Cite this article as:
González Zamora, Alberto & Olvera, Clarita & Avalos-Calleros, Aaron & García-Hernández, José & Ríos-Sánchez, Efraín & Pérez Morales, Rebeca. (2020). Genetic variability of Pun1 gene (Capsaicin synthase) in pungent cultivars of Capsicum annuum of Northern Mexico. Tropical and Subtropical Agroecosystems. 23. 1-15.
https://www.revista.ccba.uady.mx/ojs/index.php/TSA/article/view/3002

GENETIC VARIABILITY OF *Pun1* GENE (CAPSAICIN SYNTHASE) IN PUNGENT CULTIVARS OF *Capsicum annuum* OF NORTHERN MEXICO †

[VARIABILIDAD GENETICA DEL GEN *Pun1* (CAPSAICIN SINTASA) IN CULTIVARES PUNGENTES DE *Capsicum annuum* DEL NORTE DE MEXICO]

Alberto González-Zamora[1*], Clarita Olvera-Carranza[2], Jesús Aaron Avalos-Calleros[3], José Luis García Hernández[4], Efraín Ríos-Sánchez[5] and Rebeca Pérez-Morales[5*].

[1] Laboratorio de Biología Evolutiva, Facultad de Ciencias Biológicas, Universidad Juárez del Estado de Durango, Av. Universidad s/n. Fracc. Filadelfia, C.P. 35010. Gómez Palacio, Durango., México. Email: agzfc@ujed.mx.
[2] Departamento de Ingeniería Celular y Biocatálisis, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Avenida Universidad 2001, Chamilpa, C.P. 62210. Cuernavaca, Morelos, México.
[3] Laboratorio de Biología Molecular de Plantas, Instituto Potosino de Investigación Científica y Tecnológica, Camino a la Presa San José 2055. Col. Lomas 4 sección. CP. 78216. San Luis Potosí, San Luis Potosí, México.
[4] Facultad de Agricultura y Zootecnia, Universidad Juárez del Estado de Durango, Km 35 Carretera Gómez Palacio-Tlahualilo, Ejido Venecia, CP 35111, Gómez Palacio, Durango, México.
[5] Laboratorio de Biología Celular y Molecular, Facultad de Ciencias Químicas, Universidad Juárez del Estado de Durango, Av. Artículo 123 s/n. Fracc. Filadelfia, C.P. 35010. Gómez Palacio, Durango., México. Email: rebecapms@ujed.mx.
**Corresponding authors*

SUMMARY

Background: Capsaicinoids are exclusive of the genus *Capsicum* and are responsible for fruit pungency. The genetic basis of pungency is not yet fully understood, although several candidate genes have been analyzed including the Pungent gene (*Pun1*). **Objective:** The main aim of this study was to identify the *Pun1* in some pungent samples of *Capsicum annuum* of Northern Mexico adapted to conditions of high temperature and analyze the variability of polymorphisms in the gene. **Methodology:** The coding sequence of the *Pun1* gene was analyzed by Polymerase Chain Reaction and Restriction Fragment Length. **Results:** The *Pun1* gene showed highly variable patterns; the variable regions were sequenced, and we found 19 single nucleotide polymorphisms in the first exon and 7 in the intron, some polymorphisms generated amino acid changes; while that of Chiltepin cultivar showed more than one sequence. **Implications:** The protein homology model suggests that these polymorphisms could affect their functionality. **Conclusion:** All the pungent cultivars of Mexican *C. annuum* analyzed were positive to *Pun1* gene and have a high variety of polymorphisms, this study provides evidence of the high variability that exists in the *Pun1* gene responsible for pungency and could allow the selection of pungent variants to obtain a greater quantity of capsaicinoids for different applications.

Keywords: Bioinformatics; capsaicinoids; chili pepper; genetic variability; pungency; single nucleotide polymorphisms.

RESUMEN

Antecedentes: Los capsaicinoides son exclusivos del género *Capsicum* y son responsables de la pungencia del fruto. Las bases genéticas de la pungencia no han sido comprendidas completamente, aunque se han analizado varios genes candidatos, incluido el gen de la pungencia (*Pun1*). **Objetivo:** El objetivo principal de este estudio fue identificar el gen *Pun1* en algunas muestras pungentes de *Capsicum annuum* del norte de México, adaptadas a condiciones de alta temperatura; y analizar la variabilidad de los polimorfismos en el gen. **Metodología:** La región codificante del gen se analizó mediante reacción en cadena de la polimerasa y los polimorfismos por análisis específicos con enzimas de restricción.**Resultados:** El gen *Pun1* mostró patrones muy variables; las regiones variables fueron secuenciadas y encontramos 19 polimorfismos de un solo nucleótido en el primer exón y 7 en el intrón, algunos polimorfismos

† *Submitted October 1, 2019 – Accepted January 10, 2020.*
ISSN: 1870-0462.

generaron cambios de aminoácidos; mientras que la del cultivar Chiltepin mostró más de una secuencia. **Implicaciones:** El modelo de homología de proteínas sugiere que estos polimorfismos podrían afectar su funcionalidad. **Conclusión:** Todos los cultivares pungentes de *C. annuum* analizados en este estudio fueron positivos para el gen *Pun1* y presentaron una gran variedad de polimorfismos. Este estudio proporciona evidencia de la alta variabilidad que existe en el gen *Pun1* responsable de la pungencia y podría permitir la selección de variantes con alta pungencia para obtener mayor cantidad de capsaicinoides para distintas aplicaciones.
Palabras clave: Bioinformática; capsaicinoides; chile; variabilidad genética; pungencia; polimorfismos de un solo nucleótido.

INTRODUCTION

Chili is the fruit of plants belonging to the genus *Capsicum* L., consisting of more than 200 varieties grouped in about 30 species, of which five are domesticated: *C. annuum* L., *C. baccatum* L., *C. chinense* Jacq. *C. frutescens* L., and *C. pubescens* Ruiz & Pav. (Hernández *et al*., 1999). Among them, *C. annuum* is the best known and most economically important species, since it is distributed worldwide and is widely used as an additive in the food and pharmaceutical industry (Tremblay *et al*., 2016).

Capsicum fruits are an excellent source of chemical compounds with beneficial effects on health, such as ascorbic acid (vitamin C), carotenoids (provitamin A), tocopherols (vitamin E), flavonoids, and capsaicinoids (Langhans, 2018). The capsaicinoids are a group of spicy chemical analogues with concentrations that depend on genotype, fruit maturity, and growing conditions. The most important capsaicinoids, considering abundance in the fruits, are capsaicin and dihydrocapsaicin (Popelka *et al*., 2017). Capsaicinoids are responsible for the hot taste of chili peppers. They are restricted to the genus *Capsicum* and are synthesized by the acylation of the aromatic compound vanillylamine, with a branched-chain fatty acid by catalysis of the putative enzyme capsaicin synthase (CS) (Arce-Rodríguez and Ochoa-Alejo, 2017). However, the genes involved in the capsaicin biosynthesis pathway have not been well characterized.

There are several studies of massive sequencing and de novo assemblages of *C. annuum* genomes in which genes and Single Nucleotide Polymorphism (SNP) associated with pungency have been found. In a Genome-wide Association Study (GWAS) the influence of multiple chromosomal regions that are associated with the content of capsaicinoids was demonstrated, including *Pun1* gene located on chromosome 2 and another acyltransferase located on chromosome 11 (Nimmakayala *et al*., 2016). In another GWAS 69 Quantitative Traits Loci (QTLs) were identified, which included candidate genes responsible for the pungency that had already been previously reported (Han *et al*., 2018). While Hulse-Kemp *et al*., (2018) reported a de novo assembly of 3.5-Gb of the genome of *C. annuum* where the validation showed insertions / deletions in the *Pun1* haplotypes, evidencing the high variability of genomic arrangements corresponding to this locus.

Another widely used strategy is the fine analysis of candidate genes that have been suggested as responsible for pungency, including the putative aminotransferase (pAMT) and Pun1 protein which are suggested to catalyze the second to last and last steps in the pathway, respectively, therefore is suggested as the putative CS. However, there is no direct evidence that Pun1 possesses capsaicin synthase activity, due to the difficulty of obtaining the necessary precursors for the enzymatic reaction (Ogawa *et al*., 2015).

