

IDENTIFICAÇÃO E ANÁLISE *IN SILICO* DO GENE *P5CS* (Δ^1 -PIRROLINA-5-CARBOXILATO SINTETASE) EM *COFFEA* SSP.

Fernanda Dolcimasculo¹, Silvia Graciele Hülse de Souza², Alessandra F. Ribas³, Luiz Gonzaga Esteves Vieira⁴, Tiago Benedito dos Santos⁵

¹Mestranda, Universidade do Oeste Paulista (UNOESTE), Presidente Prudente-SP, ferdolci@gmail.com

²Pesquisadora, DS, Universidade Paranaense (UNIPAR), Umuarama-PR, silviahulse@prof.unipar.br

³Pesquisadora, DS, Universidade do Oeste Paulista (UNOESTE), Presidente Prudente-SP, alessandra_ribas@hotmail.com

⁴Pesquisador, PhD, Universidade do Oeste Paulista (UNOESTE), Presidente Prudente-SP, luizvie@gmail.com

⁵Pesquisador, DS, Universidade do Oeste Paulista (UNOESTE), Presidente Prudente-SP, dossantos@unoeste.br

RESUMO: A enzima Δ^1 -pirrolina-5-carboxilato sintetase (P5CS) participa da biossíntese do aminoácido prolina e é amplamente estudada em monocotiledôneas e dicotiledôneas. A prolina nas plantas superiores é sintetizada a partir da via glutamato e ornitina, e atua na osmoproteção para manter a condição da homeostase, estabilizar a estrutura celular e prevenir a degradação de proteínas em condições de estresses abióticos. Este estudo teve como objetivo identificar e analisar *in silico* o gene *P5CS* em três espécies de *Coffea*. Através da análise das sequências genômicas foram identificados um total de sete genes *P5CS* putativos nos genomas de *Coffea* spp., sendo três em *C. arabica*, dois em *C. canephora* e dois em *C. eugenoides*. Adicionalmente, foi realizada uma análise detalhada sobre características físico-químicas das proteínas, estrutura gênica, relações filogenéticas e análise de sintenia entre esses genomas.

PALAVRAS-CHAVE: Bioinformática, café, prolina, genoma, P5CS.

IDENTIFICATION AND *IN SILICO* ANALYSIS OF THE *P5CS* GENE (Δ^1 -PYRROLINE-5-CARBOXYLATE SYNTHETASE) IN *COFFEA* SSP.

ABSTRACT: The Δ^1 -pyrroline-5-carboxylate synthetase (P5CS) is one of the important enzymes in proline biosynthesis and is widely studied in monocotyledons and dicotyledons. Proline in higher plants is synthesized from ornithine and glutamate pathway, and acts in osmoprotection to maintain homeostasis, stabilize the cellular structure and prevent protein degradation. This study was aimed to identify and analyze *in silico* *P5CS* genes in three species of *Coffea*. Through the analysis of genomic sequences, a total of seven putative *P5CS* genes were identified in the genomes of *Coffea* spp., three in *C. arabica*, two in *C. canephora* and two in *C. eugenoides*. In addition, a detailed analysis was performed on the physicochemical characteristics of the proteins, gene structure, phylogenetic relationships and synteny analysis among the genomes of these species of *Coffea*.

KEY WORDS: Bioinformatics, coffee, proline, genome, P5CS.

INTRODUÇÃO

O acúmulo do aminoácido prolina é uma das respostas amplamente documentada em plantas para muitos estresses bióticos e abióticos (revisado por Signorelli & Monza, 2017). Em plantas superiores a biossíntese da prolina ocorre principalmente através da via do glutamato, que inclui duas reações enzimáticas e uma não enzimática (Hayat et al., 2012). As reações enzimáticas ocorrem no citoplasma ou cloroplasto e são catalisadas pela Δ^1 -pirrolina-5-carboxilato sintetase (P5CS; EC 2.7.2.11) e Δ^1 -pirrolina-5-carboxilato redutase (P5CR; EC 1.5.1.2) (Fig. 1).

