: Little-known Plants of the Andes with Promise for Worldwide Cultivation, ed by National Research Council National Academy Press, Washington DC. 1990:241–252.
4. Yen CY, Chiu CC, Chang FR, Chen JY, Hwang CC, Hseu YC, Yang HL, Lee AY, Tsai MT, Guo ZL, Cheng YS. 4β-hydroxywithanolide E from *Physalis peruviana* (golden berry) inhibits growth of human lung cancer cells through DNA damage, apoptosis and G 2/M arrest. BMC Cancer. 2010;10:1–8.
5. Shenstone E, Lippman Z, Van Eck J. A review of nutritional properties and health benefits of *Physalis* species. Plant Foods Hum Nutr. 2020;75:316–25.
6. Golubkina NA, Kekina HG, Engalichev MR, Antoshkina MS, Nadezhkin SM, Caruso G. Yield, quality, antioxidants and mineral nutrients of *Physalis angulata L*. and *Physalis pubescens L*. fruits as affected by genotype under organic management. Adv Hortic Sci. 2018;32(4):541–8.
7. Petkova NT, Popova VT, Ivanova TA, Mazova NN, Panayotov ND, Stoyanova A. Nutritional composition of different cape gooseberry genotypes (*Physalis peruviana L*.)—A comparative study. Food Res. 2021;5:191–202.
8. Novoa RH, Bojacá M, Galvis JA, Fischer G. Fruit maturity and calyx drying influence post-harvest behavior of Cape gooseberry (*Physalis peruviana L*.) stored at 12 ºC. Agronomía Colombiana. 2006;24(1):77–86.
9. Shinwari ZK, Jan SA, Khalil AT, Khan AD, Ali M, Qaiser M, Zahra NB. Identification and phylogenetic analysis of selected medicinal plant species from Pakistan: DNA barcoding approach. Pak J Bot. 2018;50(2):553–60.
10. Feng S, Jiang M, Shi Y, Jiao K, Shen C, Lu J, Ying Q, Wang H. Application of the ribosomal DNA ITS2 region of *Physalis* (*Solanaceae*): DNA barcoding and phylogenetic study. Front Plant Sci. 2016;7:1047.
11. Menzel MY. The cytotaxonomy and genetics of Physalis. Proc Am Philos Soc. 1951;95(2):132–83.
12. Buckley DS, Zasada JC, II JC, Stone DM. Plant morphological characteristics as a tool in monitoring response to silvicultural activities. Communicating Role Silviculture Managing Natl Forests. 1997;37.
13. Muraguri PW, Gesimba RM, Wolukau JN, Miheso M. Genetic characterization of Cape gooseberry (*Physalis peruviana L*.) accessions in selected counties in Kenya using SSR markers. Afr J Plant Sci. 2021;15(1):28–38.
14. Chacón SMI, Sánchez YD, Barrero MLS. Genetic structure of a Colombian cape gooseberry (*Physalis peruviana L*.) collection by means of microsatellite markers. Agron Colombiana. 2016;34(1):5–16.
15. Hebert PD, Cywinska A, Ball SL, DeWaard JR. Biological identifications through DNA barcodes. Proceedings of the Royal Society of London. Series B: Biological Sciences. 2003;270(1512):313–21.
16. Barcaccia G, Lucchin M, Cassandro M. DNA barcoding as a molecular tool to track down mislabeling and food piracy. Diversity. 2015;8(1):2.
17. Elansary HO, Ashfaq M, Ali HM, Yessoufou K. The first initiative of DNA barcoding of ornamental plants from Egypt and potential applications in horticulture industry. PLoS ONE. 2017;12(2):e0172170.
18. Lahaye R, Van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, Maurin O, Duthoit S, Barraclough TG, Savolainen V. DNA barcoding the floras of biodiversity hotspots. Proc Natl Acad Sci. 2008;105(8):2923–8.
19. Chen S, Yao H, Han J, Liu C, Song J, Shi L, Zhu Y, Ma X, Gao T, Pang X, Luo K. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS ONE. 2010;5(1):e8613.
20. Yao H, Song J, Liu C, Luo K, Han J, Li Y, Pang X, Xu H, Zhu Y, Xiao P, Chen S. Use of ITS2 region as the universal DNA barcode for plants and animals. PLoS ONE. 2010;5(10):e13102.
21. Ralte L, Singh YT. Use of *rbcL* and ITS2 for DNA barcoding and identification of *Solanaceae* plants in hilly state of Mizoram, India. Res Crops. 2021;22(3):616–23.
22. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. 2014;42(1):320–4.
23. Vargas O, Martínez M, Dávila P. Two new species of *Physalis* (*Solanaceae*) endemic to Jalisco, Mexico. Brittonia. 2001:505–10.
24. Whitson M, Manos PS. Untangling *Physalis* (*Solanaceae*) from the Physaloids: a two-gene phylogeny of the Physalinae. Syst Bot. 2005;30(1):216–30.
25. Feng SG, Lu JJ, Gao L, Liu JJ, Wang HZ. Molecular phylogeny analysis and species identification of *Dendrobium* (*Orchidaceae*) in China. Biochem Genet. 2014;52:127–36.
26. Li Y, Zhou XI, Feng G, Hu H, Niu L, Hebert PD, Huang D. COI and ITS2 sequences delimit species, reveal cryptic taxa and host specificity of fig-associated *Sycophila* (*Hymenoptera*, *Eurytomidae*). Molecular Ecology Resources. 2010;10(1):31–40.
27. Pang X, Song J, Zhu Y, Xu H, Huang L, Chen S. Applying plant DNA barcodes for *Rosaceae* species identification. Cladistics. 2011;27(2):165–70.
28. Zhang W, Yang S, Zhao H, Huang L. Using the ITS2 sequence-structure as a DNA mini-barcode: a case study in authenticating the traditional medicine Fang Feng. Biochem Syst Ecol. 2016;69:188–94.
29. Mishra P, Kumar A, Sivaraman G, Shukla AK, Kaliamoorthy R, Slater A, Velusamy S. Character-based DNA barcoding for authentication and conservation of IUCN Red listed threatened species of genus *Decalepis* (*Apocynaceae*). Sci Rep. 2017a;7(1):14910.
30. Kang Y, Deng Z, Zang R, Long W. DNA barcoding analysis and phylogenetic relationships of tree species in tropical cloud forests. Sci Rep. 2017;7(1):1–9.
31. Zhang W, Yuan Y, Yang S, Huang J, Huang L. ITS2 secondary structure improves discrimination between medicinal Mu Tong species when using DNA barcoding. PLoS ONE. 2015;10(7):e0131185.
32. Huang XC, Ci XQ, Conran JG, Li J. Application of DNA barcodes in Asian tropical trees–a case study from Xishuangbanna Nature Reserve, Southwest China. PLoS ONE. 2015;10(6):e0129295.
33. Ross HA, Murugan S, Sibon Li WL. Testing the reliability of genetic methods of species identification via simulation. Syst Biol. 2008;57(2):216–30.
34. Li Q, Wu J, Wang Y, Lian X, Wu F, Zhou L, Huang Z, Zhu S. The phylogenetic analysis of *Dalbergia* (*Fabaceae: Papilionaceae*) based on different DNA barcodes. Holzforschung. 2017;71(12):939–49.

