

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

ORIGEM E DIVERSIFICAÇÃO DA FAMÍLIA GÊNICA P5CS

JOÃO PEDRO DO CARMO FILGUEIRAS

Dissertação submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para a obtenção do título de **Mestre em Genética e Biologia Molecular**.

Orientadora: Prof^ª Dr^ª: Andreia Carina Turchetto Zolet

Porto Alegre, março de 2021

Instituições e fontes financiadoras

O presente trabalho foi executado no Laboratório de Genética Vegetal do Departamento de Genética da Universidade Federal do Rio Grande do Sul (UFRGS) e teve como fontes financiadoras o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e o Programa de Pós-graduação em Genética e Biologia Molecular (PPGBM) da UFRGS.

Agradecimentos

Início agradecendo a minha **família**, pela confiança, por sempre incentivarem meus sonhos e a estudar cada vez mais. A “vaquinha” que meus tios e alguns primos fizeram para me ajudar na mudança de Goiás para o Rio Grande do Sul é a principal prova disso, sou muito grato e amo todos vocês. Agradeço em especial a meu Pai **Ronie** e principalmente a minha mãe **Rosana**, que de fato me ajudaram a estar aqui, pelas batalhas e por serem sempre meu exemplo de força,

Agradeço à minha orientadora **Andreia**, pela confiança que teve em mim em desvendar a evolução deste gene pelo qual ela tem um carinho especial. E que mesmo nesse ano tão turbulento como foi, estava sempre disposta a me ajudar e sempre muito preocupada com a saúde mental de seus orientados também, o que no meio científico acaba sendo deixado de lado às vezes. Então, obrigado pelos ensinamentos, pelas reuniões além do horário, pelo carinho, por ser tão querida sempre e pelo exemplo de orientadora que és.

Agradeço às minhas companheiras de apartamento e amigas **Analú e Luana**, que em meio a pandemia me ajudaram para o surto não vir. Pelas risadas, choros, confusões e aventuras que tivemos. Em especial a Luana, de veterana a colega de mestrado, que veio “lá du Goiás” comigo, e no qual conhecemos e exploramos um pouquinho de Porto Alegre. Por causa delas, me aproximei bastante do Laboratório de Evolução Molecular, que me “adotaram” e foram sempre muito carinhosos e acolhedores, sendo a companhia de RU, de festinhas e comemorações. Então agradeço também aos **amigos do LEM**.

Aos amigos do **Laboratório de Genética Vegetal** pela companhia no primeiro ano e por sempre manterem um local de estudo/trabalho agradável de se conviver. Agradeço em especial a **Drielli** pela ajuda e dicas de bioinformática e a **Ariane** por ser sempre tão querida, pelas pausas para um docinho ou um chá, e por ser essa pessoinha tão especial que irei carregar para sempre no meu coração.

E por fim, a minha família de coração, meus amigos de longa data **Ana Gabriela, Leonardo e Reiner**. Obrigado por estarem sempre comigo, mesmo a distância, por estarem na torcida, pelos conselhos, por ouvirem os desabafos e aventuras. Amo muito vocês.

Sumário

Resumo	5
Abstract	7
1. Introdução	9
1.1 Prolina e suas funções	9
1.2 Metabolismo da prolina	10
1.3 P5CS e seus homólogos γ -GK/ γ -GPR	14
2. Objetivos	19
2.1 Objetivo geral	19
2.2 Objetivos específicos	19
3. Capítulo I - Origin and diversification of P5CS gene family	20
Abstract	22
Introduction	23
Material and Methods	24
<i>Data Sources</i>	25
<i>Multiple Sequence Alignment</i>	25
<i>Phylogenetic Analysis</i>	26
<i>Selection analysis</i>	26
Results	27
<i>Global Identification of P5CS, ProA and ProB genes</i>	27
<i>Origin of P5CS and its evolutionary relationships with ProA and ProB gene</i>	28
<i>Duplication and losses of genes</i>	29
<i>Selection pressures</i>	30
Discussion	30
<i>P5CS origin</i>	30
<i>Sites under positive selection</i>	32
<i>Duplication</i>	35
Conclusion	36
References	37
Figures and Tables	42
Supplementary Material	51
4. Considerações finais	64
5. Referências Bibliográficas	66

Resumo

Em resposta a vários estresses, o acúmulo de prolina ocorre em bactérias, protozoários, invertebrados marinhos e em plantas. Em animais e plantas, a prolina também desempenha outras funções fisiológicas. A principal via de biossíntese da prolina tem o glutamato como substrato e ocorre por três reações enzimáticas. Em procariotos, alguns eucariotos unicelulares e fungos, a primeira e a segunda reação são catalisadas pelas enzimas γ -GK e γ -GPR, que são codificadas pelos genes *ProB* e *ProA*, respectivamente. Em animais, plantas e em alguns eucariotos unicelulares um único gene, denominado P5CS, codifica uma enzima bifuncional com os domínios, GK e GPR, responsável pela primeira e segunda reações na biossíntese de prolina. Tendo em vista a importância fisiológica da prolina e do seu metabolismo nos organismos vivos, o principal objetivo deste trabalho é elucidar a história evolutiva da família gênica P5CS. Apesar de serem encontrados outros estudos que abordam aspectos evolutivos do P5CS, eles não utilizam uma grande amostragem ou ficam restritos a apenas animais ou plantas, tendo assim, questões ainda em aberto, principalmente quanto a origem e diversificação deste gene. Portanto, para atingir nossos objetivos, usamos uma abordagem filogenética com uma ampla amostragem dos genes *P5CS*, *ProB* e *ProA*, incluindo espécies dos três domínios da vida. As sequências foram recuperadas dos bancos de dados genômicos Ensembl e JGI. No total, foram obtidas 479 sequências de *P5CS* de 334 espécies, 661 sequências de *ProB* de 603 espécies e 612 sequências de *ProA* de 598 espécies. Os alinhamentos e análises filogenéticas de Máxima Verossimilhança e Bayesiana foram realizadas separadamente para os genes *ProA*, *ProB* e *P5CS*. Para entender o relacionamento do gene *P5CS* com os homólogos *ProA* e *ProB*, realizamos análises filogenéticas com os domínios GK e GPR separados e incluindo em cada conjunto de dados as sequências dos genes *ProB* e *ProA*. Foram realizadas também análises de seleção com o gene *P5CS*. Em geral, a filogenia estimada dos genes *P5CS*, *ProB* e *ProA*, concordam com a filogenia das espécies. Nas filogenias dos domínios temos a formação de dois grandes clados, mostrando uma clara separação entre o gene P5CS e seus homólogos *ProB/ProA*. Analisando este resultado em conjunto com a árvore das espécies, sugerimos que o gene P5CS descende de um único evento de fusão que ocorreu no ancestral de alguma linhagem eucariótica, e sendo disseminada para outros grupos via transferência horizontal de genes, uma vez que na árvore das espécies a característica “possuir P5CS” forma um grupo polifilético. A topologia encontrada no gene *P5CS* de

plantas nos mostra que estas espécies sofreram diversos processos independentes de duplicação e perdas gênicas. Foram também identificados sítios sob seleção positiva no P5CS das plantas, no qual se encontram diferentes resíduos entre os parálogos, o que pode estar relacionado com a subfuncionalização encontrada em algumas espécies. Foram também identificados sob seleção positiva importantes sítios do P5CS de espécies de animais, no qual se tem registro de mutações com efeitos deletérios em humanos. Nosso trabalho traz novas evidências sobre a origem do gene *P5CS* e da sua diversificação na linhagem dos animais e principalmente na linhagem das plantas.