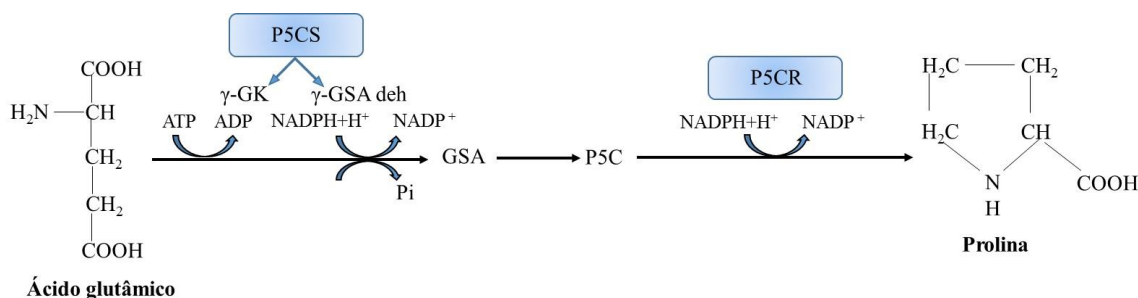


Figura 1. Via de biossíntese da prolina em plantas. Adaptado de Bandurska et al. (2017).

Evolutivamente o gene *P5CS* passou por um processo de duplicação em monocotiledôneas e dicotiledôneas (Rai & Penna, 2013). Neste sentido, a maioria das plantas apresentam 2 genes *P5CS*, um constitutivo e outro induzível por estresses. Os genes da via da prolina, especialmente *P5CS*, são comumente usados na engenharia genética para conferir às plantas tolerância sob condições de estresses ambientais (Molinari et al., 2007; Guerzoni et al., 2014; Borgo et al.,

2015). O objetivo desse estudo foi identificar e analisar os genes *P5CS* com base nas informações disponíveis nos bancos de dados dos genomas de *Coffea arabica*, *C. canephora* e *C. eugenioides*.

MATERIAL E MÉTODOS

As buscas das sequências de interesse foram feitas nas plataformas do genoma de *C. canephora* (Denoued et al., 2014) e no NCBI (*National Center for Biotechnology Information*: <http://www.ncbi.nlm.nih.gov/> Altshul et al., 1997) para as espécies *C. arabica* e *C. eugenioides*. Essas buscas foram realizadas através da ferramenta *Blast* com a utilização de sequências proteicas codificantes para a enzima *P5CS* previamente identificadas no genoma de *Arabidopsis thaliana* (www.arabidopsis.org/ AT2G39800). Em seguida, todas as sequências identificadas foram individualmente confrontadas com aquelas depositadas no banco de dados do NCBI utilizando a ferramenta *Blastp*, com a finalidade de confirmar a identidade das mesmas. As características físico-químicas das proteínas, tais como peso molecular (Mw) e ponto isoelétrico (pI), foram calculadas utilizando a ferramenta *ExPASy* - Compute pI/Mw (https://web.expasy.org/compute_pi/). As localizações subcelulares foram determinadas usando o servidor *CELLO* (*preditor de Localization subelular*) (cello.life.nctu.edu.tw/; Yu et al. 2006) e o algoritmo *Plant-mPLoc* (<http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/> Chou; Shen, 2010). A estrutura gênica foi representada de forma ilustrativa com a utilização do software *GSDS 2.0 Gene Structure Display Gene* (Hu et al., 2015). Posteriormente, foi realizado o alinhamento de aminoácidos no programa *ClustalW*, para a construção de dendrograma baseado na similaridade das sequências de ortólogos de outros organismos. O dendrograma foi gerado pelo programa *MEGA7*, através do método de *Neighbour-Joining*, com teste de confiabilidade *bootstrap* de 1000 repetições (Kumar et al., 2016). O mapa de sentença entre as três espécies de café foram construídos usando o software *Circos* (<http://circos.ca/>; Krzywinski, et al., 2009). Foram considerados genes sintênicos àqueles pares de genes que apresentaram os seguintes parâmetros: 80% de cobertura e 90% de identidade.