Among the most important candidate genes that regulate the metabolic pathway of capsaicinoid synthesis, genes that code for phenylalanine ammonia-lyase (PAL), cinnamate 4-hydroxylase (C4H), 4-coumaryl CoA ligase (4CL), hydroxy cinnamoyl transferase enzyme (HCT), coumaryl shikimate/quinate 3-hydroxylase (C3H), caffeoyl-CoA 3-O-methyltransferase (CCoAOMT), caffeic acid O-methyltransferase (COMT), hydroxy-cinnamoyl-CoA hydratase/lyase (HCHL), ketoacyl ACP synthase (KAS), acyl carrier protein (ACL), acyl-ACP thioesterase (FAT), and capsaicin synthase (CS, Pun1) have been reported. In addition, the peroxidase (POD), involved in the degradation of capsaicinoids in the mature fruits, is also an important candidate gene (Kim *et al*., 2014). The last step of the metabolic pathway is determined by CS, the acyltransferase encoded by the *Pun1* gene and catalyzes the bond between fatty acids and vanillylamine to generate capsaicinoids (Zhang *et al*., 2016).

Great interest in acyltransferase CS has been generated since Stewart *et al*., (2005) reported a deletion of approximately 2.5 kb in the *Pun1* gene, and this genotype correlates with a lack of pungency of some cultivars of Capsicum, suggesting that the deletion affects the functionality of the enzyme; later they reported the genetic control of pungency mediated by *Pun1* locus (Stewart *et al*., 2007); while another study showed that mutations in a single locus, *Pun1*, is responsible for the loss of pungency (Stellari *et al.*, 2010).

In a fine study of candidate gene, Reddy *et al.*, (2014) analyzed genes encoding the enzymes CS (*Pun1*), CCR, KAS, and HCT, and found that the mRNA levels of these enzymes correlated with the level of capsaicinoids in the fruits, suggesting that these genes are responsible for the quantitative characteristics of the pungency of the genus *Capsicum*. In the analysis of DNA sequences of *Pun1* gene, 36 single SNPs were described: 19 in the promoter region, 7 in the first exon, 7 in the intron, and 3 in the second exon. Moreover, six SNPs showed high correlation with the level of pungency: 2 in the promoter region and 4 in the coding region where 3 corresponded to non-synonymous amino acid changes, this study suggests that the presence of SNPs in the gene could affect the level of pungency, possibly due to amino acid changes in important domains that affect the activity of the CS protein.

The mechanism by which the level of pungency of *Capsicum* fruits is determined is not entirely clear; however, there is sufficient evidence to believe that CS is a key enzyme in the biosynthesis of capsaicinoids; thus, the genetic variability of CS could determine the level of pungency; additionally for many crop species, identification and utilization of SNPs have become economical because of the availability and abundance of various high-throughput technologies that allow the screening of a large number of samples. However, it is necessary to have studies that carry out a detailed analysis of the variability of the gene of interest, the levels of the desired metabolite and the growth conditions of the crop. In this study, different pungent cultivars of *C. annuum* produced and marketed in Northern Mexico were analyzed, in order to determine the genetic variability associated with SNPs in the *Pun1* gene and its possible relationship with the level of pungency.

MATERIALS AND METHODS

Plants and samples

Eighteen cultivars of *C. annuum* originating from different states of Northern Mexico and Southern United States were grown. These are traditionally used in the region of La Comarca Lagunera, a region in North-Central Mexico. A Styrofoam tray, with 200 wells disinfected with chlorinated bleach 0.05%, was used for seed germination; peat moss and vermiculite were added in a ratio of 2:1 (v/v) as substrate. For each cultivar, 10 seeds were used. One seed was planted in each well at a depth of 0.5 cm, and the tray was covered with plastic wrap for 72 h to ensure uniform germination. Seedlings were transplanted 60 days after germination at the greenhouse of the Faculty of Agriculture and Zootechnics of the Juárez University of the State of Durango. Samples of leaves of 3 – 5 individuals of each cultivar were collected 4 months after transplantation to the greenhouse; and samples were washed with distilled water to be used for DNA extraction.

DNA extraction

DNA extraction was performed using the protocol reported by Dellaporta *et al*., (1983) with modifications. High molecular weight DNA was identified by electrophoresis on a 0.8% agarose gel stained with 10 µg/ml Gel Red Nucleic Acid Gel Stain (Biotium). The genomic DNA was visualized in a Pro MiniBis photodocumentor (DNR Bio Imaging Systems). The quantification and purity were determined by the ratio of absorbance at 260/280 nm in a Nano Drop spectrophotometer (Thermo Scientific ®). DNA samples were stored at -80 °C until use.

Polymerase Chain Reaction (PCR) and Restriction Fragment Length Polymorphism (RFLP) analyses

We accessed the sequence AY819028.1 of *C. annuum* cultivar Hot 1493 acyltransferase *Pun1* gene, deposited in GenBank of the National Center for Biotechnology Information (NCBI), in order to design the following oligonucleotides: *Pun1* Fwd 5'-CCG GCC AGC AGC ATA TAT TTA-3' and *Pun1* Rev 5'-TTG TTG ACC GTA AAC TTC CGT TG-3'. These amplify a fragment of ~ 1834 bp, corresponding to the *Pun1* gene. The PCR mixture consisted of 3 µL of $MgCl_2$ (1.5 mM), 10 µL buffer (polymerase), 0.25 µL of GoTaq® DNA Polymerase (Promega), 3 µL of dNTP's (2mM), 2.5 µL of forward and reverse primers (5 µM), ~ 100 - 200 ng of template DNA, and molecular biology grade water was added to reach a final volume of 50 µL. PCR conditions were as follows: 1 cycle at 94 °C for 5 min, followed by 35 cycles at 94 °C for 45 s, 60 °C for 60 s, and 72 °C for 2 min; the final extension was at 72 °C for 5 min.

An *in-silico* restriction analysis was performed through the NEBcutter V2 server (http://nc2.neb.com/NEBcutter2/) on the theoretical 1834 bp fragment, and sites for the enzymes Alu I, Bsr I, Hinf I, Mbo I, Mbo II, Rsa I, and Sap I were detected. The experimental conditions of the restriction analyses were the following: 1.5 µL buffer (specific enzyme), 0.1 – 1 µL according to the restriction endonuclease, 4 µL of PCR product, and molecular biology grade water to reach a final volume of 15 µL. All the restriction enzymes were Cut Smart® from New England Biolabs. The PCR and restriction products were analyzed by agarose gel electrophoresis (1% for PCR product and 2% for RFLPs analyses). High DNA Mass Ladder or Low DNA Mass Ladder (2 µL) (Thermo Fisher Scientific) was used as a control of molecular weight. The staining was performed with Gel Red

Nucleic Acid Gel Stain (Biotium), and bands were visualized in a Pro MiniBis photodocumentor (DNR Bio Imaging Systems). Additionally, the oligonucleotides AGZ Fwd 5'-GGT GAT GGT TGC TCT CTG CTT-3' and AGZ Rev 5'-GCT AAC AAC TTC AGC CCT TGT TGG-3' were designed to amplify a ~ 625 bp fragment, because in this region the highest variability of RFLP on *Pun1* gene was observed and it contains the intronic region. The PCR mixture consisted of 4 µL buffer (polymerase), 0.1 µL of GoTaq® DNA Polymerase (Promega), 1.2 µL of $MgCl_2$ (1.5 mM), 1 µL of dNTP's (2 mM), 0.5 µL (5 µM) of forward and reverse primers, and ~ 200 ng of template DNA, and molecular biology grade water was added to reach a final volume of 20 µL. PCR conditions were as follows: 1 cycle at 94 °C for 3 min, followed by 35 cycles at 94 °C for 30 s, 60 °C for 30 s, and 72 °C for 60 s. The final extension was at 72 °C for 3 min. On this amplicon, the RFLP analysis was performed again only with the enzymes Alu I, Bsr I, and Hinf I. The assignment of genotypes was performed according to the banding pattern present in each sample, depending on the restriction endonuclease used.

DNA sequencing and alignments

RFLP patterns representative of each cultivar were selected, and 10 cultivars were sequenced on the highly variable coding region and the intronic region. The purified PCR products were sequenced in the Sequencing Unit of the Institute of Biotechnology of the National Autonomous University of Mexico. The sequences were analyzed in BioEdit v7.0.5. The identity of the sequences was analyzed by nucleotide BLAST (National Center for Biotechnology Information (https://blast.ncbi.nlm.nih.gov/Blast.cgi)); and BLASTx analyses were performed to visualize the sequences of proteins generated by the DNA sequences. Multiple alignments of the DNA and protein sequences were performed using the Clustal IX software.

Homology model of protein

A homology model of the acyltransferase Pun1 (AY819028.1) was constructed, through the I-tasser server (https://zhanglab.ccmb.med.umich.edu). Vinorine synthase (VS) was used as a template for modeling the structure Pun1 (CS) protein.