Abstract

In response to various stresses, the accumulation of proline occurs in bacteria, protozoa, marine invertebrates, and plants. In animals and plants, proline also performs other physiological functions. The main pathway for proline biosynthesis has glutamate as a substrate and occurs through three enzymatic reactions. In prokaryotes, some unicellular eukaryotes and fungi, the first and second reactions are catalyzed by the enzymes γ -GK and γ -GPR, encoded by the *ProB* and *ProA* genes, respectively. In animals, plants and some unicellular eukaryotes, a single gene, called *P5CS*, encodes a bifunctional enzyme with the domains, GK and GPR, responsible for the first and second reactions in proline biosynthesis. In view of the physiological importance of proline and its metabolism in living organisms, the main objective of this work is to elucidate the evolutionary history of the *P5CS* gene family. Other studies are found that address the evolutionary aspects of *P5CS*, they do not use a large sample or are restricted only to animals or plants, remaining doubts, mainly regarding the origin and diversification of this gene. Therefore, to achieve our goals, we use a phylogenetic approach with a wide sampling of the *P5CS*, *ProB* and *ProA* genes, including species from the three domains of life. The sequences were retrieved from the Ensembl and JGI genomic databases. In total, 479 *P5CS* sequences from 334 species, 661 *ProB* sequences from 603 species and 612 *ProA* sequences from 598 species were obtained. The alignments and phylogenetic analysis of Maximum Likelihood and Bayesian were performed separately for the *ProA*, *ProB* and *P5CS* genes. To understand the relationship of the *P5CS* gene with the *ProA* and *ProB* homologs, we performed phylogenetic analyzes with the separate GK and GPR domains and including the sequences of the *ProB* and *ProA* genes in each data set. Selection analyzes were also performed with the *P5CS* gene. In general, the estimated phylogeny of the *P5CS*, *ProB* and *ProA* genes agree with the specie's phylogeny. In the phylogenies of the domains, we have the formation of two large clades, showing a clear separation between the *P5CS* gene and its *ProB/ProA* homologs. Analyzing this result in conjunction with the species tree, we suggest that the *P5CS* gene descends from a single fusion event that occurred in the ancestor of some eukaryotic lineage and being disseminated to other groups via horizontal gene transfer, since in the tree of the species, the characteristic "having *P5CS*" forms a polyphyletic group. The topology found in the plant *P5CS* gene shows us that these species

have undergone several independent processes of duplication and gene losses. Sites with positive selection were also identified in the *P5CS* of the plants, in which different residues are found between the paralogs, which may be related to the subfunctionalization found in some species. Important sites of *P5CS* of animal species were also identified under positive selection, in which mutations with deleterious effects on humans are recorded. Our work brings new evidence about the origin of the *P5CS* gene and its diversification in the lineage of animals and especially in the lineage of plants.

1. Introdução

1.1 Prolina e suas funções

Os aminoácidos são definidos como compostos orgânicos que possuem ligados a um mesmo átomo de carbono um grupamento amino e um grupamento carboxílico, e são conhecidos principalmente pelo seu papel estrutural como componente básico das cadeias peptídicas (Nelson and Cox 2014). Além do papel estrutural, certos aminoácidos desempenham importantes papéis no metabolismo, participando e regulando diversas vias metabólicas necessárias para a manutenção, crescimento, reprodução e imunidade dos organismos. Tais aminoácidos são chamados de aminoácidos funcionais, e estão incluídos nesta categoria a arginina, cisteína, glutamina, leucina, prolina e triptofano (Revisado por Wu, 2009).

Dentre estes aminoácidos daremos enfoque à prolina, que possui algumas singularidades bioquímicas devido ao seu anel de pirrolidina e a ausência de um átomo de hidrogênio nas ligações peptídicas que envolvem seu nitrogênio (Adams and Frank, 1980). No quesito estrutural, a prolina é conhecida por ser um aminoácido essencial na formação do colágeno, que junto com a 4-hidroxiprolina (um aminoácido incomum derivado da prolina), correspondem a cerca de 21% da composição peptídica desta molécula, sendo importante para a torção acentuada da hélice de colágeno (Nelson and Cox 2014). O colágeno corresponde a um terço das proteínas nos mamíferos, sendo a proteína mais abundante no corpo destes animais, tanto que apenas a síntese endógena da prolina se torna insuficiente para o ótimo crescimento e produção de colágeno (revisado por Karna et al., 2020).

Como metabólito livre, a prolina é conhecida pelo seu papel fisiológico em situações de estresse, no qual diversos organismos acumulam a prolina em resposta aos mais variados tipos de estresse. Nas bactérias o acúmulo de prolina é fundamental na resposta ao estresse osmótico (Moses et al., 2012), atuando também como antioxidante em resposta à exposição de metais pesados (Ahad and Syiem, 2021). Em leveduras o acúmulo de prolina está associado com uma maior resistência ao frio (Meng et al., 2021), a dessecação (Takagi et al., 2000) e resistência a espécies reativas de oxigênio (ROS)

(Sasano et al., 2012). Em invertebrados marinhos o acúmulo de prolina ocorre quando estes animais são expostos a ambientes de hipersalinidade (Li et al., 2021). E em plantas, há estudos mostrando o aumento na concentração de prolina em resposta à seca (Harb et al., 2020), alta salinidade (Pundir et al., 2021), altas e baixas temperaturas, estresse oxidativo, radiação UV, contaminação por patógenos (Liang et al., 2013) e por metais pesados (An et al., 2020),

Nas plantas, além da resposta ao estresse, a prolina também está ligada com o desenvolvimento vegetal. Sendo importante para o processo de floração, atuando na transição do meristema apical para o meristema floral, induzindo a coflorescência e atuando no processo de *bolting* (crescimento de um caule floral alto que ocorre em um período muito curto de tempo) (Mattioli et al., 2008), essencial para o desenvolvimento e viabilidade do pólen (Funck et al., 2012), e atua no desenvolvimento da raiz, controlando a divisão celular e modulando o tamanho da zona meristemática radicular (Biancucci et al., 2015).

1.2 Metabolismo da prolina

A via biossintética da prolina é conservada desde bactérias a eucariotos (Fichman et al., 2015) e provavelmente seu surgimento foi um evento antigo na evolução, uma vez que foi proposto um papel fundamental da prolina no mundo de RNA, atuando na ligação entre a química dos carboidratos e aminoácidos (Kubyschkin and Budisa, 2019). A via foi descrita pela primeira vez em *Escherichia coli* que toma como substrato inicial o glutamato, que através de três reações enzimáticas não reversíveis, irá formar a prolina. A via se inicia com a fosforilação do glutamato em γ -glutamil-fosfato, sendo esta reação dependente de ATP e catalisada pela enzima γ -glutamil quinase (γ -GK), codificada pelo gene *ProB*. O γ -Glutamil fosfato é então reduzido para glutamato semialdeído (GSA), utilizando como doador de elétrons o NADPH e catalisada pela γ -glutamil fosfato redutase (γ -GPR), codificada pelo gene *ProA* (Baich, 1969). O GSA produzido se interconverte espontaneamente em Δ^1 -pirrolina-5-carboxilato (P5C), através de uma reação não enzimática de ciclização, existindo ambos os compostos em equilíbrio químico, no qual é favorecido a formação de P5C (Hu et al., 2008a; Nelson and Cox, 2014). A síntese é finalizada pela redução do P5C em prolina, sendo esta reação catalisada pela enzima

Δ^1 -pirrolina-5-carboxilato redutase (P5CR) e utilizando como cofator o NADPH (Ruszkowski et al., 2015) (Figura 1).