RESULTADOS E DISCUSSÃO

As informações gerais dos genes identificados são apresentadas na Tabela 1. Inicialmente, foram identificadas três sequências putativas para gene *P5CS* no genoma de *C. arabica*, duas sequências no genoma de *C. canephora* e cinco sequências no genoma de *C. eugenioides*, respectivamente. Contudo, ao analisar as sequências genômicas, proteicas e CDS, ficou evidente que existem apenas dois genes *P5CS* nesta última espécie. Portanto, as demais sequências de *C. eugenioides* não foram utilizadas para as análises posteriores (Número de acesso: XP_027158919, XP_027175339 e XP_027156561). Os genes *P5CS*, foram nomeados numericamente em ordem crescente conforme os loci identificados e anotados nos respectivos genomas (Tabela 1).

Tabela 1. Lista das principais características dos genes *P5CS* identificados nos genomas de *Coffea* ssp.

Gene	Gene ID	Localização cromossômica	Éxon	Proteína (aa)	pI/Mw	Cello	Plant-mPLoc	Subgenoma
<i>CaP5CS1</i>	XP_027110917	chr2:5762715..5773661	20	716	6.21/77126.31	Citoplasma	Citoplasma	Ca ^e
<i>CaP5CS2</i>	XP_027068820	chr6:1717505..1726360	20	713	6.26/77380.73	Citoplasma	Citoplasma	Ca ^c
<i>CaP5CS3</i>	XP_027071497	chr6:2006849..2015846	20	713	6.26/77433.61	Citoplasma	Citoplasma	Ca ^e
<i>CcP5CS1</i>	Cc02_g07340	chr2:5821657..5832197	20	726	6.38/78147.52	Mitocôndria	Citoplasma	***
<i>CcP5CS2</i>	Cc06_g02210	chr6:1777222..1785698	20	726	6.44/78950.41	Citoplasma	Citoplasma	***
<i>CeP5CS1</i>	XP_027158918	chr2:7382735..7393801	21	716	6.18/76965.12	Citoplasma	Citoplasma	***
<i>CeP5CS2</i>	XP_027175338	chr6:3088947..3103360	21	716	6.20/77795.02	Citoplasma	Citoplasma	***

De acordo com as análises realizadas, o tamanho deduzido da sequência de aminoácidos dos genes variou de 713 a 726 aa. O ponto isoelétrico (pI) variou de 6.18 a 6.44 e o peso molecular (Mw) da proteína variou de 77126.31 a 78950.41. A previsão de localização das proteínas *P5CS* apontam as suas presenças no citoplasma (Tabela 1). Em *Carica papaya* a proteína *CpP5CS* foi encontrada principalmente no retículo endoplasmático e núcleo (Zhu et al., 2012). Liu et al. (2012) acrescentam que o gene *P5CS1*, localizado no cloroplasto, é responsável pela síntese de prolina induzida por estresse, enquanto o *P5CS2*, localizado no citosol, é importante para o desenvolvimento embrionário (Szekely et al., 2008; Funck et al., 2012). A maioria das plantas possui dois genes *P5CS*, enquanto algumas espécies apresentam apenas um gene. No entanto, em *C. arabica*, foram identificados três possíveis genes - *CaP5CS1*, *CaP5CS2* e *CaP5CS3*, respectivamente, sendo possível deduzir a contribuição dos subgenomas para o genoma deste poliploide através de informações obtidas no banco de dados do NCBI (Tabela 1). As isoformas putativas *CaP5CS1* e *CaP5CS3* são originárias do subgenoma de *C. eugenioides* (Ca^e), enquanto *CaP5CS2* provém do subgenoma de *C. canephora* (Ca^c) (Tabela 1). Diversos trabalhos mostram que a conservação evolutiva entre as espécies em apresentar diferentes genes *P5CS* favorece possíveis regulações diferenciadas em resposta às condições ambientais (Turchetto-Zolet et al., 2009;

Mattioli et al., 2009a; 2009b; revisado por Signorelli & Monza, 2017). Quanto a heterogeneidade estrutural dos genes P5CS de *Coffea* ssp., esta foi variável apresentado de 20 a 21 éxons (Fig. 2).