Analysis of similarity

For the similarity analysis, DNA sequences and protein sequences were analyzed using the Mega 7.1 software. Dendrograms were then constructed to cluster groups by their similarity and clustering was inferred using the UPGMA algorithm. Distances were computed using the p-distance method. A bootstrap test with 500 replicates was performed. All positions containing gaps and missing data were eliminated.

RESULTS

***Capsicum annuum* cultivars and capsaicinoids content**

Eighteen cultivars of *C. annuum*, corresponding to the varieties Ancho, Arbol, Guajillo, Chiltepin, Jalapeño, Piquin, Puya, and Serrano were included in the present study. For Ancho, Chiltepin, Jalapeño and Serrano several cultivars were included, most from Sinaloa and Sonora, Mexico; while cultivar of Guajillo and Chiltepin were from Baja California Sur, Mexico; and Puya was from Durango, Mexico; finally, the Early Jalapeño and Piquin NuMex were from New Mexico, USA. The details are shown in the Table 1. The levels of capsaicin and total capsaicinoids were previously reported in Gonzalez-Zamora *et al*., (2013); González-Zamora *et al*., (2015); and it was observed that the amount of metabolites present in each cultivar is very different. The cultivars classified with low level of pungency were Ancho, Guajillo, Piquin, Puya, Serrano and all the varieties of Jalapeño, the cultivars with middle level of pungency were Arbol and the Chiltepin BCS; while those classified as high pungency were the Chiltepin cultivars from Sonora.

Variability in restriction patterns of *Capsicum annuum* cultivars

First, a restriction analysis *in silico* was performed and showed cleavage sites for the endonucleases AluI, BsrI, HinfI, MboI, MboII, RsaI and SapI (Figure 1a), theoretically, these sites could be identified in the *Pun1* gene, the loss or gain of restriction sites is a source of genetic variability that can be analyzed by RFLP. Therefore, the presence of the *Pun1* gene (~ 1834 bp) was analyzed in the samples and a product of the expected size was amplified, so all the individuals analyzed were positive for the presence of *Pun1*, as expected (Figure 1b). The experimental restriction assays were performed, and we found that the endonucleases MboI, MboII, RsaI, and SapI did not generate informative patterns, since no changes were observed in the restriction patterns among the different samples of the cultivars (Figure 1c).

However, the patterns generated by the endonucleases AluI, BsrI, and HinfI showed high variability among the cultivars; a detailed analysis of the polymorphic region was performed because this region contain the intronic region and the restriction assays were re-analyzed.

Table 1. Varieties and cultivars of *Capsicum annuum* included in this study, its origin and capsaicinoids content.

Variety	Cultivar	Number of plants analyzed	Origin	Capsaicin content (mg g^{-1} dry weight)	Total capsaicinoids (mg g^{-1} dry weight)
Ancho	Don Matias (DM)	5	Sinaloa, Mexico	0.29 ± 0.03[a]	1.88 ± 0.05[a]
	Mulato (ML)	5	Sinaloa, Mexico	0.08 ± 0.00	0.41 ± 0.01
Arbol	Cola de rata (CR)	4	Sonora, Mexico	20.87 ± 0.21	46.11 ± 0.23
Chiltepin	Ch2	3	Sonora, Mexico	56.68 ± 1.49[b]	118.93 ± 3.47[b]
	Ch3	3	Sonora, Mexico	47.53 ± 3.54[b]	89.77 ± 2.57[b]
	BCS	3	Baja California Sur, Mexico	15.36 ± 1.67[a]	31.84 ± 0.75[a]
Guajillo	Don Ramón	4	Sinaloa, Mexico	1.12 ± 0.05	4.61 ± 0.08
Jalapeño	Autlan (AU)	5	Sinaloa, Mexico	1.48 ± 0.04	5.85 ± 0.10
	BCS	3	Baja California Sur, Mexico	2.83 ± 0.05	11.46 ± 0.11
	Don Julio (DJ)	3	Sinaloa, Mexico	8.03 ± 0.37[a]	21.03 ± 0.27[a]
	Early Jalapeño	3	New Mexico, USA	3.20 ± 0.06	12.76 ± 0.17
	Forajido	3	Sinaloa, Mexico	0.93 ± 0.08	4.55 ± 0.44
	Euforia	3	Sinaloa, Mexico	3.36 ± 0.22	8.37 ± 0.57
	Najal	3	Sinaloa, Mexico	9.86 ± 0.55	29.51 ± 0.31
Piquin	NuMex	4	New Mexico, USA	1.87 ± 0.04	6.87 ± 0.10
Puya	Jacobo Robles	4	Durango, Mexico	1.18 ± 0.07[a]	4.80 ± 0.08[a]
Serrano	Don Diego (DD)	3	Sinaloa, Mexico	1.52 ± 0.01[a]	6.25 ± 0.04[a]
	Don Vicente	4	Sinaloa, Mexico	7.90 ± 0.02	27.80 ± 0.04

[a] Data reported in González-Zamora *et al*., (2013); [b] Data reported in González-Zamora et al., (2015).

The endonuclease AluI generated two patterns that were as follows: a) 2 bands: 300 and 250 bp; b) 2 bands: 400 and 300 bp; while BsrI generated three patterns: a) 625, 460 and 165 bp; b) 460 and 165 bp; c) 165 bp), finally HinfI generated two patterns: a) 400, 100 and 82 bp; b) 400, 200, 100, 82 bp; in the figure 2C only pattern a is observed.

The results were consistent among plants of the same cultivar; as they showed the same pattern at each restriction site. Regarding the frequency of genotypes (a, b or c, according to the enzyme used), it was found that for the AluI sites, the frequencies were a = 0.74 and b = 0.26, for BsrI, a = 0.64, b = 0.32, and c = 0.04, whereas HinfI were a = 0.56 and b = 0.44. These genotypic frequencies indicate that they are common polymorphic patterns among the cultures of *C. annuum* of Northern Mexico.

Variability in DNA sequence in cultivars of *Capsicum annuum*

According to the previous RFLP analyses, the samples with the greatest variability in the restriction sites AluI, BsrI and HinfI were selected, in order to represent all the genotypes of the cultivars analyzed. Ten samples from 8 cultivars were sequenced, obtaining reliable data for 508 nucleotides, these sequences were subjected to a multiple alignment in which 19 SNPs were observed in the region corresponding to the first exon and 7 SNPs in the region of the intron, as shown in Figure 3; the accession AY819028.1 deposited in Gene Bank (NCBI), and corresponding to *C. annuum* cultivar Hot 1493 was used as a comparison control.

By analyzing the cut-sites for the restriction enzymes used in the RFLP assay, variable regions were

identified. HinfI enzyme cuts at the 5' G ↓ ANTC 3' position and the C → T change was observed in the sequences (T allele lost the restriction site); AluI recognizes the site 5' AG ↓ CT 3', this pattern was highly variable in the RFLPs and two restriction sites were detected in the sequences: in the first one, all of the samples presented the same cut site for Alu I and in the second one, two changes were observed A → G and G → A that modified the restriction pattern. In both cases, the second position is the one that generated the loss of the restriction site. Finally, the BsrI enzyme recognized the 5' ACTGGN ↓ 3' site and in the sequences the G → C change was observed (C allele lost the restriction site).

A BLASTx analysis was carried out from the DNA sequences in order to know if SNPs generated changes at the protein level; the amino acid changes were in the positions 207, 211, 220, 227, and 228 of the protein. The cultivars included in this study differed with respect to the sequence AY819028.1 which was used as a comparison parameter, as shown in Figure 4.

SNPs that do not generate changes at the level of amino acids have been reported, so DNA sequences were analyzed in detail in order to detect SNPs that generated changes in the protein sequence. Changes in DNA that affected the amino acids 206 and 207 were observed; in the first one, the SNP generated a synonymous change for Ser → Ser (this site was recognized by the enzyme HinfI); in amino acid 207, an Ile → Val change was observed, but it was not detected with restriction enzymes. The change Gln → Glu in the amino acid 227 was detected by the enzyme BsrI. In addition, the AluI cut site was detected in exon 1, although it did not generate changes in the amino acids sequence. On the other hand, heterozygous sites were observed, generating a third restriction pattern (as was observed in the RFLPs with BsrI). Finally, the sequence obtained from the cultivar Chiltepin was not included in the alignments because the presence of more than one amplification product was observed (Figure 5).