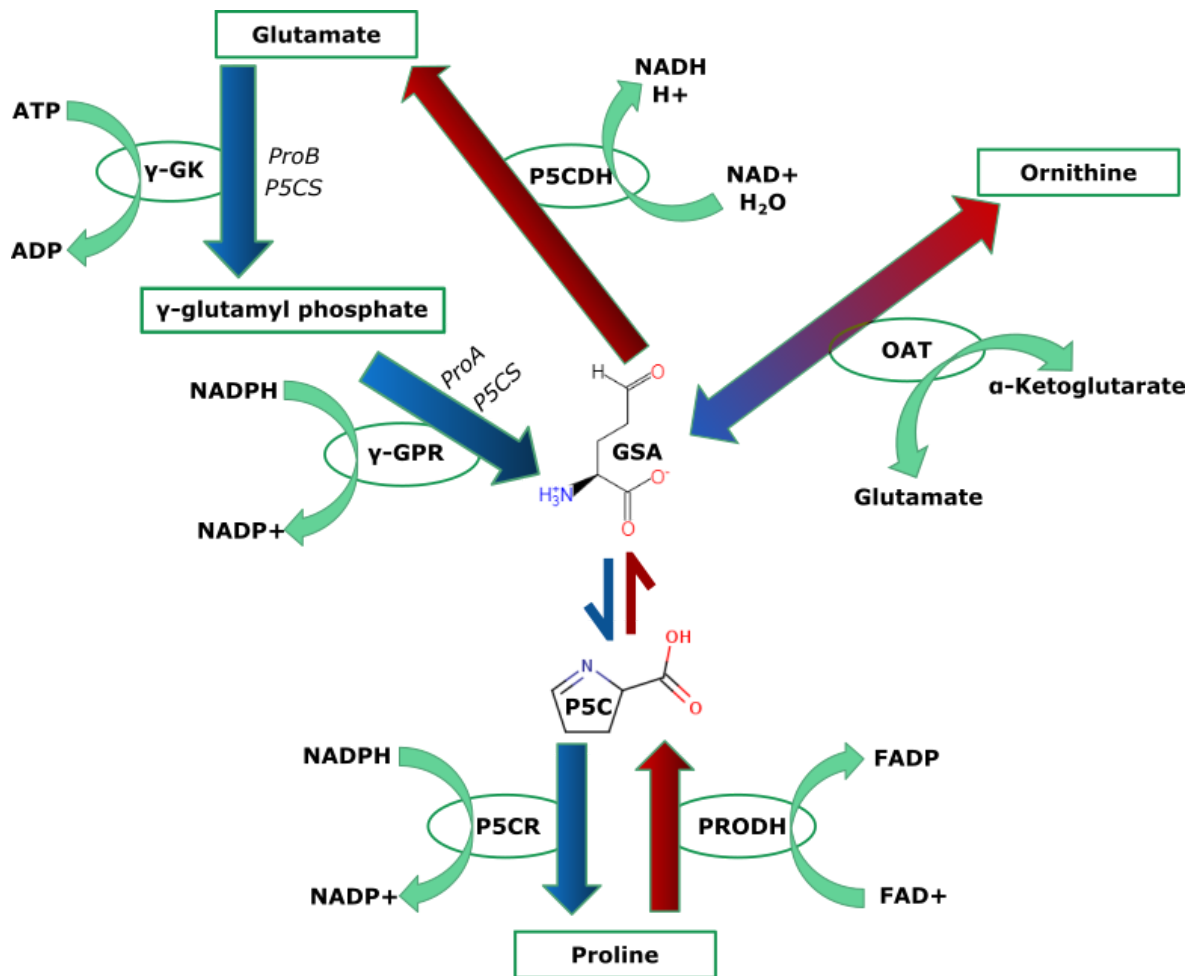


Figura 1 - Metabolismo da prolina, no qual as setas azuis indicam as reações de biossíntese e as vermelhas as reações do catabolismo. As elipses representam as enzimas que catalisam as reações. γ -GK - γ -glutamil quinase; γ -GPR - γ -glutamil fosfato redutase; GSA - glutamato-5-semialdeído; P5C - Δ^1 -pirrolina-5-carboxilato; P5CR - Δ^1 -pirrolina-5-carboxilato redutase; OAT - ornitina aminotransferase; P5CDH - pirrolina-5-carboxilato desidrogenase; PRODH - prolina desidrogenase.

É importante ressaltar que os genes *ProB* e *ProA* são encontrados apenas em bactérias, arqueas e em alguns eucariotos, como algas e fungos (Fichman et al., 2015). Em plantas (Hu et al., 1992) e animais (Smith et al., 1980), temos a presença do gene *P5CS* que codifica a enzima bifuncional, Δ^1 -pirrolina-5-carboxilato sintase (*P5CS*), que possui as porções catalíticas de γ -GK e γ -GPR, realizando assim as duas primeiras reações da síntese da prolina e convertendo o glutamato em GSA.

Embora o glutamato seja o principal substrato para a produção da prolina, a mesma pode ser também sintetizada a partir da ornitina, através de duas reações enzimáticas. Esta via se inicia com a transaminação da ornitina, catalisada pela enzima Ornitina aminotransferase (OAT), que irá transferir o grupamento amino da ornitina para o α -cetoglutarato, formando o GSA e o glutamato, respectivamente. O GSA irá então se interconverter em P5C que como já citado, será reduzido pela enzima P5CR em prolina. Como a reação catalisada pela OAT é reversível, o GSA formado pela P5CS ou GK/GPR, pode ser também utilizado para a produção de ornitina, sendo em mamíferos a única via para a síntese de ornitina (e, portanto, da arginina) quando os níveis de arginina são insuficientes para a síntese proteica (Nelson and Cox, 2014). Portanto o metabolismo da prolina está relacionado diretamente com o metabolismo de outros dois aminoácidos, o glutamato e a arginina (ornitina), se interligando assim a duas outras importantes vias, a do ciclo de Krebs e ciclo da ureia (**Figura 2**) (Hu et al., 2008b).

Em animais, o metabolismo da prolina (e não a prolina em si) desempenha funções importantes na resposta ao estresse e na sinalização celular. Mas para melhor entendimento, será necessário também explicar a via de degradação da prolina, que ocorre através de duas reações enzimáticas. A primeira reação é feita pela enzima prolina oxidase (POX), também conhecida como prolina desidrogenase (PRODH), que catalisa a oxidação da prolina em P5C, utilizando como receptor de elétrons o FAD^+ , como a POX fica localizada na membrana mitocondrial interna, esta primeira reação está acoplada à cadeia transportadora de elétrons, o que resulta na formação direta de ATP e podendo gerar também ROS (Pandhare et al., 2009). A última reação ocorre na matriz mitocondrial, sendo catalisada pela enzima pirrolina-5-carboxilato desidrogenase (P5CDH), que irá oxidar o P5C em glutamato, utilizando o NAD^+ como receptor de elétron (Rizzi et al., 2015).

Sendo o ciclo da prolina, como é conhecida a interconversão de P5C-prolina, importante para a produção de ROS, que atuam como sinalizador no processo de apoptose (Phang et al., 2008). O metabolismo da prolina (tendo como substrato o glutamato) também tem se mostrado importante para se manter a homeostasia oxirredutora das células, se conectando também com a porção oxidativa da via das pentose-fosfato (**Figura 2**), e vem se mostrando importante para o crescimento de tumores (Liu et al., 2015; Tanner et al., 2018). A prolina é também muito utilizada para a produção de ATP nos músculos de

Hymenoptera, moscas tsé-tsé e Mosquitos (Teulier et al., 2016), o que se relaciona também com os tripanosomatídeos, que quando se encontram no corpo de seus vetores, a degradação da prolina é essencial para a geração de ATP e consequente sobrevivência destas espécies (Bringaud, 2012; Mantilla et al., 2017).