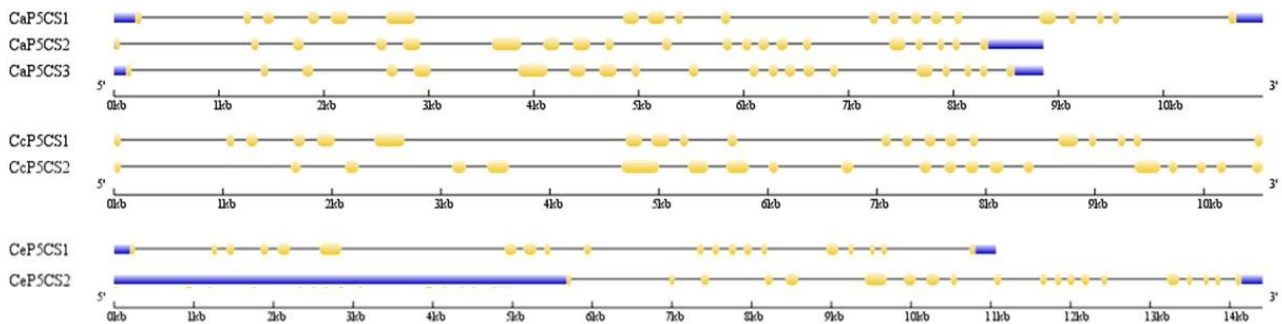


Figura 2. Representação da estrutura gênica dos genes P5CS de *Coffea* ssp. Os íntrons foram representados pela linha preta e os éxons pela caixa amarela.

A relação filogenética entre as proteínas P5CS das três espécies de *Coffea* e de outras espécies foi analisada (Fig. 3) e revelou dois grandes grupos: o grupo das monocotiledôneas e grupo das dicotiledôneas. Esse fato era esperado considerando que eventos de duplicação ocorreram após a divergência de monocotiledôneas e dicotiledôneas (Turchetto-Zolet et al., 2009). No geral, todas essas observações indicam que o P5CS de *Coffea* ssp. compartilham características comuns com os P5CSs obtidos de outras plantas.

