Original article

Development of stages-specific markers for in vitro morphogenetic effects by genes in silico candidates on fruit species of the Cerrado

Desenvolvimento de marcadores estádios-específicos para efeitos morfogenéticos in vitro por genes candidatos in silico em espécies frutíferas do cerrado

Luana Cristina Diniz Santos1, Victor Soares Rodrigues1, Márcio Antônio Pimenta1, Sônia Ribeiro Arrudas1 e Guilherme Araújo Lacerda1

1State University of Montes Claros (Unimontes), Montes Claros-MG, Brazil.

Abstract

Objective: to analyze the silicon candidates for the genes of interest SERK, BBM and psbA and their possible occurrence in Cerrado species. Materials and Methods: three different genes were tested as a way of expanding the studies and so that it was possible to frame which gene contains more sequences related to Cerrado fruits. Results: the sequences of interest of the fruits of the Cerrado Barbatimão (Stryphnodendron Mart.), Buriti (Mauritia flexuosa L.f.), Cagaita (Eugenia dysenterica (Mart.) DC.), Cajuzinho (Brysonima intermedia A.Juss.), Guacatonga (Casearia syestris Sw.), Jenipapo (Genipa americana L.), Joá-bravo (Solanum viarum Dunal), Pau-terra (Qualea grandiflora Mart.), Pequi (Caryocar brasiliense Cambess.), all identified with their respective botanical descriptors. Conclusion: enables discussion and research focus supporting the development of this analytical analysis of such analysis: histo-analysts of processes and molecular-specific processes of embryogenesis, as a strategy for the study of zygotic and somatic, as well as organogenic for agronomic and native species.

Key words: Somatic embryogenesis - related kinase. Babyboom. Photosystem II protein D1.

Resumo

Objetivo: analisar os candidatos in silico aos genes de interesse SERK, BBM e psbA e sua possível ocorrência em espécies do Cerrado. Materiais e Métodos: foram testados três diferentes genes como forma de expansão dos estudos e para que fosse possível enquadrar qual gene continha mais sequências relacionadas às frutíferas do Cerrado. Resultados: as sequências de interesse das frutíferas do Cerrado encontradas e analisadas foram: Barbatimão (Stryphnodendron Mart.), Buriti (Mauritia flexuosa L.f.), Cagaita (Eugenia dysenterica [Mart.] DC.), Cajuzinho (Brysonima intermedia A. Juss.), Guacatonga (Casearia syestris Sw.), Jenipapo (Genipa americana L.), Joá-bravo (Solanum viarum Dunal), Pau-terra (Qualea grandiflora Mart.), Pequi (Caryocar brasiliense Cambess.), todas identificadas com seus respectivos descritores botânicos. Conclusão: tal análise possibilitou a discussão e enfoque desta pesquisa subsidiando o desenvolvimento futuro de marcadores fisiológicos, histo-anatômicos e moleculares estádios-específicos, como estratégia para o estudo dos processos da embriogênese zigótica e somática, assim como organogênicos para espécies agronômicas e nativas.


Corresponding author: Guilherme Araújo Lacerda | guilherme.lacerda@unimontes.br
Received: 09/29/2022. Approved: 02/09/2023.
Introduction

The Cerrado biome is considered the richest in biodiversity in the world, but is constantly threatened by human activities\(^1\). In it, approximately 44% of the flora is endemic, making this biome the most diverse savanna of the planet\(^2\,^3\). The second largest biome in South America belongs to the Brazilian Cerrado, with an extension of 2 million km\(^2\), being about 23% of the national territory, considered one of the hotspots of world Biodiversity\(^4\). The plants of the Cerrado undergo several changes in land use, which causes the loss of habitat, and a large part of their vegetation cover, end up becoming pastures for agriculture\(^5\). In vitro plant tissue cultivation systems are increasingly being used due to industrial, medicinal and agricultural challenges. The increase of genotypes of plants of interest with adapted agronomic characteristics is a widely used form for the rapid propagation of valuable clones through somatic embryogenesis\(^6\). Due to the constant environmental degradation, ex situ and in situ approaches become increasingly efficient, to contain the loss of genetic diversity in various areas and habitats. Environmental changes determine which strategy becomes best employed\(^7\).