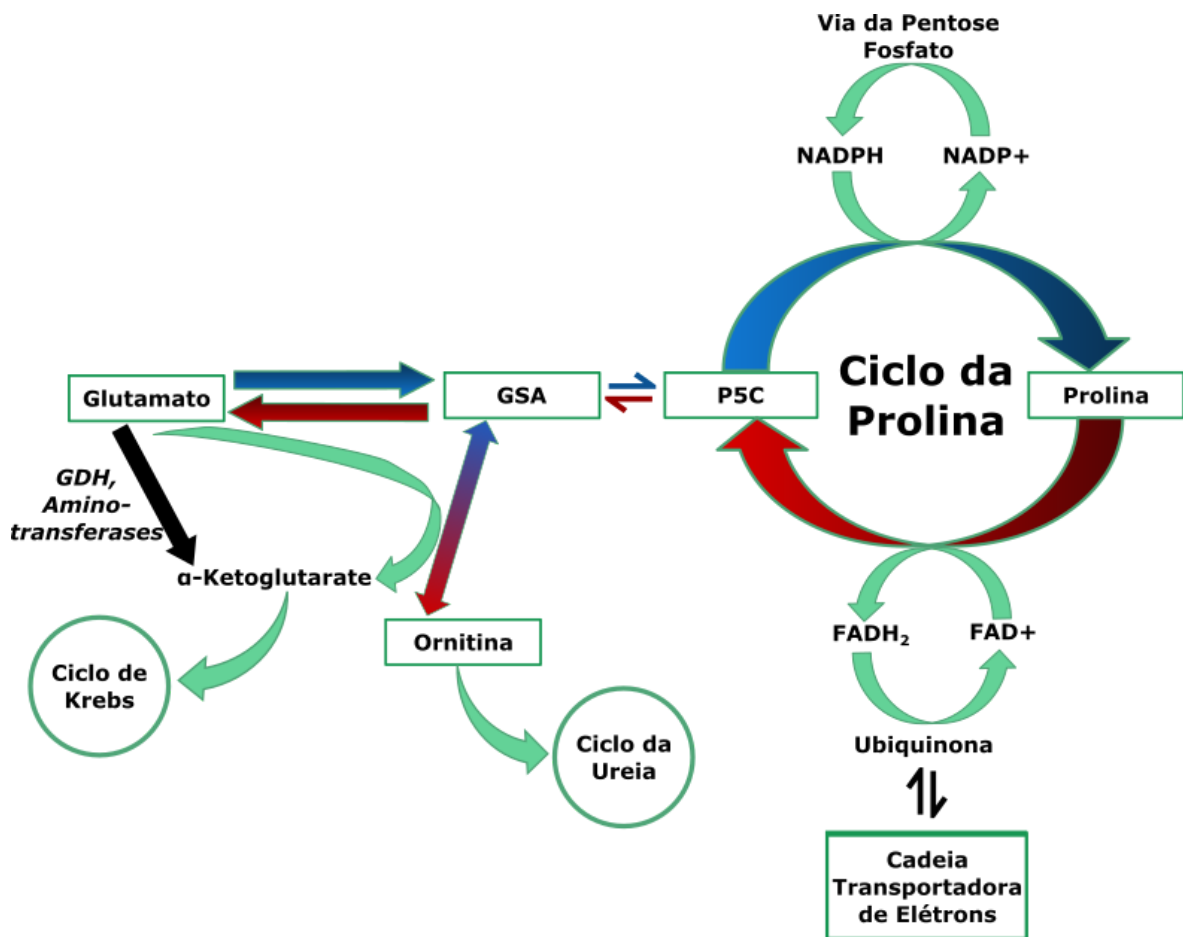


Figura 2 - Representação das interconexões do metabolismo da prolina com outras vias celulares. Localizações intracelulares das vias não são mostradas na imagem. *GDH* - Glutamato desidrogenase; GSA - Glutamato-5-semialdeído; P5C - Pirrolina-5-carboxilato.

O principal ponto de controle da biossíntese de prolina nos organismos ocorre na enzima γ -GK, sendo ela inibida alostericamente pela prolina (Smith et al., 1984). O mesmo vale para as plantas, no qual a prolina se liga na porção do domínio GK da enzima P5CS, atuando como um regulador alostérico e inibindo o anabolismo da prolina (Hu et al., 1992). Em mamíferos a prolina atua como inibidor alostérico da enzima P5CR, sendo a forma nativa da enzima P5CS insensível à uma regulação alostérica pela prolina ou ornitina. Porém, nas células intestinais o mRNA do gene P5CS sofre um processo de

splicing alternativo, codificando uma enzima com menos dois aminoácidos, o que torna o P5CS sensível apenas à ornitina, sendo então inibida alostericamente por ela (Hu et al., 1999). Esse mecanismo evoluiu possivelmente devido ao papel essencial do intestino na produção de citrulina/arginina em neonatos, uma vez que a síntese *de novo* (através do P5C formado pela P5CS) é a principal fonte de ornitina (Hu et al., 2008a; Marini et al., 2012).

1.3 P5CS e seus homólogos γ -GK/ γ -GPR

Tendo em vista que o P5C/GSA é a principal ligação entre o metabolismo do glutamato, prolina e arginina, e que o glutamato é o principal substrato para a síntese *de novo* da prolina (Delauney and Verma, 1993), as enzimas γ -GK e γ -GPR, ou somente a P5CS, são de maior interesse para via. Além de ainda existirem questões em aberto quanto à evolução e diversificação dos genes responsáveis em produzir estas enzimas.

A enzima γ -GK, codificada pelo gene *ProA*, é encontrada na forma de homodímero em *Campylobacter jejuni* (2AKO), enquanto em *Escherichia coli* (Marco-Marín et al., 2007) e *Burkholderia thailandensis* (4Q1T) na forma de homotetrâmero, apesar de experimentos em *E. coli* mostrarem que na forma de homodímeros a enzima já se encontra totalmente ativa (Pérez-Arellano et al., 2010). Para leveduras não se tem estudos mostrando a formação de dímeros ou se a enzima é funcional na forma de monômero.

As enzimas γ -GK além do domínio GK, possuem um outro domínio denominado PUA (pseudouridina sintase e a arqueosina transglicosilase). O domínio PUA geralmente se encontra associado a enzimas que catalisam modificações pós-transcricionais em tRNA e rRNA. Mas o domínio PUA não se encontra presente em todas as enzimas GK, cerca de 20% das leveduras e bactérias naturalmente não o possuem, o que sugere que a presença do domínio PUA não é essencial para o funcionamento da enzima γ -GK (Pérez-Arellano et al., 2007). Entretanto, em *E. coli*, na ausência do domínio PUA a enzima γ -GK necessitou do dobro de prolina para ser inibida e apresentou uma menor estabilidade a variações de temperatura, quando comparada com a forma selvagem de sua enzima (Pérez-Arellano et al., 2005). Um estudo de viabilidade em *Saccharomyces cerevisiae* mostrou que o domínio PUA em si não é necessário para a atividade catalítica da enzima, mas sim a região que liga os dois domínios na enzima γ -GK, atuando possivelmente na conformação/estabilidade da enzima (Kaino et al., 2012). Já foi proposto a participação do

domínio PUA na regulação/estabilidade do mRNA do gene *ProB*, mas a função deste domínio na enzima γ -GK ainda não foi completamente compreendida (Fichman et al., 2015).

A enzima γ -GPR também apresenta formação de estruturas oligoméricas, sendo encontrada na forma de homotetrâmero nas espécies em *B. thailandensis* (4GHK) e *Thermotoga maritima* (Page et al., 2003), enquanto em *S. cerevisiae* a estrutura encontrada no *Protein Data Bank* é de um homodímero (1VLU). Mas isso é esperado para γ -GPR, uma vez que ela faz parte da superfamília ALDH (aldeído desidrogenase), sendo característico a presença de um domínio de oligomerização (Stiti et al., 2021). Como a enzima P5CS possui uma região homóloga a γ -GPR, ela também é um membro desta superfamília, com seu gene sendo também nomeado como ALDH18. Para animais é mais comum encontrar trabalhos com a nomenclatura ALDH18 ao invés de P5CS.

