

WELLINGTON SILVA GOMES

**CARACTERIZAÇÃO *IN SILICO* E LOCALIZAÇÃO SUBCELULAR DAS  
ENZIMAS DA VIA DAS PENTOSE FOSFATO EM MILHO**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

VIÇOSA  
MINAS GERAIS - BRASIL  
2013

**Ficha catalográfica preparada pela Seção de Catalogação e  
Classificação da Biblioteca Central da UFV**

T

G633c  
2013

Gomes, Wellington Silva, 1983-

Caracterização *in silico* e localização subcelular das enzimas da via das pentoses fosfato em milho / Wellington Silva Gomes. – Viçosa, MG, 2013.

ix, 79f. : il. (algumas color.) ; 29cm.

Inclui apêndice.

Orientador: Aluizio Borém

Tese (doutorado) - Universidade Federal de Viçosa.

Inclui bibliografia.

1. Enzimas. 2. Bioinformática. 3. Células. 4. Genética molecular. 5. Genes. 6. Filogenia. 7. Milho. I. Universidade Federal de Viçosa. Departamento de Fitotecnia. Programa de Pós-Graduação em Genética e Melhoramento. II. Título.

CDD 22. ed. 572.7

WELLINGTON SILVA GOMES

**CARACTERIZAÇÃO *IN SILICO* E LOCALIZAÇÃO SUBCELULAR DAS  
ENZIMAS DA VIA DAS PENTOSSES FOSFATO EM MILHO**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

APROVADA: 7 de março de 2013

---

Wendel Mattos Pompilho

---

Roberto Fritsche Neto

---

Juliana Rocha Lopes Soares Ramos  
(Coorientadora)

---

Eveline Teixeira Caixeta  
(Coorientadora)

---

Aluizio Borém  
(Orientador)

A minha mãe, Maura Luiza da Silva, por todo o amor, carinho, dedicação e por me ensinar a vencer, com honestidade e humildade, todos os desafios da vida.

Mãe, Obrigado por ser essa mulher forte e trabalhadora que, mesmo com pouca instrução e sem a ajuda de outras pessoas, conseguiu educar-nos em toda nossa plenitude.

Com muito amor, dedico.

## AGRADECIMENTOS

Inúmeras pessoas fizeram parte do meu crescimento pessoal e profissional durante o doutorado o que culminou nesse trabalho. Impossível agradecê-las como merecido, porém faço àquelas que se tornaram mais envolvidas nessa pesquisa.

Inicialmente agradeço ao meu Pai Celestial pela modesta consciência que adquiri ao longo desta vida tão efêmera e tão dinâmica, que nunca me deixou só e pelo grande amor dado sem a espera de recompensa.

A minha gloriosa mãe: as vezes que você falou, em cada lágrima sentida que por mim já derramou; obrigado pelos beijos e conselhos que me deu e a vida que você me ofereceu. Obrigado minha mãe por tudo que eu sou.

Agradeço a meu irmão, tios e primos pela cumplicidade que sempre me trataram, pelas horas em que necessitei apenas de um olhar de aprovação. Muito obrigado a todos.

Obrigado a minha grande amiga Marilda, a qual considero minha segunda mãe. Obrigado por todo o amor dedicado e por todos os momentos compartilhados.

À Universidade Federal de Viçosa, particularmente ao Programa de Pós Graduação em Genética e Melhoramento.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES e Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq pela concessão das bolsas de estudo durante o curso.

Ao meu orientador, Prof. Aluizio Borem, pela orientação, conselhos, ensinamentos e incentivo por todos esses anos.

Ao Dr. Mark Settles pela oportunidade de desenvolver meu trabalho em seu laboratório. Obrigado por toda sua atenção durante o período que estive em Gainesville, Florida, como parte do meu doutorado sanduíche.

Agradeço também aos meus amigos de longe e de perto, que sempre me acompanharam e fizeram melhores os meus dias. Em especial, cito aqueles que compartilharam de minhas alegrias e tristezas durante meu período no exterior, Carla Cristina Rosado e Marcos Gomide, sem vocês eu não teria conseguido.

Não posso deixar de citar ao meu colega de trabalho e de casa, Federico Martin, por toda a ajuda prestada, seja com o inglês, ou com os trabalhos desenvolvidos no laboratório.

Ao meu amigo Thiago Otávio, por todos os momentos compartilhados durante o doutorado. Obrigado pelas discussões filosóficas, científicas e pelo companheirismo. Espero que possamos ser amigos por toda a vida.

A minha amiga Paula, pela amizade e por ter aberto sua casa para me receber. Obrigado por tudo.

Aos amigos de longe, minha paixão (Patrícia Thieme), Michelle, Mariana, Lorena, Isabel, Gustavo, Maurício, Preta (Liliane), Carina Bubu (Carina), Renata, Letícia, Edson, Sílvia, Luciano, Wagner, Wendel e Giovanni. Distantes na presença, mas pertos pelo coração. Obrigado pelos anos de amizade.