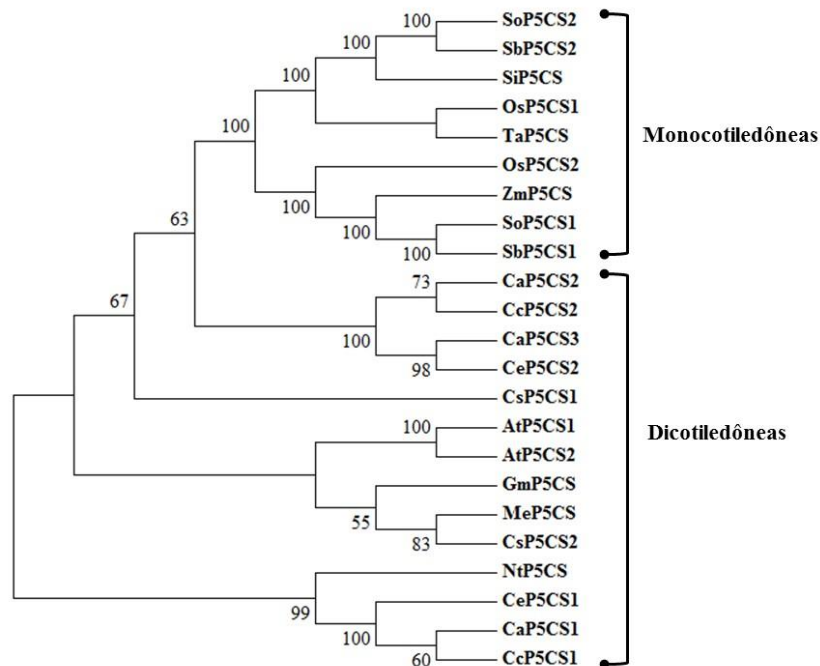


Figura 3. Análise filogenética das sequências P5CS (*Coffea* ssp) e suas homólogas. Monocotiledôneas: *Zea mays* (ZmP5CS - XP008656219), *Sorghum bicolor* (SbP5CS1 - ACU65226; SbP5CS2 - ACU65227), *Oriza sativa* (OsP5CS1 - BAA19916; OsP5CS2 - BAG94966), *Saccharum officinarum* (SoP5CS1 - AGW20993.1; SoP5CS2 - AGW20994.1), *Setaria italica* (SiP5CS - XP004961886), *Triticum aestivum* (TaP5CS - BAD97364); Dicotiledôneas: *C. arabica* (CaP5CS1 - XP_027110917; CaP5CS2 - XP_027068820; CaP5CS3 - XP_027071497), *C. canephora* (CcP5CS1 - Cc02_g07340; CcP5CS2 - Cc06_g02210), *C. eugenoides* (CeP5CS1 - XP_027158918; CeP5CS2 - XP_027175338), *Manihot esculenta* (MeP5CS - XP_021626266.1), *Citrus sinensis* (CsP5CS1 - KDO48125.1; CsP5CS2 - KDO69189.1), *A. thaliana* (AtP5CS1 - NP_001198715.1; AtP5CS2 - NP_191120.2), *Nicotiana tabacum* (NtP5CS - AGW30209), *Glycine max* (GmP5CS - AAR86688), pelo método de *Neighbour-Joining* com *bootstrap* de 1000 replicatas (Kumar et al., 2016).

Posteriormente, uma análise de sintenia foi realizada em *Coffea* ssp., gerando uma tabela e mapas comparativos, consistindo de: *C. arabica* x *C. canephora*; *C. arabica* x *C. eugenoides*; *C. canephora* x *C. eugenoides*; *C. arabica* x *C. canephora* x *C. eugenoides* (Fig. 4). O mapa (A) *C. arabica* x *C. canephora* apresentaram, três pares de relações de sintenia; o mapa (B) *C. arabica* x *C. eugenoides* apresentaram três pares de relações de sintenia; o mapa (C) *C. canephora* x *C. eugenoides*, apresentou dois pares de sintenia, respectivamente (Fig. 4). Na sintenia conjunta entre *C. arabica* x *C. canephora* x *C. eugenoides* apresentaram oito pares de sintenia (Fig. 4 D).

De modo geral, considerando os resultados das análises *in silico* e com base nas informações dos genomas dos diploides, *C. canephora* e *C. eugenioides*, a espécie *C. arabica* apresenta um evento de duplicação. No genoma das plantas é comum observar a ocorrência de eventos de duplicação, sendo estes eventos considerados um importante processo evolutivo, resultando em variabilidade entre as plantas (Jiao et al., 2011). Em alguns casos, essas duplicações conferiram o que se chama neofuncionalizações dos genes, sendo relacionado a um caráter adaptativo (Panchy et al., 2016). Renny-Byfield et al. (2014) ressalta ainda a importância da ocorrência do processo de poliploidia em plantas de interesse agrônomo. Neste sentido, pode-se propor que os genes *P5CS* identificados no genoma de *C. arabica* passaram pelo processo de duplicação gênica (Tabela 1).