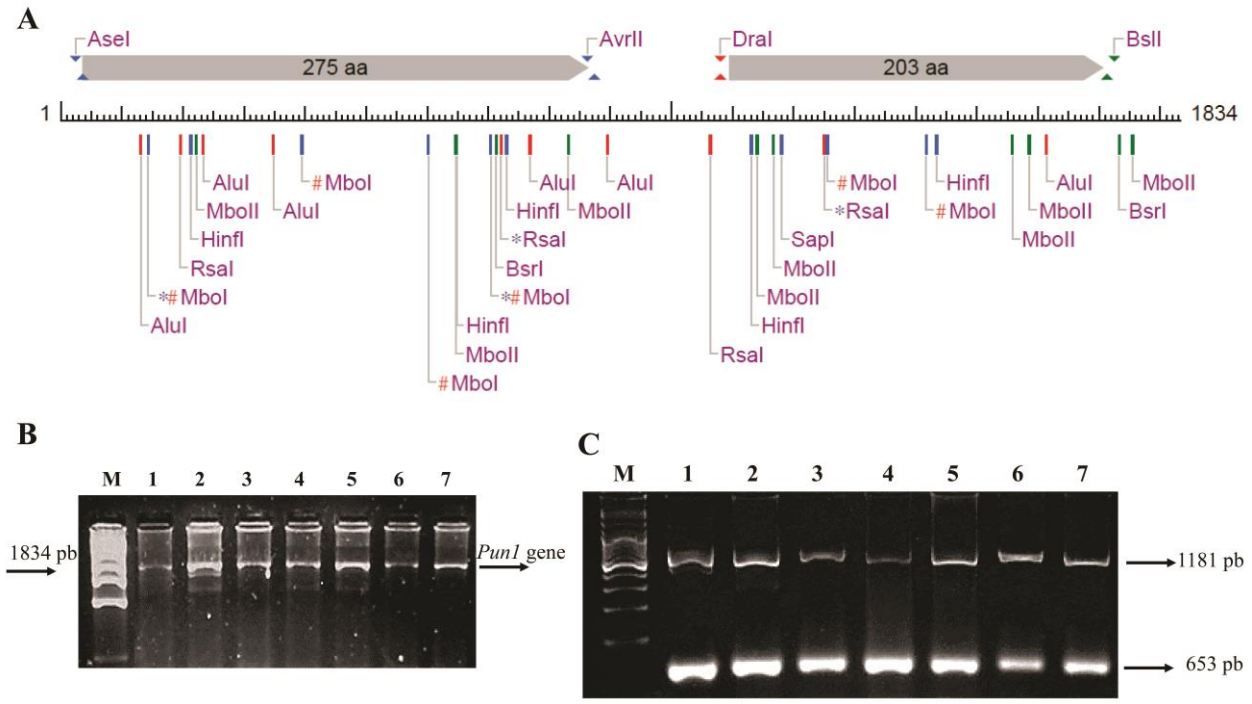

Figure 1. *Pun1* gene analysis. A) *In silico* analysis of RFLPs on *Pun1* gene. B) Amplification product corresponding to *Pun1* gene (1834 pb). C) Representative no variable patterns identified by RFLPs with SapI enzyme. In the electrophoresis lane M corresponds to molecular weight ladder and lanes 1-7 are the samples.

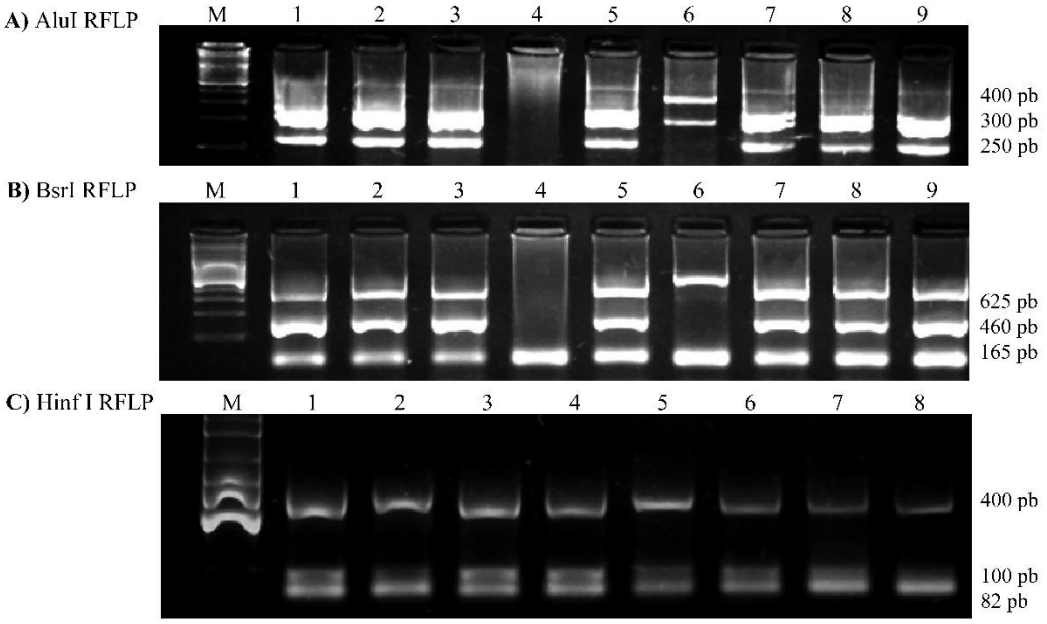

Figure 2. Restriction Fragment Length Polymorphism on *Pun1* gene. In all electrophoresis lane M corresponds to molecular weight ladder. **A**) RFLP with AluI enzyme: lanes 1-3, 5, 7-9 are restriction pattern a, while lane 6 is restriction pattern b. **B**) RFLP with BsrI enzyme: lanes 1-3, 5, 7-9 are restriction pattern a, lane 6 corresponds to restriction pattern b, and 4 lane correspond to restriction pattern c. **C**) RFLP with HinfI enzyme where all lanes shown are restriction pattern a.

Similarity analyses among the cultivars of *Capsicum annuum*

Similarity analyses were performed based on 10 nucleotide sequences and 10 protein sequences, respectively, to determine the cultivars that were more similar. Accession AY819028.1 was used as a comparison parameter, this accession sequence differs from the sequences of cultivars from Mexico and New Mexico; different relationships were found according to the analyzed sequence (Figure 6). At the DNA level, a matrix consisted of a total of 508 positions in the final dataset and two groups were observed, in the first one the cultivars Serrano, Ancho ML, and Puya were found, while in the second one, the other cultivars were sub-grouped.

The matrix of protein analysis had a total of 48 positions in the final dataset and two groups were observed. The first group included the cultivars Guajillo, Jalapeño DJ, Serrano DD, and Ancho ML, with a high level of similarity; in the second group, Jalapeño AU and Puya were found, while the cultivars Ancho DM, Piquin, and Arbol CR had a high level of similarity between them. The difference between the dendrograms was because in the DNA sequence, variability present in the intron is included, while at the protein level, only the variability of the SNPs that impact on the amino acid sequence are visualized. Within this variability, the non-synonymous changes could have a high potential to modify the functionality of the protein, so the non-synonymous changes were of great interest.

Homology model of Capsaicin synthase

Was constructed a homology model of the acyltransferase Pun1 (CS) (AY819028.1 obtaining a C-score, a TM score and RMSD of 0.57, 0.75 and 5.86 respectively (Figure 7A). These scores suggest that the model is confident. Vinorine synthase (VS) was used as a template for modeling the structure CS (cyan) despite the low sequence identity between CS and VS (31% of identity). Two-domain structure of acyltransferase was observed in CS that shown similarity with structures of VS. In fact, conserved residues typically present in the BAHD family, to which VS belongs, are found in both CS domains. One of them is the HXXXD169-173 motif (red) where it is found the catalytic residue His169 that functions as a general base during catalysis and is situated in the center of the reaction channel at the interface of both domains; and the DFGWG 384-388 motif (green) is far from the catalytic site, but it is indispensable for the catalysis and has importance for maintaining the integrity of the Coenzyme A binding pocket of all enzymes of the BAHD family. Due to all these results, we also suggest that CS is a member of the BAHD family, as it had been previously suggested.