Agradeço a todos os professores desta Universidade, bem como os demais servidores, que de maneira direta ou indireta contribuíram para minha formação.

## **BIOGRAFIA**

WELLINGTON SILVA GOMES, filho de Maura Luiza da Silva e Geraldo Gomes Ribeiro, nasceu na cidade de Uberlândia, do dia 14 de maio de 1983.

Em 2000, conclui o ensino médio na Escola Estadual Teotônio Vilela, em Uberlândia.

Em setembro de 2002, ingressou na Universidade Federal de Uberlândia, diplomando-se Biólogo em Fevereiro de 2007.

Em março de 2008, iniciou na pós-graduação em Genética e Melhoramento, em nível de mestrado, na Universidade Federal de Viçosa (UFV), recebendo título em julho de 2009.

Em agosto de 2009, ingressou no Programa de Pós-Graduação, em nível de doutorado, em Genética e Melhoramento da UFV e em cooperação com a Universidade da Flórida, desenvolveu seu doutorado sanduíche em 2012, submetendo-se à defesa de tese em março de 2013.

## SUMÁRIO

<b>RESUMO</b> .....	<b>X</b>
<b>ABSTRACT</b> .....	<b>XI</b>
<b>1. INTRODUÇÃO</b> .....	<b>1</b>
<b>2. REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	<b>3</b>
<b>3. REVISÃO DE LITERATURA</b> .....	<b>5</b>
3.1. IMPORTÂNCIA E ORIGEM DO MILHO .....	5
3.2. METABOLISMO DE CARBOIDRATOS .....	6
3.3. A VIA DAS PENTOSE FOSFATO .....	6
3.4. A VIA DAS PENTOSE FOSFATO E A PRODUÇÃO DE AMIDO EM MILHO .....	11
<b>4. REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	<b>13</b>
<b>5. ARTIGOS</b> .....	<b>15</b>
<b>5.1. IDENTIFICAÇÃO E CARACTERIZAÇÃO IN SILICO DOS GENES E ENZIMAS DA VIA DAS PENTOSE FOSFATO EM MILHO</b> .....	<b>16</b>
<b>RESUMO</b> .....	<b>17</b>
<b>ABSTRACT</b> .....	<b>18</b>
5.1.2. INTRODUÇÃO .....	19
5.1.3. MATERIAL E MÉTODOS.....	21
5.1.3.1. Identificação de sequências da via das pentose fosfato em milho.....	21
5.1.3.2. Alinhamento das sequências .....	21
5.1.3.3. Estimativa da distância das sequências .....	22
5.1.3.4. Análise filogenética.....	22
5.1.4. RESULTADOS E DISCUSSÃO .....	23
5.1.5. CONCLUSÕES.....	35
5.1.5. REFERÊNCIAS BIBLIOGRÁFICAS .....	36
<b>5.2. ANÁLISES COMPARATIVAS SOBRE MÉTODOS EXPERIMENTAIS E IN SILICO PARA A DETERMINAÇÃO DA LOCALIZAÇÃO SUBCELULAR DAS ENZIMAS DA VIA DAS PENTOSE FOSFATO DE MILHO</b> .....	<b>40</b>
<b>RESUMO</b> .....	<b>41</b>
<b>ABSTRACT</b> .....	<b>42</b>
5.2.1. INTRODUÇÃO .....	43
5.2.2. MATERIAL E MÉTODOS.....	45
5.2.2.1 Predição da localização subcelular das enzimas da via das pentose fosfato em milho.....	45

5.2.2.2 Avaliação experimental da localização subcelular das enzimas da via das pentose fosfato em milho .....	45
5.2.2.3.1. Material genético .....	45
5.2.2.3.2. Extração de RNA .....	46
5.2.2.3.3. Obtenção de cDNA .....	46
5.2.2.3.4. Amplificação dos genes da PPP .....	47
5.2.2.3.5. Clonagens dos genes de interesse visando fusão com o gene GFP .....	49
5.2.2.3.6. Sequenciamento .....	50
5.2.2.3.7. Preparo das plantas de Nicotiana benthamiana .....	50
5.2.2.3.8. Transformação de Agrobacterium tumefaciens pelo método de choque térmico .....	51
5.2.2.3.9. Transformação das plantas via Agroinfiltração .....	51
5.2.2.3.10. Análises microscópicas .....	52
5.2.3. RESULTADOS E DISCUSSÃO .....	53
5.2.4. CONCLUSÕES .....	67
5.2.5. REFERÊNCIAS BIBLIOGRÁFICAS .....	68
<b>5. APÊNDICE.....</b>	<b>73</b>

## RESUMO

GOMES, Wellington Silva. D.Sc., Universidade Federal de Viçosa, março de 2013. **Caracterização *in silico* e localização subcelular das enzimas da via das pentoses fosfato em milho.** Orientador: Aluizio Borém de Oliveira. Coorientadores: Juliana Rocha Lopes Soares Ramos e Eveline Teixeira Caixeta.