Alguns estudos propuseram que as enzimas γ -GK e γ -GPR podem formar um complexo *in vivo*, sendo essencial para atividade catalítica das mesmas, pois ensaios com *E. coli* mostraram que a enzima γ -GK só se encontra ativa com a presença da enzima γ -GPR (Smith et al., 1984; Kaino et al., 2012). A capacidade destas enzimas formarem complexos parece ocorrer até mesmo com enzimas de diferentes organismos, pois foi detectada atividade catalítica da γ -GK de levedura quando expressa em *E. coli* (Kaino et al., 2012). A formação de um complexo γ -GK- γ -GPR, pode possibilitar a canalização de substrato (no qual o produto de uma reação é transportado para um segundo sítio ativo, sem se difundir no meio), que seria vantajoso devido à alta instabilidade do γ -glutamil fosfato, formado pela γ -GK (Arentson et al., 2012). Um modelo de como seria este complexo foi proposto, no qual ele seria composto por um homotetrâmero de γ -GK interagindo com um dímero de γ -GPR. Este modelo se reforça, pois, a cristalografia da enzima γ -GK revelou que o γ -glutamil fosfato pode permanecer intacto no sítio ativo da γ -GK (Marco-Marín et al., 2007).

A hipótese para a origem da enzima bifuncional P5CS, é que ela tenha surgido pela fusão dos genes *ProA* e *ProB* (**Figura 3**). No entanto, ainda não está claro se este foi um único evento que se espalhou por meio de transferência horizontal de genes ou se ocorreu várias vezes de forma independente no ancestral de cada grupo (Fichman et al., 2015). Fusões gênicas ocorrem quando duas ou mais ORFs se unem em uma só, e podem ocorrer por translocações ou inversões cromossômicas ou também por deleções intersticiais

(Leonard and Richards, 2012). No trabalho de Enright et al., 1999, a maioria das proteínas fusionadas identificadas se tratava de enzimas metabólicas, e hipotetizaram que isso poderia ser devido ao efeito de canalização de substrato. Efeito este, que como discutido anteriormente, parece ser vantajoso para a reação GK e GPR, o que pode ter levado a fixação deste gene fusionado nos animais, plantas e alguns eucariotos unicelulares.

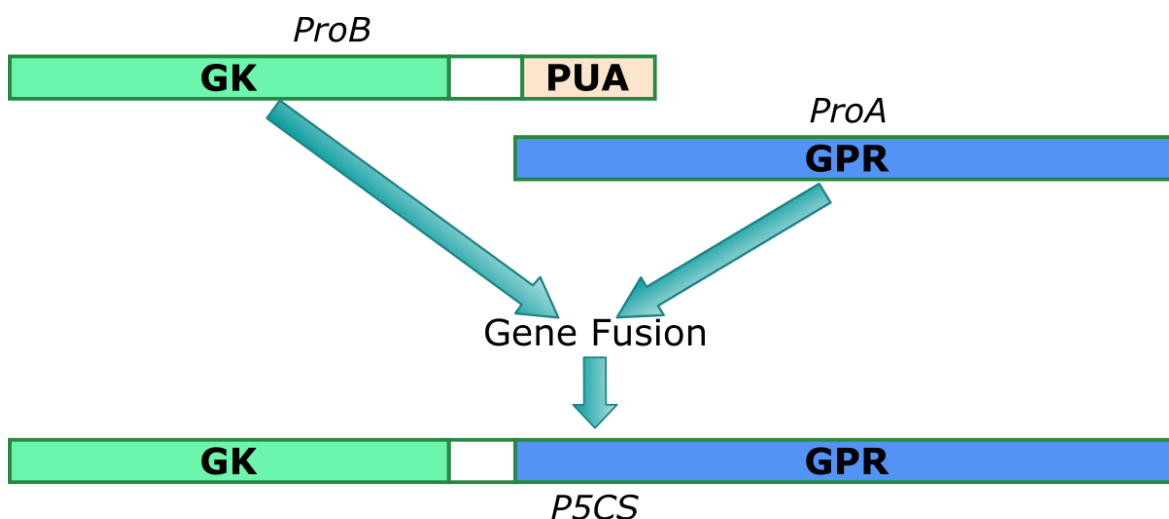


Figura 3 - Representação dos domínios encontrados nos genes *ProB*, *ProA* e sua homologia com os domínios encontrados no gene *P5CS*. A principal hipótese para a origem do gene *P5CS* é que ela tenha se originado de um evento de fusão entre os genes *ProB* e *ProA*, sendo o domínio PUA perdido durante este evento.

Apesar disso, ainda não existem estudos que comprovem que a canalização de substrato de fato ocorre na enzima *P5CS*, ou se apenas a proximidade dos dois sítios ativos já provoca uma maior eficiência da via (Arentson et al., 2012). A fusão artificial dos genes *ProB/ProA* em *E. coli* mostrou uma maior atividade enzimática e produção de prolina quando comparada com a cepa selvagem que expressa as enzimas por dois genes distintos (Chen et al., 2006). A fusão dos genes *ProB/ProA* em *S. cerevisiae* também se mostrou funcional, entretanto foi analisada apenas a atividade do domínio GK e no qual a forma fusionada apresentou uma menor atividade (cerca de 70%) quando comparado com a forma isolada da enzima γ -GK. Eles também analisaram a fusão dos genes com a deleção do domínio PUA do gene *ProB*, como ocorre na forma bifuncional *P5CS* em que não encontramos o domínio PUA, porém a atividade da enzima caiu para 0,58% da forma selvagem da γ -GK (Kaino et al., 2012).

Além disso, a fusão gênica é também um fator que garante a co-localização e a co-expressão dos domínios (Lees et al., 2016), o que pode facilitar também o direcionamento para organelas específicas como mitocôndria e cloroplasto. Como ocorre com o P5CS de mamíferos que se encontra na membrana mitocondrial interna (Hu et al., 2008a). Um estudo inicial também mostrou um direcionamento específico no P5CS de plantas sendo localizadas no citoplasma e em situações de estresse se localizava também no cloroplasto, apesar de não ter sido encontrado nenhuma sequência sinal em sua cadeia peptídica. (Székely et al., 2008). Entretanto, um novo estudo contradiz este achado, no qual seus resultados mostram uma localização exclusivamente citoplasmática do P5CS (Funck et al., 2020).

Além da localização intracelular e do inibidor alostérico, o gene *P5CS* de animais e plantas possuem diferenças quanto ao número de parálogos que são encontrados em seus genomas. Nas angiospermas, é comum encontrar pelo menos dois genes que codificam a enzima P5CS, e no qual um estudo anterior mostrou que possivelmente estes parálogos surgiram de vários processos independentes de duplicação (Turchetto-Zolet et al., 2009). Enquanto em animais geralmente encontramos um único gene (Holmes, 2017).

Estudos funcionais com *Arabidopsis thaliana* demonstraram haver certa diferenciação entre as funções de seus dois genes parálogos, sendo o *AtP5CS1* mais responsivo ao estresse e o *AtP5CS2* se expressa de forma constitutiva, atuando mais no desenvolvimento da planta (Székely et al., 2008; Funck et al., 2020). Na família Poaceae, trabalhos com *Oryza sativa* e *Sorghum bicolor* também mostraram que seus parálogos possuem diferentes padrões de expressão nos tecidos e que podem desempenhar papéis não redundantes no desenvolvimento das plantas (Hur et al., 2004; Su et al., 2011). A subfuncionalização parece ocorrer também entre os parálogos da bactéria *Bacillus subtilis*, que possui dois genes codificantes da enzima γ -GK, sendo a expressão de um deles induzida por estresse osmótico (Brill et al., 2011; Hoffmann et al., 2017).