Somatic embryogenesis and plant regeneration processes are part of the changes in gene expression and are being related to changes in DNA methylation\(^6\). The SERK gene (Somatic Embryogenesis Kinase Receptor) consists of plasma membrane receptor genes that have been studied in several species and, however, being associated with some characteristics, such as reproduction or plant development. They are a small group of transmembrane proteins of plant species belonging to subgroup II of the kinasis\(^8\). The first SERK gene discovered was through cDNA library of a culture of embryogenic cells of carrot species (\textit{Daucus carota})\(^9\). SERKs genes control several aspects of plant development, such as brassinosteroid and phytosulfokine dependent growth, somatic embryogenesis, male sporogenesis and stomatal standardization. SERKs are an integral part of the plant’s immune system and regulate cell death\(^10\).

The Baby Boom gene (BBM) was discovered through embryogenic cell markers of tissue culture in a species of Canola (\textit{Brassica napus}), when it was shown to induce somatic embryos. The BBM gene is capable of inducing embryogenesis in differentiated cells under culture conditions and is also a possible regulator of embryonic development\(^11\). The BBM gene can increase regeneration capacity in tissue culture and is involved in the conversion of vegetative to embryogenic state\(^12\). This gene potentially activates signal transduction pathways that lead to the induction of embryo development from differentiated somatic cells\(^13\). The BBM gene is one of the central regulators of plant cell development and performs functions in embryonic development\(^14\). The \textit{psbA} gene...
(Photosystem II protein D1 (psbA)) encodes the polypeptide D1 of the photosynthesis 2 gene in plants, thus being a gene that can help through the encoded proteins\textsuperscript{15}. More information about this gene and its association with native plants of the Cerrado could contribute to research in the area of bioinformatics.

The selection of improved cultivars, the prospection of extracts of fruit species for pest control and diseases of crops of economic importance, ornamentation and new food products add value to plant products. It is known that, generally, native fruit species do not have in-depth studies as it happens, for example, with commercial fruits such as grape or apple. The fruits of the Cerrado do not present uniformity of the fruit among the plants, thus the innovative uses of the fruit species of the Cerrado are based on other ways of exploitation than extractivism.

This study aimed to analyze the candidates in silico to the genes of interest SERK, BBM and psbA and their possible occurrence in native plant species of the Cerrado.

**Materials and Methods**

**Selection of the evaluated sequences**

The sequences identified as SERK and psbA were obtained, between 2017 and 2018, from the nucleotide database of the NCBI (National Center for Biotechnology Information), the largest biological database in the world, which functions as a collection of various sources. After being saved in FASTA format, they were subjected to the Clusterization process, a computational process that separates the data into similar groups and subsequently submitted to manual annotation.

After this step, all biological sequences identified as SERK and psbA were used for a new search, in order to obtain an optimization of the database saturation, using the BLAST tool (Basic Local Alignment Search Tool). By statistics, this tool finds regions of similarity between biological sequences in order to find new ones for the genes of interest, as well as to correct incomplete clusters.

This process, known as saturation, was repeated until no new significant sequence was found. Two different softwares were used, which resulted in two diversified results.

**Sequence translation and phylogenetic analysis**

After the sequences were translated by the EMBOSS transeq tool, phylogenetic trees were constructed from Mega 11 (Molecular Evolutionary Genetics Analysis), software that performs genetic, evolutionary and molecular analysis\textsuperscript{16,17}. The Mega software has several versions, in which
all have similar functions of creation and editing, in dendrograms, of phylogenetic trees, from nucleotide or polypeptide sequences. From the sequences provided by databases, the Mega software makes a comparative analysis of molecular sequence data, which helps to reconstruct evolutionary histories and to understand the evolution of genes and species of interesse\textsuperscript{18}. The alignment of polypeptide sequences for SERK and psbA genes was performed by the ClustalW program with the standard parameters (default)\textsuperscript{19}. The criteria used were the Neighbor-Joining comparison model, p-Distance method and pair-wise suppression\textsuperscript{20}. The validity of the tree regarding the phylogenetic distance of the clusters was measured by the probabilistic test of bootstrap\textsuperscript{21}.