A disponibilidade dos genomas sequenciados tem provocado um aumento no volume de informação e o desafio atual é desenvolver uma análise detalhada e interpretar, funcionalmente, os dados gerados. Tradicionalmente, o enfoque das pesquisas consistia em conhecer uma determinada função e buscar o gene responsável. Atualmente, dispõe-se de um enorme número de genes desconhecidos que demandam caracterização funcional. Enquanto essa caracterização ainda é incipiente, *softwares* têm ajudado cada vez mais a identificar e prever, funcionalmente, genes por meio da comparação de sequências gênicas de espécies conhecidas e caracterizadas. Nesse sentido, este trabalho teve como objetivo a caracterização *in silico* dos genes e enzimas da via das pentoses fosfato em milho e a comparação dos métodos experimentais e por meio de ferramentas de predição *online*, visando identificar os melhores *softwares* disponíveis para prever a localização subcelular das enzimas dessa via. Foram identificados 25 genes em *Arabidopsis thaliana* e 23 em milho, que codificam as sete enzimas da via das pentoses fosfato, sendo que, cada uma das enzimas dessa via são codificadas por pelo menos dois genes, em ambas as espécies avaliadas. Nas avaliações de distância entre as sequências nucleotídicas e análises filogenéticas, distintos graus de similaridade entre os genes foram identificados, apresentando, em geral, maior similaridade entre os genes parálogos. Dez enzimas da via foram experimentalmente caracterizadas e os resultados da comparação dos métodos para a localização subcelular sugerem que os preditores com múltiplas previsões proporcionaram desempenhos significativamente melhores quando comparados aos que se baseiam apenas em sequências protéicas ou peptídeos sinais. Dessa forma, além da caracterização de genes e enzimas da via das pentoses fosfato, foram identificados os *softwares* mais acurados para determinar a localização subcelular das enzimas.

## ABSTRACT

GOMES, Wellington Silva. D.Sc., Universidade Federal de Viçosa, march, 2013. **In silico characterization and subcellular localization of the pentose phosphate pathway in maize.** Adviser: Aluizio Borem de Oliveira. Co-advisers: Juliana Rocha Lopes Soares Ramos e Eveline Teixeira Caixeta.

The availability of sequenced genomes generated a huge volume of information and the current challenge is develop detailed analysis and interpretate, functionally, all data generated. Traditionally, the focus of the research was to learn a particular function and seek the gene responsible. Currently, the situation is reversed. The huge number of unknown genes requires functional characterization. While this characterization is still incipient, softwares have helped increasingly to identify and functionally predict genes by comparing the sequences of known genes from species already characterized. In this study, we sought to develop *in silico* characterization of genes and enzymes of the Pentose Phosphate Pathway in maize and make the comparison of experimental methods and tools with online predictions, aiming to identify the best available softwares for the predicting the localization of these enzymes. Twenty-three genes were identified in maize, encoding the seven enzymes of the pentose phosphate pathway, where each enzyme in this pathway are encoded by at least two genes. In the evaluations of the distance between the nucleotide sequences and phylogenetic analyzes, different degrees of homology among genes were identified, with generally high similarity between paralogous genes. Ten pathway enzymes have been experimentally characterized and the results of the comparison of methods for subcellular localization suggests that the predictors with multiple projections provided significantly better performance compared to those based on only protein sequences or signal peptides. Thus, besides the characterization of genes and enzymes of the pentose phosphate pathway, It was possible identify the most accurate software to determine the subcellular localization of the enzymes.

## 1. INTRODUÇÃO

O milho (*Zea mays* L.) é considerado uma das principais espécies vegetais cultivadas no mundo, fornecendo produtos largamente utilizados para a alimentação humana, animal e para a indústria. Em função da sua multiplicidade de aplicações, o milho assume papel relevante socioeconômico, além de ser matéria-prima indispensável para vários complexos agroindustriais (FANCELLI; DOURADO NETO, 2000, OGLIARI; ALVES, 2007).

A produtividade de híbridos de milho existentes poderá diminuir em cenários de alterações climáticas, principalmente devido ao calor intenso, necessitando assim, maior tolerância térmica nas variedades de milho cultivadas (SOUTHWORTH et al., 2000). Nesse contexto, as discussões sobre o estresse por altas temperaturas e hídrico tem aumentado, e uma maior ênfase é dada normalmente ao último. Talvez isso se deva ao fato de que a perda de água seja imediatamente aparente ou porque métodos agronômicos para reduzir ou eliminar o estresse hídrico estejam bem estabelecidos. Em contraste, na redução do rendimento de grãos, o efeito da alta temperatura não é facilmente perceptível durante o período de estresse (ARNOLDO et al., 2006). Assim, o melhoramento para às mudanças climáticas requer modificação na fisiologia da planta para a produção contínua do estoque de sementes com alta produtividade.