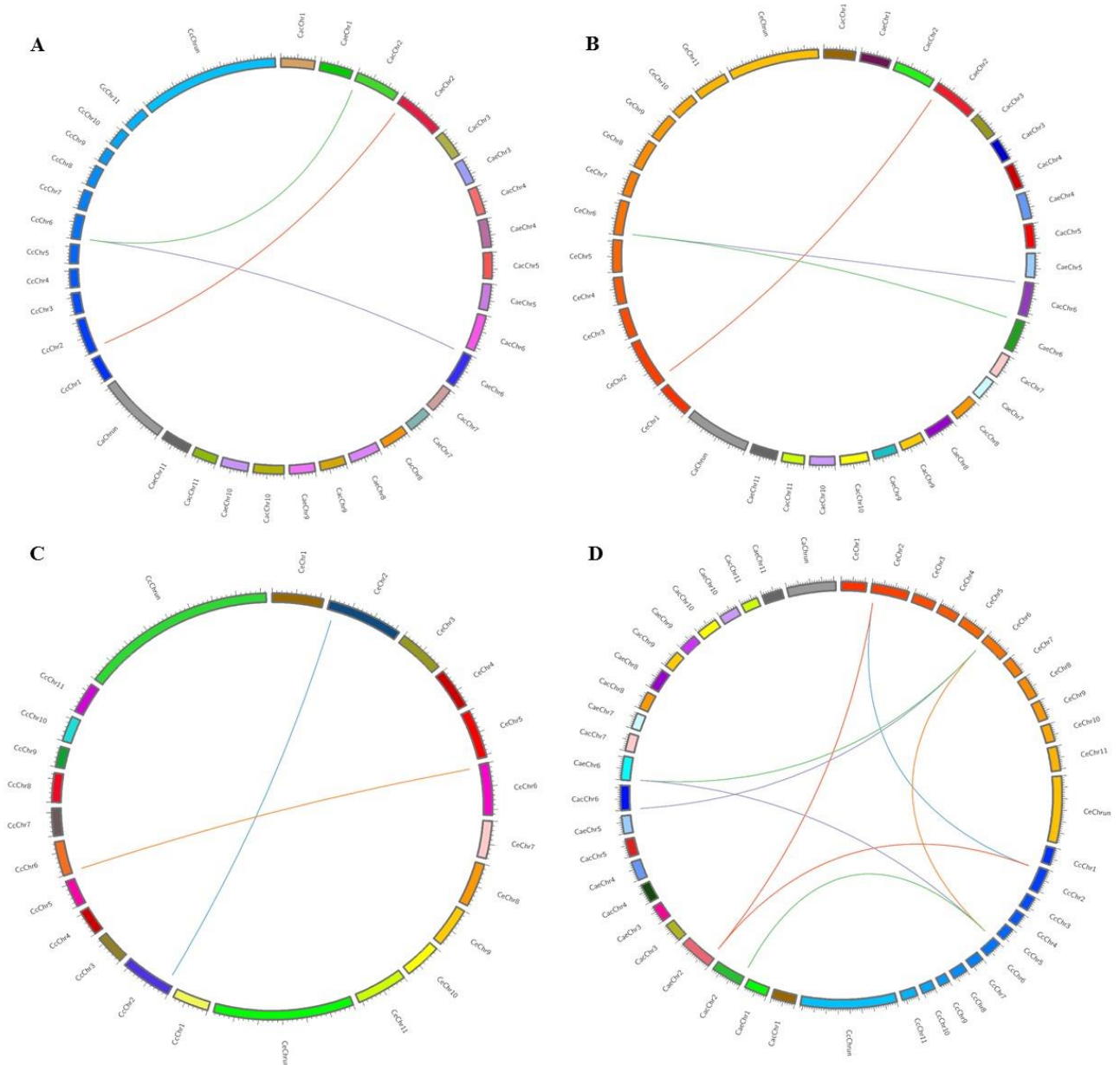


Figura 4. Análise de sintenia do gene *P5CS* de *C. arabica*, *C. canephora* e *C. eugenioides*. (A) Mapa comparativo entre *C. arabica* e *C. canephora*. (B) Mapa comparativo entre *C. arabica* e *C. eugenioides*. (C) Mapa comparativo entre *C. canephora* e *C. eugenioides*. (D) Sintenia conjunta com *C. arabica*, *C. canephora* e *C. eugenioides*.

Tabela 2. Sintonias entre genes *P5CS* de *C. arabica*, *C. canephora* e *C. eugenioides*.