The evolutionary history was inferred using the Neighbor-Joining method and the ideal tree contains the percentage of replicates in which the associated taxa grouped in the bootstrap test (500 replicats) are shown above the ramus\textsuperscript{20,22}. The tree is drawn in scale, with lengths of branches in the same units of evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were calculated using the p-Distance method and are in the units of the number of amino acid differences per site.

**Identification of common grouping reasons**

To find reasons for grouping between sequences for SERK and psbA, we used the program MEME (Multiple Expectation Minimization for Motif Elicitation, version 5.4.1, using the criteria: zero or one occurrence per sequence, maximum of five motifs, and size for each of them between 6 and 200 aminoacids\textsuperscript{24}).

**Results**

Evolutionary analyses were performed in MEGA11\textsuperscript{17} (Figure 1), whose analysis involved 29 amino acid sequences\textsuperscript{23}. All ambiguous positions were removed for each sequence pair (option of deletion in pairs). There were a total of 731 positions in the final dataset.

No fruit species of the Cerrado was found in the database to correlate to the SERK gene. However, at the end of processo, 29 biological sequences of agroeconomic interest were obtained, such as cotton (*Gossypium hirsutum*), vine (*Vitis vinifera*), pineapple (*Ananas comosus*), rice (*Oryza sativa*), coffee (*Coffea canephora*), lettuce (*Lactuca sativa*), wheat (*Triticum aestivum*), maize (*Zea mays*), soya (*Glycine max*), cocoa (*Theobroma cacao*), carrot (*Daucus Carota*) and potato (*Solanum tuberosum*).
Figure 1. Tree generated in bootstrap analysis, neighbor clustering method (NJ) demonstrating phylogenetic relationships, for SERK gene candidate sequences, among several agronomic species: cotton (Gossypium hirsutum), grapevine (Vitis vinifera), pineapple (Ananas comosus), rice (Oryza sativa), coffee (Coffea canephora), lettuce (Lactuca sativa), wheat (Triticum aestivum), maize (Zea mays), soybean (Glycine max), cocoa (Theobroma cacao), carrot (Daucus carota) and potato (Solanum tuberosum).

The MEME for probable sequences for SERK (Figure 2) was well characterized for the agronomic species: pineapple (Ac - Ananas comosus), carrot (Dc - Daucus Carota), soybean (Gm - Glycine max), cotton (Gh - Gossypium suthirum), English potato (St - Solansum tuberoum) (Tc - Theobrama cacao), maize (Zm - Zea mays), wheat (Ta - Triticum aestivum), rice (Os - Oryza sativa), lettuce (Ls - Lactuca sativa).
Santos LCD, Rodrigues VS, Pimenta MA, Arrudas SR, Lacerda GA.

**Figure 2.** MEME (Multiple Expectation Minimization for Motif Elicitation [http://meme-suite.org/tools/meme]) for probable sequences for SERK in agronomic species: pineapple (Ac - *Ananas comosus*), carrot (De - *Daucus carota*), soybean (Gm - *Glycine max*), cotton (Gh - *Gossypium hirsutum*), potato (St - *Solanum tuberosum*), cocoa (Tc - *Theobroma cacao*), maize (Zm - *Zea mays*), wheat (Ta - *Triticum aestivum*), rice (Os - *Oryza sativa*), lettuce (Ls - *Lactuca sativa*), coffee (Cc - *Coffea canephora*). The parameters used were: any number of repetitions, maximum number of motifs 20 and optimal amplitude between 6 and 200.