Tendo em vista a importância da prolina para todos os organismos vivos, a compreensão sobre a história evolutiva dos genes envolvidos com a biossíntese desse aminoácido torna-se de extrema relevância. Embora estudos prévios tenham abordado alguns aspectos evolutivos da família gênica P5CS, que codifica a principal enzima da via, algumas questões sobre a origem e diversificação dessa família gênica ainda permanecem sem resposta. Sendo assim, esta dissertação visa melhor estabelecer o relacionamento

filogenético do gene *P5CS* com seus homólogos monofuncionais *ProB/ProA*, através de uma ampla amostragem abordando os três domínios da vida (Archaea, Bacteria e Eukarya). Quanto a origem do gene *P5CS*, nós hipotetizamos que se deu por apenas um único evento de fusão que ocorreu em uma das linhagens iniciais dos eucariotos, sendo disseminada via transferência horizontal de genes, entre os ancestrais dos animais, plantas e dos eucariotos unicelulares que o possuem. Sendo esperado nesta hipótese, encontrar nas filogenias todas as sequências do gene *P5CS* agrupadas formando um único clado e seus homólogos *ProA/ProB* formando um outro clado independente. Para o padrão de duplicação em plantas, como suposto em um trabalho anterior (Turchetto-Zolet et al., 2009), hipotetizamos que ocorreram múltiplos processos independentes de duplicação do gene *P5CS*. Sendo esperado encontrar na filogenia das plantas, diversos clados distintos com todos as sequências das espécies que compartilham uma duplicação. Realizamos também testes de seleção positiva, tentando identificar sítios importantes que podem estar relacionados com as subfuncionalizações encontradas entre os parálogos de certas espécies de plantas e sítios importantes para a diferenciação da regulação alostérica encontrada em animais.

2. Objetivos

2.1 Objetivo geral

O principal objetivo deste trabalho é elucidar a relação evolutiva do gene *P5CS* com seus homólogos monofuncionais *ProB* e *ProA*, com enfoque em sua origem e diversificação.

2.2 Objetivos específicos

- Identificar homólogos de *P5CS*, *ProB* e *ProA* no genoma de representantes de todos os domínios da árvore da vida, e verificar a partir de quais grupos encontramos o gene *P5CS*;
- Identificar a origem do gene *P5CS* e compreender o relacionamento com os genes ancestrais *ProA* e *ProB*;
- Compreender a diversificação do gene *P5CS* em plantas e o relacionamento de seus parálogos;
- Revelar novos *insights* sobre a evolução do gene *P5CS* durante a diversificação dos animais e plantas;
- Verificar sinais de seleção positiva no gene *P5CS* dentro das plantas e entre as sequências de animais e das plantas.

3. Capítulo I - Origin and diversification of *P5CS* gene family

João Pedro do Carmo Filgueiras e Andreia Carina Turchetto-Zolet

O primeiro e único capítulo desta dissertação está apresentado no formato de artigo científico, que se encontra em preparação, e será posteriormente submetido a revista *Molecular Phylogenetics and Evolution*, contemplando os objetivos da dissertação.

4. Considerações finais

A prolina é um aminoácido no qual bactérias, fungos, plantas e invertebrados marinhos, aumentam sua produção e a acumulam em resposta aos mais diversos tipos de estresse. Além disso, atua no desenvolvimento vegetal e seu metabolismo em animais cada vez mais vêm se mostrando importante, ligando-se a vias essenciais como do Ciclo de Krebs e da ureia e atuando na sinalização celular e na apoptose. A principal via para a síntese da prolina tem como substrato o glutamato e ocorre através de três reações enzimáticas. No qual em alguns organismos as duas primeiras reações ocorrem através de duas enzimas (GK e GPR), codificadas por dois genes distintos (*ProB* e *ProA*, respectivamente). Em outros organismos como plantas e animais elas ocorrem por meio de uma única enzima bifuncional (possuindo os domínios GK e GPR), denominada P5CS, que surgiu provavelmente pela fusão dos genes *ProB* e *ProA*, apesar de ainda não se ter claro em que momento ocorreu e se foi através de um único ou vários eventos de fusão.

A forma bifuncional da enzima P5CS, em acordo com outros estudos, é exclusiva dos eucariotos, presente nos animais, plantas, Stramenopiles, Rhizaria e *Acanthamoeba*. Enquanto as enzimas monofuncionais, são encontradas nas arqueas, bactérias, algas verdes e vermelhas, nos Alveolatas e nos fungos. Nas filogenias dos domínios, combinando o *P5CS* com seus homólogos *ProB* e *ProA*, temos a formação de dois grandes clados, um formado pelas sequências do gene P5CS e o outro pelos genes *ProA* e *ProB*, mostrando uma clara separação entre a enzima bifuncional dos seus homólogos monofuncionais. Resultado que nos indica uma origem comum para o gene P5CS, sendo provável então que tenha surgido de um único evento de fusão. Entretanto, quando analisamos as relações filogenéticas das espécies não temos um padrão monofilético de distribuição do gene *P5CS*, sendo então disseminada entre as diferentes linhagens através da transferência horizontal de genes. Nas linhagens das algas verdes, tripanosomatídeos e oomicetos, foram identificados um ou os dois genes monofuncionais. Baseado em nossa hipótese e nas filogenias estimadas, sugerimos que estas linhagens sofreram deleções em um dos domínios da enzima bifuncional P5CS, sendo então a enzima GPR das algas e tripanosomatídeos, codificada por um gene P5CS incompleto e não por um gene *ProA*, e para a enzima GK dos oomicetos, seguimos a mesma lógica.

Quanto à diversificação do gene P5CS nas plantas, nossos resultados indicam que ocorreram múltiplos processos independentes de duplicação e perda do gene P5CS neste

grupo. Rastreamos os grupos de parálogos que possuem uma origem comum, notamos uma forte associação com antigos eventos de duplicação de genoma completo e de poliploidias. Sugerimos então que esta possa ser a principal fonte geradora de parálogos do gene P5CS nas plantas. Análises de seleção revelaram que 11 códons do gene *P5CS* estão evoluindo sob seleção positiva, sendo que alguns destes possuem uma diferenciação de resíduos entre os grupos de parálogos. Estes sítios podem estar envolvidos na subfuncionalização encontrada em algumas espécies, no qual um parálogo atua mais na resposta ao estresse e o outro está ligado a desenvolvimento vegetal. Estudos com mutações pontuais nestes resíduos, seriam interessantes para comprovar se eles atuam ou não na subfuncionalização destes parálogos.

O P5CS nos animais não possui uma diversificação quanto ao número de genes encontrado em seus genomas, como observado nas plantas, a maioria das espécies amostradas possui apenas um único gene. Apenas a linhagem dos Teleostei possui dois genes codificantes da enzima P5CS. Porém, quanto a funcionalidade, na linhagem animal encontramos uma importante diversificação, pois sua porção GK do P5CS, é a única a não ser inibida alostericamente pela prolina. Enquanto a enzima γ -GK e o domínio GK do P5CS vegetal, são todos responsivos a prolina. Na linhagem dos animais foram identificadas 12 códon do gene P5CS que estão evoluindo sob seleção positiva. Destes, um que merece destaque é o sítio T221 (tendo como referência o P5CS de *Homo sapiens*), pois este é um sítio altamente conservado, sendo os animais os únicos a não possuírem um glutamato neste sítio, até mesmo quando olhamos a nível de domínio e comparamos também com o gene *ProB*. Este sítio na enzima γ -GK de *Escherichia coli*, está envolvido na interação com a prolina. Podendo então, ser uma das variações que permitiram a perda da capacidade da enzima P5CS dos animais de serem inibida pela prolina.

Sendo assim, a dissertação cumpriu com seus principais objetivos trazendo *insights* sobre a origem e a diversificação do gene P5CS em animais e principalmente em plantas. Nosso trabalho também identificou sítios sob seleção positiva, que mostram variações internas ou externas ao grupo no qual foram identificadas. Abrindo assim espaço para estudos funcionais através de mutações de ponto, para melhor entender a estrutura e ação da enzima P5CS. Para alguns sítios, já é encontrado na literatura registros e estudos dos impactos causados por mutações não-sinônimas.