<i>Coffea arabica</i>			<i>Coffea canephora</i>				
Gene	Gene ID	Localização cromossômica	Gene	Gene ID	Localização cromossômica	Query cover (%)	Identity (%)
1 <i>CaP5CS1</i>	XP_027110917	chr2:5762715..5773661	<i>CcP5CS1</i>	Cc02_g07340	chr2:5821657..5832197	95	97,2
2 <i>CaP5CS2</i>	XP_027068820	chr6:1717505..1726360	<i>CcP5CS2</i>	Cc06_g02210	chr6:1777222..1785698	94	99,5
3 <i>CaP5CS3</i>	XP_027071497	chr6:2006849..2015846	<i>CcP5CS2</i>	Cc06_g02210	chr6:1777222..1785698	94	97,8
<i>Coffea arabica</i>			<i>Coffea eugenioides</i>				
Gene	Gene ID	Localização cromossômica	Gene	Gene ID	Localização cromossômica	Query cover (%)	Identity (%)
1 <i>CaP5CS1</i>	XP_027110917	chr2:5762715..5773661	<i>CeP5CS1</i>	XP_027158918	chr2:7382735..7393801	100	99,1
2 <i>CaP5CS2</i>	XP_027068820	chr6:1717505..1726360	<i>CeP5CS2</i>	XP_027175338	chr6:3088947..3103360	100	98,1
3 <i>CaP5CS3</i>	XP_027071497	chr6:2006849..2015846	<i>CeP5CS2</i>	XP_027175338	chr6:3088947..3103360	100	99,8
<i>Coffea canephora</i>			<i>Coffea eugenioides</i>				
Gene	Gene ID	Localização cromossômica	Gene	Gene ID	Localização cromossômica	Query cover (%)	Identity (%)
1 <i>CcP5CS1</i>	Cc02_g07340	chr2:5821657..5832197	<i>CeP5CS1</i>	XP_027158918	chr2:7382735..7393801	100	97,2
2 <i>CcP5CS2</i>	Cc06_g02210	chr6:1777222..1785698	<i>CeP5CS2</i>	XP_027175338	chr6:3088947..3103360	100	97,9

CONCLUSÕES

1 - Neste estudo, foram identificados os genes putativos que codificam a enzima P5CS nas espécies de *C. arabica*, *C. canephora* e *C. eugenioides*.

2 - A caracterização funcional de P5CS em *Coffea* ssp. pode ser importante em futuros estudos que visem auxiliar a compreensão dos mecanismos de modulação da resposta adaptativa dessas espécies, e suas cultivares, aos estresses abióticos.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALTSCHUL, S.F.; MADDEN, T.L.; SCHAFER, A.A.; ZHANG, J.; ZHANG, Z.; MILLER, W.; LIPMAN, D. Gapped BLAST and PSIBLAST: a new generation of protein database search programs. **Nucleic Acids Research**, v. 25, p. 3389–3402, 1997.
- BANDURSKA, H.; NIEDZIELA, J.; PIETROWSKA-BOREK, M.; NUC, K.; CHADZINKOLAU, T.; RADZIKOWSKA, D. Regulation of proline biosynthesis and resistance to drought stress in two barley (*Hordeum vulgare* L.) genotypes of different origin. **Plant Physiology and Biochemistry**, v. 118, p. 427–437, 2017.
- BORGIO, L.; MARUR, C.J.; VIEIRA, L.G.E. Effects of high proline accumulation on chloroplast and mitochondrial ultrastructure and on osmotic adjustment in tobacco plants. **Acta Sci Agron**, v. 37, p. 191–199, 2015.
- DENOEUDE, F.; CARRETERO-PAULET, L.; DEREEPER, A.; DROC, G.; GUYOT, R.; PIETRELLA, M.; ZHENG, C.; ALBERTI, A.; ANTHONY, F.; APREA, G. et al. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. **Science**, v. 345, n. 6201, p. 1181–1184, 2014.
- FUNCK, D.; WINTER, G.; BAUMGARTEN, L.; FORLANI, G. Requirement of proline synthesis during Arabidopsis reproductive development. **BMC Plant Biol.** v. 12, p. 91, 2012.
- GUERZONI, J.T.S.; BELINTANI, N.G.; MOREIRA, R.M.P.; HOSHINO, A.A.; DOMINGUES, D.S.; BESPALHOK, FILHO, J.C.; VIEIRA, L.G.E. Stress-induced Δ^1 -pyrroline-5-carboxylate synthetase (P5CS) gene confers tolerance to salt stress in transgenic sugarcane. **Acta Physiol Plant**, v. 36, p. 2309–2319, 2014.
- HAYAT, S.; HAYAT, Q.; ALYEMENI, M. N.; WANI, A. S.; PICHTEL, J.; AHMAD, A. Role of proline under changing environments: a review. **Plant Sign. Behav.**, v. 7, p. 1456–1466, 2012.
- HU, B.; JIN, J.; GUO, A.-Y.; ZHANG, H.; LUO, J. GAO. GSDB 2.0: an upgraded gene feature visualization server. **Bioinformatics**, v. 31, n. 8, p. 1296–1297, 2015.
- KUMAR, S.; STECHER, G.; TAMURA, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. **Mol Biol Evol.** v. 33, n.7, p. 1870–1874, 2016.
- KUO-CHEN.; CHOU.; HONG, BIN-SHEN, "Plant-mPLOC: a top-down strategy to augment the power for predicting plant protein subcellular localization", **PLoS ONE**, v. 5, p. 11335, 2010.