![Motif locations](image)

<table>
<thead>
<tr>
<th>Name</th>
<th>p-value</th>
<th>Motif Locations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ac1</td>
<td>0.00e+0</td>
<td></td>
</tr>
<tr>
<td>Ac2</td>
<td>0.00e+0</td>
<td></td>
</tr>
<tr>
<td>Dc1</td>
<td>0.00e+0</td>
<td></td>
</tr>
<tr>
<td>Gm1</td>
<td>0.00e+0</td>
<td></td>
</tr>
<tr>
<td>Gh1</td>
<td>1.51e-99</td>
<td></td>
</tr>
<tr>
<td>Gh2</td>
<td>0.00e+0</td>
<td></td>
</tr>
<tr>
<td>Gh3</td>
<td>0.00e+0</td>
<td></td>
</tr>
<tr>
<td>Gh4</td>
<td>0.00e+0</td>
<td></td>
</tr>
<tr>
<td>Gh5</td>
<td>0.00e+0</td>
<td></td>
</tr>
<tr>
<td>Gh6</td>
<td>0.00e+0</td>
<td></td>
</tr>
<tr>
<td>Ls1</td>
<td>3.73e-121</td>
<td></td>
</tr>
<tr>
<td>Os1</td>
<td>3.61e-257</td>
<td></td>
</tr>
<tr>
<td>St1</td>
<td>0.00e+0</td>
<td></td>
</tr>
<tr>
<td>St2</td>
<td>0.00e+0</td>
<td></td>
</tr>
<tr>
<td>Tc1</td>
<td>0.00e+0</td>
<td></td>
</tr>
<tr>
<td>Ta1</td>
<td>0.00e+0</td>
<td></td>
</tr>
<tr>
<td>Zm1</td>
<td>0.00e+0</td>
<td></td>
</tr>
<tr>
<td>Zm2</td>
<td>0.00e+0</td>
<td></td>
</tr>
<tr>
<td>Zm3</td>
<td>0.00e+0</td>
<td></td>
</tr>
<tr>
<td>Zm4</td>
<td>0.00e+0</td>
<td></td>
</tr>
<tr>
<td>Zm5</td>
<td>0.00e+0</td>
<td></td>
</tr>
<tr>
<td>Zm6</td>
<td>0.00e+0</td>
<td></td>
</tr>
</tbody>
</table>

From the parameters used, it is observed that the polypeptide sequences of the SERK protein (Figure 2) are actually different from the others, since they do not present common motifs, validating the external group of the phylogenetic tree (Figure 1). For the psbA gene, the analysis involved 54 amino acid sequences. All ambiguous positions were removed for each sequence pair (option of deletion in pairs). There were a total of 132 positions in the final dataset. At the end of the analysis, the psbA gene proved to be the best candidate, since it presents a great diversity of
fruit species of the Cerrados, such as Barbatimão (Stryphnodendron Mart.), Buriti (Mauritia flexuosa L.f.), Cagaita (Eugenia enterdysica (Mart.) DC.), Muricio dwarf (Byrsonima intermedia A.Juss.), Guacatonga (Casearia sylvestris Sw.), Jenipapo (Genipa americana L.), Joá-bravo (Solanum viarum Dunal), Pau-terra (Qualea grandiflora Mart.), and Pequi (Caryocar brasiliense Cambess). Only one cultivated species, Tamarindo (Tamarindus indica L.), was obtained in the period studied (Figure 3). The sequences for this species, however, were not used in the construction of phylogeny, since they present such an extensive divergence that would restrict the analysis of the data.

Figure 3. Consensus tree generated in bootstrap analysis, Neighbor Clustering (NJ) method demonstrating phylogenetic relationships for psbA sequences among fruit species from the Cerrado: Barbatimão (Stryphnodendron Mart.), Buriti (Mauritia flexuosa L.f.), Cagaita (Eugenia enterdysica (Mart.) DC.), Murici-dwarf (Byrsonima intermedia A.Juss.), Guacatonga (Casearia Sylvestris Sw.), Jenipapo (Genipa Americana L.), Joá-bravo (Solanum viarum Dunal), Woodwood (Qualea grandiflora Mart.), Pequi (Caryocar brasiliense Cambess.).