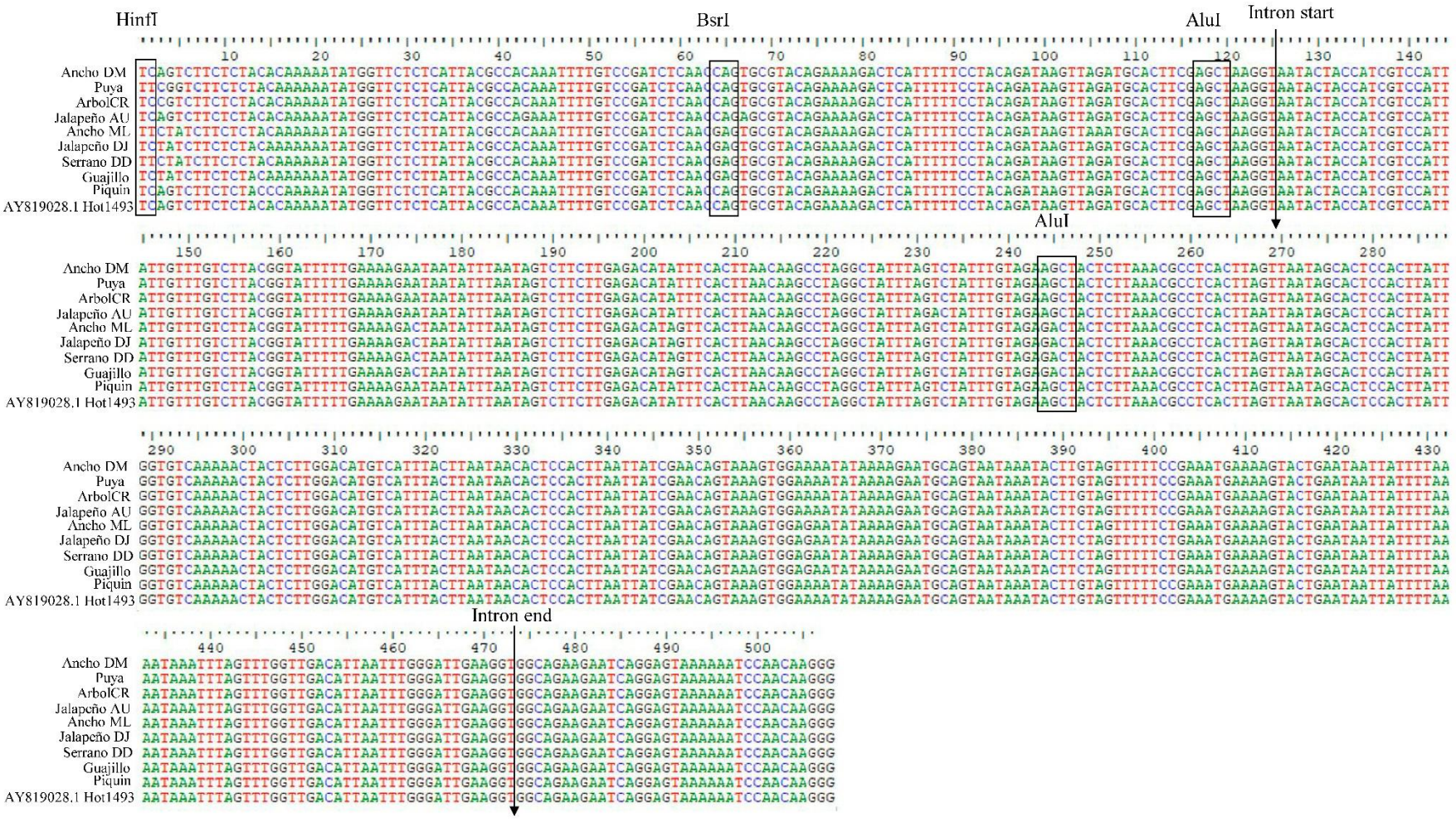

Figure 3. Alignment of DNA sequences in the variable region of the *Pun1* gene of *C. annuum*, the arrows indicate the start and end of the intron. Sites recognized by AluI, BsrI and HinfI restriction enzymes are indicated. The AY819028.1 accession was used as a comparison control.

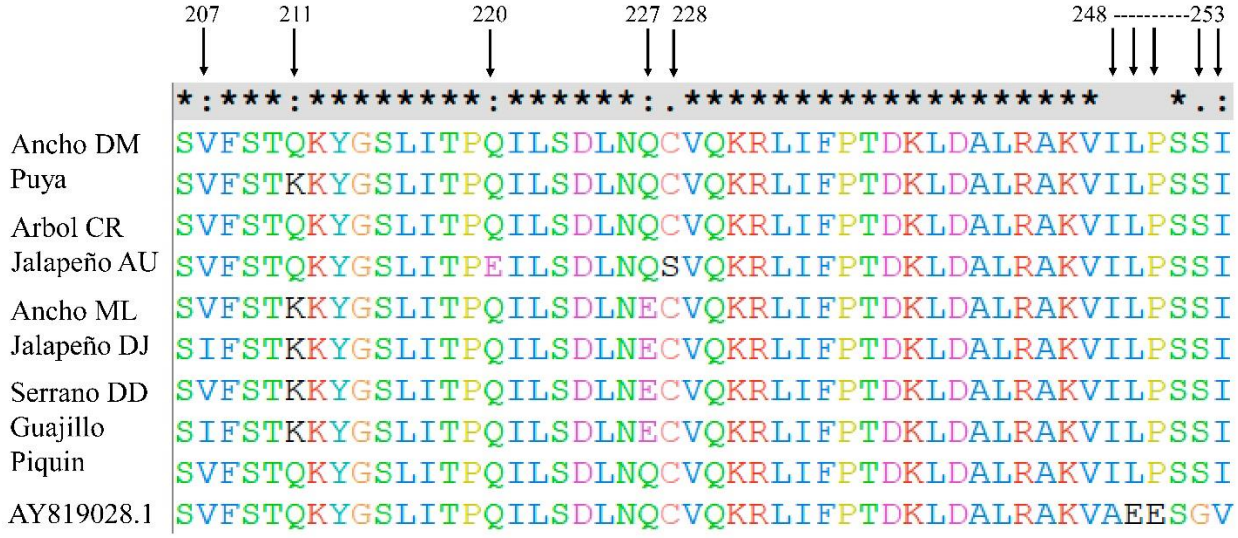

Figure 4. Alignment of DNA sequences of the *Pun1* gene on translated protein. Changes in the amino acids are shown. AY819028.1 accession was used as comparison control.

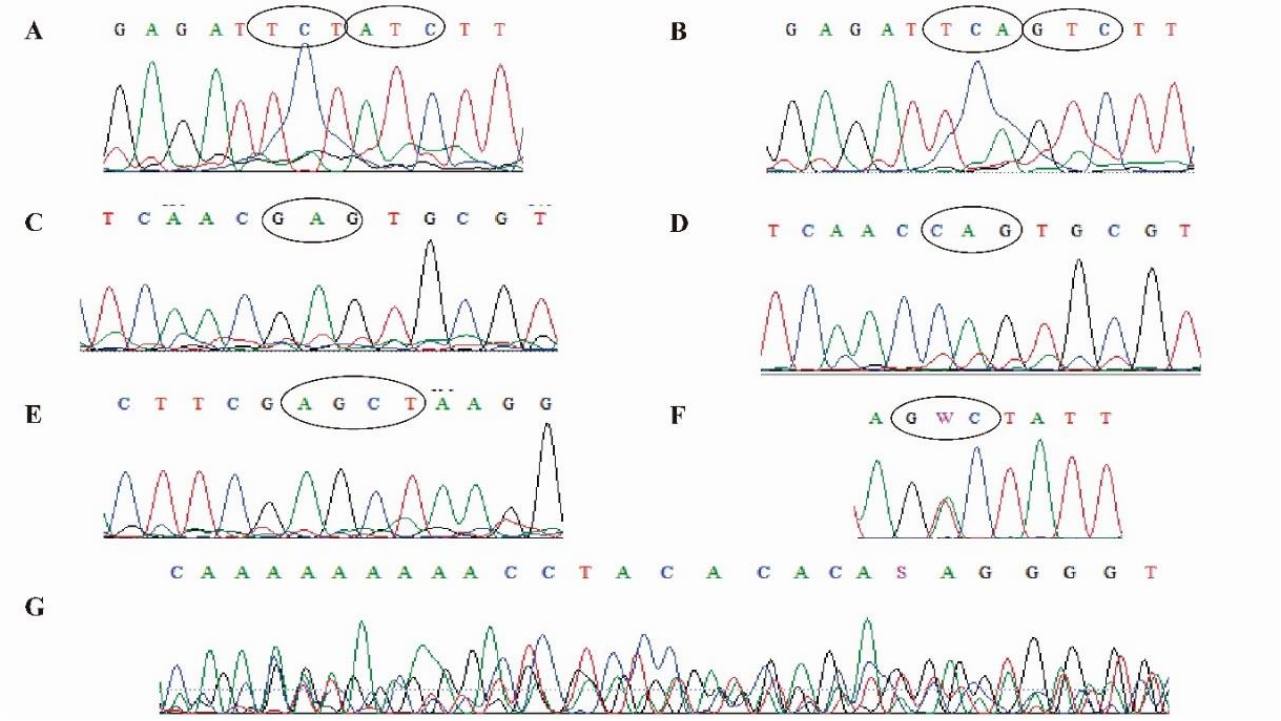

Figure 5. SNPs found on DNA sequence of the *Pun1* gene of *C. annuum*. Sequences in DNA that generate changes to protein are shown. (**A, B**) Codons for amino acids 206 (UCU (TCT) and UCA (TCA) coded to serine) and 207 (AUC (ATC) coded to isoleucine and GUC (GTC) coded to valine). (**C, D**) Codons for amino acid 227 (GAG coded to glutamic acid and CAG coded to glutamine). (**E**) Sequence of AluI site on first exon. (**F**) Representative heterozygosity in some samples. (**G**) Multiple products of sequencing in the Chiltepin sample.