35. Chen SL. Traditional Chinese medicine molecular identification by DNA barcode. Beijing: People's Medical Publishing House. 2012:18:32–38.
36. Garzón-Martínez GA, Osorio-Guarín JA, Delgadillo-Durán P, Mayorga F, Enciso-Rodríguez FE, Landsman D, Mariño-Ramírez L, Barrero LS. Genetic diversity and population structure in *Physalis peruviana* and related taxa based on InDels and SNPs derived from COSII and IRG markers. Plant Gene. 2015;4:29–37.
37. Carlson CS, Thomas DJ, Eberle MA, Swanson JE, Livingston RJ, Rieder MJ, Nickerson DA. Genomic regions exhibiting positive selection identified from dense genotype data. Genome Res. 2005;15(11):1553–65.
38. Hajibabaei M, Singer GA, Hickey DA. Benchmarking DNA barcodes: an assessment using available primate sequences. Genome. 2006;49:851–4.
39. Queiroz Cde S, Batista FR, deOliveira LO. Evolution of the 5.8S nrDNA gene and internal transcribed spacers in *Carapichea ipecacuanha* (*Rubiaceae*) within a phylogeographic context. Mol Phylogenet Evol. 2011;59:293–302.
40. Pere K, Mburu K, Muge EK, Wagacha JM, Nyaboga EN. Molecular discrimination and phylogenetic relationships of *Physalis* species based on ITS2 and *rbcL* DNA barcode sequence. Crops. 2023;3:302–19.
41. Yao H, Song J. Use of ITS2 region as the universal DNA barcode for plants and animals. PLoS ONE. 2010;5, e13102.
42. Hall TA. BioEdit: a user-friendly biological sequence alignment editorand analysis program for Windows 95/98/NT. In Nucleic Acids Symposium Series 1999;41(41);95–8.
43. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–7.
44. Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61(3):539–42.
45. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 2003;19(12):1572–4.
46. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 2001;17(8):754–5.
47. Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc Natl Acad Sci. 2004;101(19):7287–92.
48. Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. The Vienna Websuite. Nucleic Acids Res. 2008;36(2):70–4. https://doi.org/10.1093/nar/gkn188
49. Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. ViennaRNA Package 2.0. Algorithms for Molecular Biology. 2011;6:1–4.
50. Kartavtsev YP. Divergence at Cyt-b and Co-1 mtDNA genes on different taxonomic levels and genetics of speciation in animals. Mitochondrial DNA. 2011;22(3):55–65.
51. Troshin PV, Procter JB, Barton GJ. Java bioinformatics analysis web services for multiple sequence alignment—JABAWS. MSA Bioinf. 2011;27(14):2001–2.
52. Troshin PV, Procter JB, Sherstnev A, Barton DL, Madeira F, Barton GJ. JABAWS 2.2 distributed web services for Bioinformatics: protein disorder, conservation and RNA secondary structure. Bioinformatics. 2018;34(11):1939–40.
53. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35:1547–9.
54. Tamura K, Stecher G, Kumar S. MEGA 11: molecular evolutionary genetics analysis version 11. Mol Biol Evol. 2021;38(7):3022–7.
55. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989;123(3):585–95.
56. Puillandre N, Lambert A, Brouillet S, Achaz GJ. ABGD, automatic barcode gap discovery for primary species delimitation. Mol Ecol. 2012;21(8):1864–77.

Publisher's note

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).
Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.
These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.
We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.
While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.
These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.
To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.
Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.
If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com