The MEME for probable sequences for psbA (Figure 4) was well characterized for native species of the Cerrado, such as Pequi, Cagaita, Buriti and Pau-terra. And although Pau-terra was
characterized in this process, unlike the observed one, it does not share any structural motif with the other species, which justifies its position as belonging to the external group (Figure 3).

**Figure 4.** MEME (Multiple Expectation Minimization for Motif Elicitation (http://meme-suite.org/tools/meme)) for probable sequences for psbA among the fruit species of the Cerrado: Barbatimão (*Stryphnodendron* Mart.), Buriti (*Mauritia flexuosa* L.f.), Cagaita (*Eugenia dysenterica* (Mart.) DC.), Murici-anão (*Brysonima intermedia* A.Juss.), Guacatonga (*Casearia sylvestris* Sw.), Jenipapo (*Genipa Americana* L.), Hornbeam (*Solanum viarum* Dunal), Pau terra (*Qualea grandiflora* Mart.), Pequi (*Caryocar brasiliense* Cambess.).
Discussion

In the present work, the SERK sequences were beyond the agronomic species most described in the literature, such as: carrot, soybean, cotton, English potato, cocoa, maize, wheat and lettuce. SERK is part of the Leurich repeat, receptor-like kinases (LRR-RLKs), subgroup of protein kinases, whose characteristic is an extracellular domain with at least five replications rich in leucine, a transmembrane and another of intracellular kinasis. It has the potential to be a marker of cellular competence to form somatic embryos cultured in vitro, since they are able to stimulate somatic embryogenesis, as studied in pineapples, coffer and colza.

In addition, SERK is not exclusively involved in plant signaling, but in other functions, as a response to biotic and abiotic stress. SERK is able to inhibit organ abscission and restrict the size of floral abscission zones, modulating the location or activity of signaling components responsible for this process.

In studies with Oryza sativa, they realized that the overexpression of this gene was able to increase the resistance of cultivars to brusone, suggesting that it can partially mediate the transduction of the defense signal.

The genes in this group show a complex expression pattern throughout plant development. Both are expressed on the anther carpet and are mediated by precursors. Single knockout mutants of SERK do not show obvious phenotypes, but double mutants give rise to sterile males due to a flaw in carpet specification. Fertility can be restored by a single copy of any of these genes. In addition, they may form homodimers or heterodimers in vivo, suggesting that they are interchangeable in the SERK1/SERK2 signaling complex.

Chloroplasts of higher plants and green algae contain a single psbA gene that is responsible for a set of transcriptions. In response to the action of light, this gene encodes the polypeptide D1 of the photosystem II (PSII) of the plants. Such protein is subject to photodamage and its repair requires degradation of damaged D1 and its replacement by nascent D1. The mechanisms that couple the synthesis of D1 with the assembly and repair of PSII, poorly understood, require assistance by a multitude of proteins encoded by the nucleus and their expression seems to be strongly regulated by translation control. For the psbA, such correlation with the results was not expected, denoting the importance of greater approaches of this gene.

This study presented as a limiting factor the access to sequences related to genes in public
databases, since bioinformatics studies with native plants of the Cerrado are still incipient.

**Conclusion**

The psbA gene proved to be the best candidate, since it presents great diversity of native species of the Cerrados, such as Barbatimão, Buriti, Cagaita, Murici-anão, Guaçatonga, Jenipapo, Joá-bravo, Pau-terra and Pequi. This analysis enabled the discussion and focus of this research, subsidizing the future development of physiological, histoanatomical and molecular markers stages-specific as a strategy for the study of the processes of zygotic and somatic embryogenesis, as well as organogenic for agronomic and native species.

**Authors’ contributions**

All authors approved the final version of the manuscript and declared themselves responsible for all aspects of the work, including ensuring its accuracy and completeness.

**Conflict of interest**

The authors declare that there are no conflicts of interest.

**References**