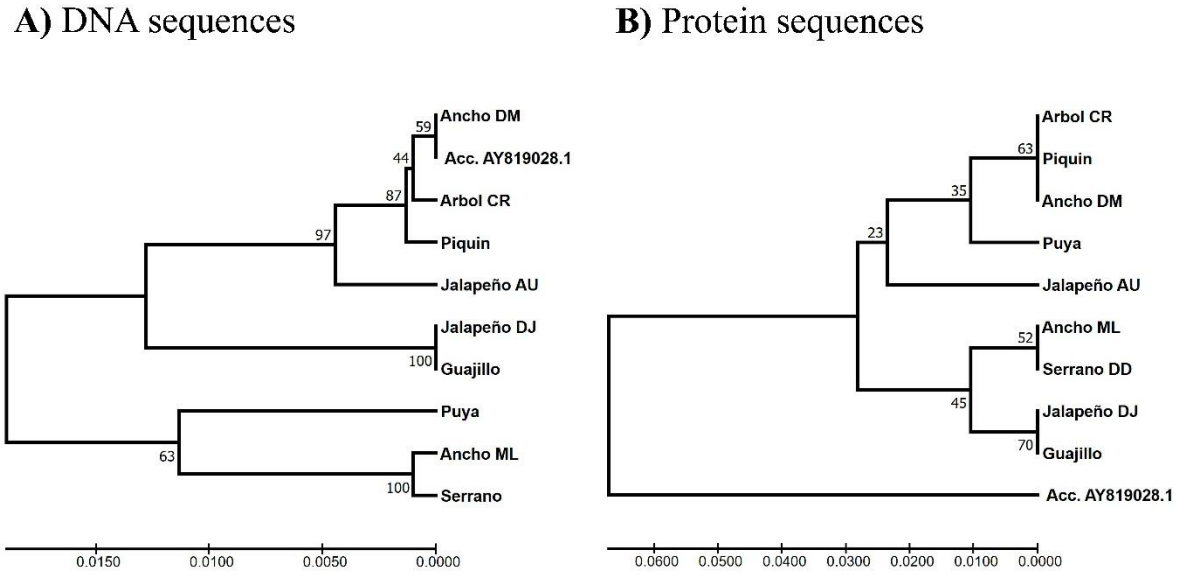

Figure 6. Similarity relationships among cultivars of Northern Mexico and New Mexico. (**A**) Based in DNA sequences. The optimal tree with the sum of branch length = 0.07 is shown. (**B**) Based on protein sequences. The optimal tree with the sum of branch length = 0.20 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test is shown next to the branches.

The variable region (yellow) found in the Mexican cultivars is indicated in the model of CS (cyan) (Figure 7B), it was observed that this region is part of the reaction channel entrance, close to the border that leads to the catalytic site of the enzyme; it could be suggested that some residues belonging to the variable region may affect the entrance of the substrate to the active sites influencing the enzyme activity. Moreover, based on the prediction of ligand binding sites from I-tasser, several residues (blue) could be involved in the substrate interaction, these residues could be able to interact with the variable residues (pink) since its arrangement aimed its lateral chains to the center of the catalytic channel; therefore, the changes in the residues interaction would affect its interaction with the substrate (Figure 7C).

DISCUSSION

Capsicum species are frequently described in terms of genetic diversity, considering morphological, agronomic, and molecular databases because *Capsicum spp*. plays an important economic role in many countries. The American continent is the center of origin of the genus *Capsicum* and Mexico has cultivars with unique characteristics, some of them are grown in extreme conditions of temperature and humidity, favoring a higher production of secondary metabolites such as capsaicinoids (Sarpras *et al*., 2016). Most studies have shown that there is great genetic variability among cultivars of *Capsicum*, probably because it is a cultivar that has been widely handled during thousands of years of domestication, although it has also been found high variability and genetic redundancy in semi-wild and wild populations (Kraft *et al*., 2014). Several studies have focused on whole genome variability through low-throughput technologies as AFLP, RAPDs, micro and mini-satellites, SSR, and RFLPs on genomic DNA, and recently high-throughput technologies with approaches in genomic, transcriptomic and metabolomic have been carried out in order to have reference data (genome and transcriptome) and understand the biosynthesis of capsaicinoids, even the evolution of pungency in the *Capsicum* genus.

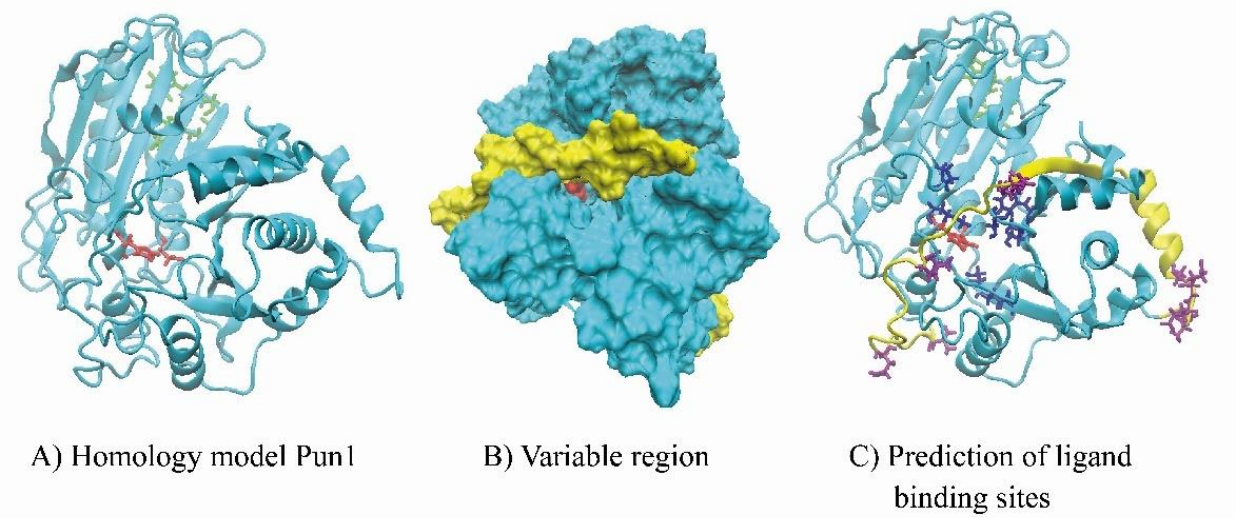

Figure 7. Homology model of Capsaicin synthase (CS) protein constructed using the I-tasser server. **A**). Model of the acyltransferase CS, the catalytic residues present in the HXXXD169-173 motif are shown in red, while the DFGWG 384-388 motif indispensable for the catalysis and maintaining of the integrity of the Coenzyme A binding pocket is shown in green. **B**) The variability region found in Mexican cultivars is indicated in yellow on the model of CS in cyan. The variability region is part of the reaction channel border that leads to the catalytic residues (red) of the enzyme. **C)** The prediction of ligand binding sites from I-tasser, the residues probably involved in the interaction with the substrate (blue) and the variable residues present inside of the variable region (pink) are close and could be interacting, the variable residues aim to the center of the catalytic channel.

It is widely documented that the pungency level of *Capsicum* fruits depends on environmental factors such as light, temperature, CO_2, altitude, availability of water and nutrients (Rezende-Naves *et al*., 2019). Although these factors can be controlled, it has been observed that the response to environmental conditions is very variable, depending on the genotype or cultivated variety, so that the final content of capsaicinoids can be highly variable, even under controlled conditions. In a previous study we observed that, in spite that all the varieties were cultivated in the same environmental conditions, the level of capsaicinoids was very variable among them, so this variability in pungency could be associated to the genotype, as it has been described in review reports (Arce-Rodríguez *et al*., 2019).