5. Referências Bibliográficas

- Adams E and Frank L (1980) Metabolism of Proline and the Hydroxyprolines. *Annu Rev Biochem* 49:1005–1061. doi: 10.1146/annurev.bi.49.070180.005041
- Ahad RIA and Syiem MB (2021) Analyzing dose dependency of antioxidant defense system in the cyanobacterium *Nostoc muscorum* Meg 1 chronically exposed to Cd²⁺. *Comp Biochem Physiol Part C Toxicol Pharmacol* 242:108950. doi: 10.1016/j.cbpc.2020.108950
- Arentson BW, Sanyal N and Becker DF (2012) Substrate channeling in proline metabolism. *Front Biosci* 17:375. doi: 10.2741/3932
- An Q, He X, Zheng N, Hou S, Sun S, Wang S, Li P, Li X and Song X (2020) Physiological and genetic effects of cadmium and copper mixtures on carrot under greenhouse cultivation. *Ecotoxicol Environ Saf* 206:111363. doi: 10.1016/j.ecoenv.2020.111363
- Baich A (1969) Proline synthesis in *Escherichia coli* a proline-inhibitable glutamic acid kinase. *Biochim Biophys Acta - Gen Subj* 192:462–467. doi: 10.1016/0304-4165(69)90395-X
- Biancucci M, Mattioli R, Moubayidin L, Sabatini S, Costantino P and Trovato M (2015) Proline affects the size of the root meristematic zone in *Arabidopsis*. *BMC Plant Biol* 15:263. doi: 10.1186/s12870-015-0637-8
- Brill J, Hoffmann T, Putzer H and Bremer E (2011) T-box-mediated control of the anabolic proline biosynthetic genes of *Bacillus subtilis*. *Microbiology* 157:977–987. doi: 10.1099/mic.0.047357-0
- Bringaud F (2012) Multiple roles of proline transport and metabolism in trypanosomatids. *Front Biosci* 17:349. doi: 10.2741/3931
- Chen M, Cao J, Zheng C and Liu Q (2006) Directed evolution of an artificial bifunctional enzyme, γ -glutamyl kinase/ γ -glutamyl phosphate reductase, for improved osmotic tolerance of *Escherichia coli* transformants. *FEMS Microbiol Lett* 263:41–47. doi: 10.1111/j.1574-6968.2006.00397.x
- Delauney AJ and Verma DPS (1993) Proline biosynthesis and osmoregulation in plants. *Plant J* 4:215–223. doi: 10.1046/j.1365-313X.1993.04020215.x
- Enright AJ, Iliopoulos I, Kyrpidis NC and Ouzounis CA (1999) Protein interaction maps for complete genomes based on gene fusion events. *Nature* 402:86–90. doi: 10.1038/47056
- Fichman Y, Gerdes SY, Kovács H, Szabados L, Zilberstein A and Csonka LN (2015) Evolution of proline biosynthesis: enzymology, bioinformatics, genetics, and transcriptional regulation. *Biol Rev* 90:1065–1099. doi: 10.1111/brv.12146
- Funck D, Winter G, Baumgarten L and Forlani G (2012) Requirement of proline synthesis during *Arabidopsis* reproductive development. *BMC Plant Biol* 12:191. doi: 10.1186/1471-2229-12-191
- Funck D, Baumgarten L, Stift M, von Wirén N and Schönemann L (2020) Differential Contribution of P5CS Isoforms to Stress Tolerance in *Arabidopsis*. *Front Plant Sci*. doi: 10.3389/fpls.2020.565134
- Harb A, Simpson C, Guo W, Govindan G, Kakani VG and Sunkar R (2020) The Effect of Drought on Transcriptome and Hormonal Profiles in Barley Genotypes With Contrasting Drought Tolerance. *Front Plant Sci*. doi: 10.3389/fpls.2020.618491

- Hoffmann T, Bleisteiner M, Sappa PK, Steil L, Mäder U, Völker U and Bremer E (2017) Synthesis of the compatible solute proline by *Bacillus subtilis* : point mutations rendering the osmotically controlled proHJ promoter hyperactive. *Environ Microbiol* 19:3700–3720. doi: 10.1111/1462-2920.13870
- Holmes RS (2017) Comparative and evolutionary studies of ALDH18A1 genes and proteins. *Chem Biol Interact* 276:2–8. doi: 10.1016/j.cbi.2016.12.012
- Hu CA, Delauney AJ and Verma DP (1992) A bifunctional enzyme (Δ^1 -pyrroline-5-carboxylate synthetase) catalyzes the first two steps in proline biosynthesis in plants. *Proc Natl Acad Sci* 89:9354–9358. doi: 10.1073/pnas.89.19.9354
- Hu CA, Lin W-W, Obie C and Valle D (1999) Molecular Enzymology of Mammalian Δ^1 -Pyrroline-5-carboxylate Synthase. *J Biol Chem* 274:6754–6762. doi: 10.1074/jbc.274.10.6754
- Hu CA, Khalil S, Zhaorigetu S, Liu Z, Tyler M, Wan G and Valle D (2008a) Human Δ^1 -pyrroline-5-carboxylate synthase: function and regulation. *Amino Acids* 35:665–672. doi: 10.1007/s00726-008-0075-0
- Hu CA, Phang JM and Valle D (2008b) Proline metabolism in health and disease. *Amino Acids* 35:651–652. doi: 10.1007/s00726-008-0102-1
- Hur J, Jung K-H, Lee C-H and An G (2004) Stress-inducible OsP5CS2 gene is essential for salt and cold tolerance in rice. *Plant Sci* 167:417–426. doi: 10.1016/j.plantsci.2004.04.009
- Kaino T, Tasaka Y, Tatehashi Y and Takagi H (2012) Functional Analysis of the C-Terminal Region of γ -Glutamyl Kinase of *Saccharomyces cerevisiae*. *Biosci Biotechnol Biochem* 76:454–461. doi: 10.1271/bbb.110682
- Karna E, Szoka L, Huynh TYL and Palka JA (2020) Proline-dependent regulation of collagen metabolism. *Cell Mol Life Sci* 77:1911–1918. doi: 10.1007/s00018-019-03363-3
- Kubyshekin V and Budisa N (2019) The Alanine World Model for the Development of the Amino Acid Repertoire in Protein Biosynthesis. *Int J Mol Sci* 20:5507. doi: 10.3390/ijms20215507
- Lees JG, Dawson NL, Sillitoe I and Orengo CA (2016) Functional innovation from changes in protein domains and their combinations. *Curr Opin Struct Biol* 38:44–52. doi: 10.1016/j.sbi.2016.05.016
- Leonard G and Richards TA (2012) Genome-scale comparative analysis of gene fusions, gene fissions, and the fungal tree of life. *Proc Natl Acad Sci* 109:21402–21407. doi: 10.1073/pnas.1210909110
- Li Y, Niu D, Wu Y, Dong Z and Li J (2021) Integrated analysis of transcriptomic and metabolomic data to evaluate responses to hypersalinity stress in the gill of the razor clam (*Sinonovacula constricta*). *Comp Biochem Physiol Part D Genomics Proteomics* 38:100793. doi: 10.1016/j.cbd.2021.100793
- Liang X, Zhang L, Natarajan SK and Becker DF (2013) Proline Mechanisms of Stress Survival. *Antioxid Redox Signal* 19:998–1011. doi: 10.1089/ars.2012.5074
- Liu W, Hancock CN, Fischer JW, Harman M and Phang JM (2015) Proline biosynthesis augments tumor cell growth and aerobic glycolysis: involvement of pyridine nucleotides. *Sci Rep* 5:17206. doi: 10.1038/srep17206
- Mantilla BS, Marchese L, Casas-Sánchez A, Dyer NA, Ejeh N, Biran M, Bringaud F, Lehane MJ, Acosta-Serrano A and Silber AM (2017) Proline Metabolism is Essential