- KRZYWINSKI, M.; SCHEIN, J.; BIROL, I.; CONNORS, J.; GASCOYNE, R.; HORSMAN, D.; JONES, S.J.; MARRA, M.A. Circos: An Information Aesthetic for Comparative. **Genomics. Genome Research**, v. 19, p. 1639–1645, 2009.
- LIU, W.; LE, A.; HANCOCK, C.; LANE, A.N.; DANG, C.V.; FAN, T.W.; et al. (). Reprogramming of proline and glutamine metabolism contributes to the proliferative and metabolic responses regulated by oncogenic transcription factor c-MYC. **Proc. Natl. Acad. Sci. USA**, v. 109, p. 8983–8988, 2012.
- MATTIOLI, R.; COSTANTINO, P.; TROVATO, M. Proline accumulation in plants: not only stress. **Plant Signal Behav**, v. 2, p. 1016, 2009.
- MATTIOLI, R.; FALASCA, G.; SABATINI, S.; ALTAMURA, M.M.; COSTANTINO, P.; TROVATO, M. The proline biosynthetic genes P5CS1 and P5CS2 play overlapping roles in Arabidopsis flower transition but not in embryo development, **Physiol Plant**, 2009.
- MOLINARI, H.B.; MARUR, C.J.; BESPALHOK, J.C.F.; KOBAYASHI, A.K.; PILEGGI, M.; LEITE-JUNIOR, R.P.; PEREIRA, L.F.P.; VIEIRA, L.G.E. Osmotic adjustment in transgenic citrus rootstock Carrizo citrange (*Citrus sinensis* Osb. x *Poncirus trifoliata* L. Raf.) overproducing proline. **Plant Sci**, 167:1375–1381, 2004.
- RAI, A.N.; PNNA, S. Molecular evolution of plant P5CS gene involved in proline biosynthesis. **Mol Biol Rep**, v. 4, p. 6429–6435, 2013.
- SIGNORELLI, S.; MONZA, J. Identification of Δ^1 -pyrroline 5-carboxylate synthase (P5CS) genes involved in the synthesis of proline in *Lotus japonicus*. **Plant signaling & behavior**, v. 12, n. 11, e 1367464, 2017.
- SZEKELY, G.; ABRAHAM, E.; CSEPLO, A.; RIGO, G.; ZSIGMOND, L.; CSISZAR, J. et al. Duplicated P5CS genes of Arabidopsis play distinct roles in stress regulation and developmental control of proline biosynthesis. **Plant J**, v. 53, p. 11–28, 2008.
- TURCHETTO-ZOLET, A.C.; MARGIS-PINHEIRO, M.; MARGIS, R. The evolution of pyrroline-5-carboxylate synthase in plants: A key enzyme in proline synthesis. **Mol Genet Genomics**, v. 281, p. 87–97, 2009.
- YU, C. S.; CHEN, Y. C.; LU, C. H.; HWANG, J. K. Prediction of protein subcellular localization. **Proteins: Structure, Function, and Bioinformatics**, v. 64, p. 643–651, 2006.
- ZHU, X.Y.; LI, X.P.; ZOU, Y, et al. Cloning, characterization and expression analysis of Δ^1 -pyrroline-5-carboxylate synthetase (P5CS) gene in harvested papaya (*Carica papaya*) fruit under temperature stress. **Food Res Int.**, v. 49, p. 272–279, 2012.