Also, there is sufficient experimental evidence showing that capsaicinoid biosynthesis is related to the expression of several candidate genes; Keyhaninejad *et al*., (2014) reported that levels of mRNA expression of Pal, Kas, At3 (Pun1), and FatA genes in the placenta of the fruit of *Capsicum* was directly related to the level of pungency, however the polymorphic variability of these genes that may affect mRNA expression levels and/or activity of the enzymes was not analyzed. In this regard, the high pungency levels might be due to the higher level of candidate gene(s) expression, but nucleotide variation in pungency related genes may also be involved in the variations in level of capsaicinoids, as has been suggested. Variations of SNP in the DNA sequence can have different implications; if it is found in a non-coding site it can have a neutral effect, affect the recognition site of promoters, other regulatory elements or alternative splicing sites; while the SNP in coding sequences can generate synonymous changes (encodes the same amino acid) or generate non-synonymous changes capable of altering the functionality of the protein.

In this study, we conducted a PCR-RFLP analysis on the *Pun1* gene and detected three restriction enzymes that generated highly variable patterns, as was observed in the figure 2. By carrying out a detailed analysis on the highly variable region and sequencing the most representative cultivars, as is observed in the figures 3 – 5, several SNPs were identified in the coding region and the region corresponding to intron. In this respect, the similarity found between the DNA sequences and the prediction of the amino acid sequence was distinct, grouping the cultivars differently (Figure 6). Using the DNA sequences, the cultivar Arbol, with medium level of pungency, is grouped near the reference used as a control, while when using the protein sequences they are grouped differently, at the base we find the reference sequence and at the other end the cultivar Arbol, all the other cultivars with low level of pungency were grouped between these two. The above shows that there is variable similarity at the level of DNA and proteins sequences, in addition this variability in DNA reflects the presence of the intron which is a non-translated region.

The SNPs found in the coding region and generate amino acid change could alter the functionality and/or interaction of the CS enzyme, affecting the synthesis of capsaicinoids. In this respect, Reddy *et al*., (2014) reported a high correlation between the presence of polymorphisms and the level of pungency, suggesting that the presence of SNPs may modulate the activity of the enzyme. The authors reported 36 SNPs: 19 in the promoter region, 7 in the first exon, 7 in the intron and 3 in the second exon; regarding the polymorphisms in the coding region, they found 6 non-synonymous changes at the protein level and 4 synonymous changes. In this study we found 26 SNPs; 19 in the first exon and 7 in the intron; the SNPs that showed amino acid changes in the cultivars of Mexico and New Mexico were Leu206Ser, Val207Ile, Gln211Lys, Gln220Glu, Gln227Glu, and Cys228Ser, of which 206, 207, 211, and 227 had already been reported previously, while 220 and 228 that were found in our cultivars had not been reported.

Although experimental studies on CS are limited due to the impossibility of performing *in vitro* synthesis, there are studies in other species where it has been possible to analyze the importance of polymorphisms in the expression of genes and the functionality of proteins (Wen *et al*., 2017). In this respect, it is possible that SNPs in the intronic region can affect the level of mRNA splicing, resulting in variability in the synthesis of capsaicinoids, therefore, in the pungency of the fruits. Meanwhile, the non-synonymous SNPs are of great interest, since they could modify the functionality of the enzyme, for example the change Gln → Lys, where the glutamine is a polar amino acid without charge and the lysine has a positive charge, or the change Gln → Glu, where the glutamate has a negative charge, the sites can be observed in the figure 7. These amino acid changes with different polarity and charge could generate structural alterations that affect the protein and, therefore, its efficiency in the synthesis of capsaicinoids. The homology model using the VS showed that the variable region found in CS is part of the reaction channel entrance, close to the border that leads to the catalytic site of the enzyme, and it could be suggested that some residues belonging to the variable region affect the entrance of the substrate to the active sites influencing the enzyme activity, and that changes in the residues interaction would affect its interaction with the substrate. The capsaicin biosynthesis pathway has not been well characterized, although its metabolites are of great importance for different industrial applications the substrate specificity and enzymatic function of CS in the capsaicinoid biosynthetic pathway remains unknown, but all evidence to date indicates that this enzyme is essential for capsaicinoid biosynthesis. Although, there is no available data of another protein that has greater homology with the CS, we believe that possibly the amino acid changes in the variable region can impact the CS enzyme activity, however, the correlation of these modifications with pungency level is difficult because protein function is unknown.

On the other hand, several studies have analyzed the level of pungency in the different cultivars of *Capsicum* and its relationship with genetic and environmental characteristics. In our previous study we reported the pungency level of the sequenced cultivars and found that the cultivars Ancho, Puya, Guajillo, Jalapeño and Serrano had a low level of pungency (determined by the level of capsaicinoids); Arbol and Chiltepin from Baja California, Mexico, had a medium pungency level; while Chiltepin from Sonora had a high level of pungency. In the DNA sequencing, Chiltepin showed several products which could indicate that this cultivar has more copies of the *Pun1* gene (Figure 5G). It has already been suggested that some cultivars may have multiple copies of genes related to pungency. Qin *et al*., (2014) reported three copies of *Pun1* in a Chiltepin cultivar from Querétaro, Mexico, and Egan *et al*., (2019) reported the tandem duplication of *Pun1* and found recombination and divergence in other cultivars and related species; which shows the complexity of the genes related to pungency.

Chiltepin cultivars are semi wild with a high level of pungency and could have more than one copy of *Pun1* gene; however, the conditions in which they grow also affect the synthesis of capsaicinoids. In Northern Mexico, Chiltepin grows in extreme conditions of temperature and humidity, in this respect Kirke *et al.*, (2018) reported the functionality of the promoter of *Pun1* in a heterologous system and demonstrated the activation of the promoter by wounding and heat, so it is likely that the environment and the duplication of *Pun1* are responsible for the high pungency observed in Chiltepin cultivar. To date it is known that the *Pun1* encodes an acyltransferase, the identity of *Pun2* is currently unknown, and Han *et al*., (2019) reported the identification of a novel pungency-controlling gene *Pun3*, they suggest that the gene product acts as a master regulator of capsaicinoid biosynthetic genes in *C. annuum*. The above data show that there is a great variability in the *Pun* sequences, in our study we found several SNPs in cultivars of Northern Mexico and New Mexico, and possibly duplication of *Pun1* in the cultivar Chiltepin, we also show that this variability can also be analyzed with restriction enzymes on the specific gene, reducing the screening costs.

Although more detailed studies are necessary this study provides evidence that it is possible changes in the DNA sequence affect at the level of amino acids generating alterations that could potentially modify the activity of the protein and affect the amount of

capsaicinoids present in the cultivars, therefore in addition to the variety of *Capsicum* cultivar, temperature and nutrients, the genetic variability present in key genes of capsaicinoid biosynthesis can be analyzed to select the variables of agronomic interest that lead to the highest production of capsaicinoids for its use in different applications.

CONCLUSIONS

Cultivars of *C. annuum* of Northern Mexico have high variability of SNPs in the *Pun1* gene that regulates the biosynthetic pathway of major capsaicinoids, this variability can be analyzed by PCR-RFLP, in addition some SNPs generated changes in the amino acid sequence and could potentially affect the functionality of the protein, therefore the efficiency of capsaicinoid synthesis.

Acknowledgements
We thank Eliab Misael González Olvera for critical read of this manuscript and invaluable job in correcting the style of the English language.

Funding. AGZ, COC, JLGH, RPM, received support of Sistema Nacional de Investigadores (SNI, CONACyT).