- for *Trypanosoma brucei brucei* Survival in the Tsetse Vector. *PLOS Pathog* 13:e1006158. doi: 10.1371/journal.ppat.1006158
- Marco-Marín C, Gil-Ortiz F, Pérez-Arellano I, Cervera J, Fita I and Rubio V (2007) A Novel Two-domain Architecture Within the Amino Acid Kinase Enzyme Family Revealed by the Crystal Structure of *Escherichia coli* Glutamate 5-kinase. *J Mol Biol* 367:1431–1446. doi: 10.1016/j.jmb.2007.01.073
- Marini JC, Stoll B, Didelija IC and Burrin DG (2012) De novo synthesis is the main source of ornithine for citrulline production in neonatal pigs. *Am J Physiol Metab* 303:E1348–E1353. doi: 10.1152/ajpendo.00399.2012
- Mattioli R, Marchese D, D'Angeli S, Altamura MM, Costantino P and Trovato M (2008) Modulation of intracellular proline levels affects flowering time and inflorescence architecture in *Arabidopsis*. *Plant Mol Biol* 66:277–288. doi: 10.1007/s11103-007-9269-1
- Meng L, Yang X, Lin X, Jiang H-Y, Hu X-P and Liu S-X (2021) Effect of overexpression of SNF1 on the transcriptional and metabolic landscape of baker's yeast under freezing stress. *Microb Cell Fact* 20:10. doi: 10.1186/s12934-020-01503-0
- Moses S, Sinner T, Zapras A, Stoveken N, Hoffmann T, Belitsky BR, Sonenshein AL and Bremer E (2012) Proline Utilization by *Bacillus subtilis*: Uptake and Catabolism. *J Bacteriol* 194:745–758. doi: 10.1128/JB.06380-11
- Nelson DL and Cox MM (2014) *Principios de bioquímica de Lehninger*, 6th ed. Artmed, Porto Alegre.
- Page R, Nelson MS, von Delft F, Elsliger M-A, Canaves JM, Brinen LS, Dai X, Deacon AM, Floyd R, Godzik A et al. (2003) Crystal structure of γ -glutamyl phosphate reductase (TM0293) from *Thermotoga maritima* at 2.0 Å resolution. *Proteins Struct Funct Bioinforma* 54:157–161. doi: 10.1002/prot.10562
- Pandhare J, Donald SP, Cooper SK and Phang JM (2009) Regulation and function of proline oxidase under nutrient stress. *J Cell Biochem* 107:759–768. doi: 10.1002/jcb.221747
- Pérez-Arellano I, Rubio V and Cervera J (2005) Dissection of *Escherichia coli* glutamate 5-kinase: Functional impact of the deletion of the PUA domain. *FEBS Lett* 579:6903–6908. doi: 10.1016/j.febslet.2005.11.037
- Pérez-Arellano I, Gallego J and Cervera J (2007) The PUA domain – a structural and functional overview. *FEBS J* 274:4972–4984. doi: 10.1111/j.1742-4658.2007.06031.x
- Pérez-Arellano I, Carmona-Álvarez F, Gallego J and Cervera J (2010) Molecular Mechanisms Modulating Glutamate Kinase Activity. Identification of the Proline Feedback Inhibitor Binding Site. *J Mol Biol* 404:890–901. doi: 10.1016/j.jmb.2010.10.019
- Phang JM, Pandhare J and Liu Y (2008) The Metabolism of Proline as Microenvironmental Stress Substrate. *J Nutr* 138:2008S-2015S. doi: 10.1093/jn/138.10.2008S
- Pundir P, Devi A, Krishnamurthy SL, Sharma PC and Vinaykumar NM (2021) QTLs in salt rice variety CSR10 reveals salinity tolerance at reproductive stage. *Acta Physiol Plant* 43:35. doi: 10.1007/s11738-020-03183-0
- Rizzi YS, Monteoliva MI, Fabro G, Grosso CL, Laróvere LE and Alvarez ME (2015) P5CDH affects the pathways contributing to Pro synthesis after ProDH activation by biotic and abiotic stress conditions. *Front Plant Sci* 6:759–768. doi: 10.3389/fpls.2015.00572

- Ruszkowski M, Nocek B, Forlani G and Dauter Z (2015) The structure of *Medicago truncatula* δ 1-pyrroline-5-carboxylate reductase provides new insights into regulation of proline biosynthesis in plants. *Front Plant Sci*. doi: 10.3389/fpls.2015.00869
- Sasano Y, Haitani Y, Ohtsu I, Shima J and Takagi H (2012) Proline accumulation in baker's yeast enhances high-sucrose stress tolerance and fermentation ability in sweet dough. *Int J Food Microbiol* 152:40–43. doi: 10.1016/j.ijfoodmicro.2011.10.004
- Smith RJ, Downing SJ, Phang JM, Lodato RF and Aoki TT (1980) Pyrroline-5-carboxylate synthase activity in mammalian cells. *Proc Natl Acad Sci* 77:5221–5225. doi: 10.1073/pnas.77.9.5221
- Smith RJ, Deutch AH and Rushlow KE (1984) Purification and characteristics of a gamma-glutamyl kinase involved in *Escherichia coli* proline biosynthesis. *J Bacteriol* 158:545–551.
- Stiti N, Giarola V and Bartels D (2021) From algae to vascular plants: The multistep evolutionary trajectory of the ALDH superfamily towards functional promiscuity and the emergence of structural characteristics. *Environ Exp Bot* 185:104376. doi: 10.1016/j.envexpbot.2021.104376
- Su M, Li X-F, Ma X-Y, Peng X-J, Zhao A-G, Cheng L-Q, Chen S-Y and Liu G-S (2011) Cloning two P5CS genes from bioenergy sorghum and their expression profiles under abiotic stresses and MeJA treatment. *Plant Sci* 181:652–659. doi: 10.1016/j.plantsci.2011.03.002
- Székely G, Ábrahám E, Cséplő Á, Rigó G, Zsigmond L, Csiszár J, Ayaydin F, Strizhov N, Jásik J, Schmelzer E et al. (2008) Duplicated P5CS genes of *Arabidopsis* play distinct roles in stress regulation and developmental control of proline biosynthesis. *Plant J* 53:11–28. doi: 10.1111/j.1365-3113.2007.03318.x
- Takagi H, Sakai K, Morida K and Nakamori S (2000) Proline accumulation by mutation or disruption of the proline oxidase gene improves resistance to freezing and desiccation stresses in *Saccharomyces cerevisiae*. *FEMS Microbiol Lett* 184:103–108. doi: 10.1111/j.1574-6968.2000.tb08998.x
- Tanner JJ, Fendt S-M and Becker DF (2018) The Proline Cycle As a Potential Cancer Therapy Target. *Biochemistry* 57:3433–3444. doi: 10.1021/acs.biochem.8b00215
- Teulier L, Weber J-M, Crevier J and Darveau C-A (2016) Proline as a fuel for insect flight: enhancing carbohydrate oxidation in hymenoptera. *Proc R Soc B Biol Sci* 283:20160333. doi: 10.1098/rspb.2016.0333
- Turchetto-Zolet AC, Margis-Pinheiro M and Margis R (2009) The evolution of pyrroline-5-carboxylate synthase in plants: a key enzyme in proline synthesis. *Mol Genet Genomics* 281:87–97. doi: 10.1007/s00438-008-0396-4
- Wu G (2009) Amino acids: metabolism, functions, and nutrition. *Amino Acids* 37:1–17. doi: 10.1007/s00726-009-0269-0