Conflict of interest statement. The authors declare no conflict of interest

Compliance with ethical standards. The study complies with the ethical and biosafety standards of the participating institutions

Data availability. Data is available if necessary

REFERENCES

Arce-Rodríguez, M.L.; Ochoa-Alejo, N. 2017. An R2R3-MYB Transcription factor regulates capsaicinoid biosynthesis. Plant Physiology. 174:1359-1370. doi: https://doi.org/10.1104/pp.17.00506

Arce-Rodríguez, M.L.; Ochoa-Alejo, N. 2019. Biochemistry and molecular biology of capsaicinoid biosynthesis: recent advances and perspectives. Plant Cell Reports. 38(9):1017-1030. https://doi.org/10.1007/s00299-019-02406-0

Dellaporta, S.L.; Wood, J.; Hicks, J.B. 1983. A plant DNA minipreparation: Version II. Plant Molecular Biology Reporter. 1:19-21. https://doi.org/10.1007/BF02712670

Egan, A.N.; Moore, S.; Stellari, G.M.; Kang B.C.; Jahn, M.M. 2019. Tandem gene duplication and recombination at the *AT3* locus in the Solanaceae, a gene essential for capsaicinoid biosynthesis in *Capsicum*. PLoS ONE. 14(1): e0210510. https://doi.org/10.1371/journal.pone.0210510

González-Zamora, A.; Sierra-Campos, E.; Luna-Ortega, J.G.; Pérez-Morales, R.; Rodríguez, J.C.; García-Hernández, J.L. 2013. Characterization of different Capsicum varieties by evaluation of their capsaicinoids content by High Performance Liquid Chromatography, determination of pungency and effect of high temperature. Molecules. 18:13471-13486. https://doi: 10.3390/molecules181113471

González-Zamora, A.; Sierra-Campos, E.; Pérez-Morales, R.; Vázquez-Vázquez, C.; Gallegos-Robles, M.A.; López-Martínez, J.D.; García-Hernández, J.L. 2015. Measurement of Capsaicinoids in Chiltepin Hot Pepper: A Comparison Study between Spectrophotometric Method and High-Performance Liquid Chromatography Analysis. Journal of Chemistry. ID 709150, 10 pages. http://dx.doi.org/10.1155/2015/709150

Han, K.; Jang, S.; Lee, J.; Lee, D.; Kwon, J.; Kang, B. 2019. A MYB transcription factor is a candidate to control pungency in *Capsicum annuum*. Theoretical and Applied Genetics. 132(4):1235-1246. https://doi.org/10.1007/s00122-018-03275-z

Han, K.; Lee, E.Y., Ro, N.Y.; Hur, O.S.; Lee, J.H.; Kwon, J.K.; Kang, B.C. 2018. QTL mapping and GWAS reveal candidate genes controlling capsaicinoid content in *Capsicum*. Plant Biotechnol Journal. 16: 1546–1558 doi: 10.1111/pbi.12894

Hernández-Verdugo, S.; Dávila-Aranda, P.; Oyama, K. 1999. Taxonomical synthesis of the knowledge, origin and domestication of the genus *Capsicum*. Síntesis del conocimiento taxonómico, origen y domesticación del género *Capsicum*. Boletin de la Sociedad Botánica de México. 64:65-84. doi: 10.17129/botsci.1583

Hulse-Kemp, A.M.; Maheshwari, S.; Stoff, K.; Hill, T.A.; Jaffe, D.; Williams, S.R.; Weisenfeld, N.; Ramakrishnan, S.; Kumar, V.; Shah, P.; Schatz, M.; Church, D.; Deynze, A.V. 2018. Reference quality assembly of the 3.5-Gb genome of *Capsicum annuum* from a single linked-read library. Horticulture Research. 5(4):1 – 13. doi:10.1038/s41438-017-0011-05.

Keyhaninejad, N.; Curry, J.; Romero, J.; O´Connell, M.A. 2014. Fruit specific variability in capsaicinoid accumulation and transcription of structural and regulatory genes in *Capsicum* fruit. Plant Science. 59-68. doi: 10.1016/j.plantsci.2013.10.013

Kim, S.; Park, M.; Yeom, S.I.; Kim, Y.M.; Lee, J.M.; Lee, H.A. 2014. Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. Nature Genetics. 46(3):270-279. doi:10.1038/ng.2877

Kirke, J.; Kaplan, N.; Velez, S.; Jin, X.L.; Vichyavichien, P.; Zhang, X.H. 2018. Tissue-preferential activity and induction of the pepper capsaicin synthase PUN1 promoter by wounding, heat and metabolic pathway precursor in tobacco and tomato plants. Molecular Biotechnology. 60(3):194-202. https://doi.org/10.1007/s12033-018-0060-0

Kraft, K.; Brown, C.; Nabhan, G.; Luedeling, E.; Luna-Ruiz, J.J.; d'Eeckenbrugg, G.C.; Hijmans R.; Gepts, P. 2014. Multiple lines of evidence for the origin of domesticated chili pepper, *Capsicum annuum*, in Mexico. Proceedings of the National Academy of Sciences of the United States of America. 111(17):6165-6170. https://doi.org/10.1073/pnas.1308933111

Langhans, W. 2018. Food Components in health promotion and disease prevention. Journal of Agricultural and Food Chemistry. 66:2287-2294. doi: 10.1021/acs.jafc.7b02121.

Nimmakayala, P.; Abburi, V.; Saminathan, T.; Alaparthi, S.; Almeida, A.; Davenport, B.; Nadimi, M.; Davidson, J.; Tonapi, K.; Yadav, L.; Malkaram, S.; Vajja, G.; Hankins, G.; Harris, R.; Park, M.; Choi, D.; Stommel, J.; Reddy, UK. 2016. Genome-wide diversity and association mapping for capsaicinoids and fruit weight in *Capsicum annuum* L. Scientific Reports (Nature), 6. 38081; doi: 10.1038/srep38081

Ogawa, K.; Murota, K.; Shimura, H.; Furuya, M.; Togawa, Y.; Matsumura, T.; Masuta, C. 2015. Evidence of capsaicin synthase activity of the Pun1-encoded protein and its role as a determinant of capsaicinoid accumulation in pepper. BMC Plant Biology. 15, 93. https://doi.org/10.1186/s12870-015-0476-7

Popelka, P.; Jevinová, P.; Šmejkal, K.; Roba, P. 2017. Determination of capsaicin content and pungency level of different fresh and dried chilli peppers. Folia Veterinaria. 61, 11-16. doi: https://doi.org/10.1515/fv-2017-0012

Qin, C.; Yu, C.; Shen, Y.; Fang, X.; Chen, L.; Min, J.; Cheng, J.; Zhao, S.; Xu, M.; Luo, Y. 2014. Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. Proceedings of the National Academy of Sciences of the United States of America. 111, 5135-5140. https://doi.org/10.1073/pnas.1400975111

Reddy, U.K.; Almeida, A.; Abburi, V.L.; Alaparthi, S.B.; Unselt, D.; Hankins, G.; Park, M.; Choi, D.; Nimmakalaya, P. 2014. Identification of gene-specific polymorphisms and association with capsaicin pathway metabolites in *Capsicum annuum* L. Collections. PLoS ONE. 9, e86393. https://doi.org/10.1371/journal.pone.0086393

Rezende-Naves, E.; de Ávila-Silva L.; Sulpice, R.; Araújo, W.; Nunes-Nesi, A.; Peres, L.; Zsögön, A. 2019. Capsaicinoids: Pungency beyond *Capsicum*. Trends in Plant Science. 24 (2). 109-120. https://doi.org/10.1016/j.tplants.2018.11.001

Sarpras, M.; Rashmi G.; Vineet, S.; Sushil, S.C.; Jharna, D.; Ajay, K. 2016. Comparative analysis of fruit metabolites and pungency candidate genes expression between Bhut Jolokia and other *Capsicum* species. PLoS ONE. 11(12). e0167791. https://doi.org/10.1371/journal.pone.0167791

Stellari, G.M.; Mazourek, M.; Jahn, M.M. 2010. Contrasting modes for loss of pungency between cultivated and wild species of *Capsicum*. Heredity. 104. 460–471. https://doi.org/10.1038/hdy.2009.131

Stewart, C.; Mazourek, M.; Stellari, G.M.; O'Connell, M.; Jahn, M. 2007. Genetic control of pungency in *C. chinense* via the *Pun1* locus. Journal of Experimental Botany. 58 (5). 979–991. doi:10.1093/jxb/erl243

Stewart, C.; Kang, B.C.; Liu, K.; Mazourek, M.; Moore, S.L.; Yoo, E.Y.; Kim, B.D.; Paran, I.; Jahn, M.M. 2005. The *Pun1* gene for pungency in pepper encodes a putative acyltransferase. The Plant Journal. 42, 675-688. doi: 10.1111/j.1365-313X.2005.02410.x

Tremblay, A.; Arguin, H.; Panahi, S. 2016. Capsaicinoids: a spicy solution to the management of obesity? International Journal

of Obesity (Lond). 40, 1198-1204. doi: 10.1038/ijo.2015.253

Wen, W.; He, Z.; Gao, F.; Liu, J.; Jin, H.; Zhai, S.; Qu, Y.; Xia, X. 2017. A High-density consensus map of common wheat integrating four mapping populations scanned by the 90K SNP array. Frontiers in Plant Science. 8, 1389. doi: 10.3389/fpls.2017.01389.

Zhang, Z.X.; Zhao, S.N.; Liu, G.F.; Huang, Z.M.; Cao, Z.M.; Cheng, S.H.; Lin, S.S. 2016. Discovery of putative capsaicin biosynthetic genes by RNA-Seq and digital gene expression analysis of pepper. Scientific Reports (Nature). 6, 34121. doi: 10.1038/srep34121.