Embora o aumento na temperatura terrestre possa ameaçar a produtividade do milho, pesquisas demonstram que uma variação estável de calor potencializa a atividade da ADP-glicose pirofosforilase (AGPase) aumentando assim, a produção de amido (GREENE; HANNAH, 1998; SMIDANSKY et al., 2003). Entretanto, esse amido produzido não é acumulado em quantidades normais nas sementes (JONES et al., 1984).

Dados experimentais apontam que o acúmulo de amido nas sementes de milho é determinado por etapas fundamentais no desenvolvimento de sementes e em sua fisiologia. O modelo convencional para a biossíntese de amido fornece um caminho direto das hexoses-fosfatos para sua incorporação nas sementes, independentemente da via das pentoses fosfato (HANNAH; JAMES, 2008). Entretanto, Settles e colaboradores (2007)

demonstraram que o metabolismo do carbono central desempenha um papel mais direto na síntese do amido. Mutações no gene que codifica uma das enzimas da via oxidativa das pentoses fosfato, 6-fosfogliconato desidrogenase, ocasionou grave diminuição no acúmulo de amido no endosperma, bem como no desenvolvimento do embrião, sugerindo-se que a via das pentoses fosfato seja importante no acúmulo de amido nas sementes.

Assim, a assimilação de carbono pela fotossíntese e a produção de amido não são prejudicados em climas quentes, entretanto, as alterações nas vias metabólicas, devido ao aumento da temperatura, modificam o fluxo do carbono central a partir da sacarose, reduzindo o enchimento dos grãos pelo amido.

Nesse sentido, um projeto, consórcio entre universidades americanas e o Departamento de Agricultura dos Estados Unidos (USDA) está realizando pesquisas para gerar informações sobre genes e enzimas da via das pentoses fosfato envolvidos no metabolismo de amido, bem como, na alteração do rendimento dos grãos em ambientes com altas temperaturas.

No contexto específico desse projeto, objetivou-se desenvolver a caracterização *in silico* dos genes e enzimas da via das pentoses fosfato e suas relações evolutivas. Além disso, buscou-se ampliar o entendimento da função e interação das enzimas da via das pentoses fosfato por meio da identificação subcelular de cada enzima. Não obstante, buscou-se identificar os preditores mais acurados para determinar a localização subcelular das enzimas em questão, bem como sugestões de *softwares* para serem usados em outras predições para plantas.

## 2. REFERÊNCIAS BIBLIOGRÁFICAS

ARNOLDO, J.R., S., C.-N., LÓPEZ-SANTILLÁN, J., HUERTA, A., TREJO-LÓPEZ, C.; BRIONES-ENCINIA, F. High temperature and water stress during flowering in tropical corn populations. **Phyton-Intl. J. Exp. Bot.** V. 73, p. 31-40. 2006.

FANCELLI, A.L.; DOURADO NETO, D. Produção de milho. Guaíba: **Agropecuária**, 360p. 2000.

GREENE, T.W.; HANNAH, L.C. Enhanced stability of maize endosperm ADP-glucose pyrophosphorylase is gained through mutants that alter subunit interactions. **Proc. Natl. Acad. Sci.** 1998.

HANNAH, L.C.; JAMES, M.G. The complexities of starch biosynthesis in cereal endosperms. **Curr. Opin. Biotechnol.** v. 19, p. 160-165. 2008.

JONES, R.J., QUATTAR, S.; CROOKSTON, R.K. Thermal environment during endosperm cell division and grain filling in maize: Effects on kernel growth and development in vitro. **Crop Sci.** v. 24, p. 133-137. 1984.

OGLIARI J.B; ALVES A.C. Manejo e uso de variedades de milho com estratégia de conservação em Anchieta. In: de Boef WS, Thijssen MH, Ogliari JB, Stapit BR (eds) **Biodiversidade e agricultores: fortalecendo o manejo comunitário**. L&PM, Porto Alegre, p 226–234. 2007.

SETTLES, A.M.; HOLDING, D.R.; TAN, B.C.; LATSHAW, S.P.; LIU, J.; SUZUKI, M.; LI, L.; O'BRIEN, B.A.; FAJARDO, D.S.; WROCLAWSKA, E.; TSEUNG, C.W.; LAI, J.; HUNTER, C.T.; AVIGNE, W.T.; BAIER, J.; MESSING, J.; HANNAH, L.C.; KOCH, K.E.; BECRAFT, P.W.; LARKINS, B.A.; MCCARTY, D.R. Sequence-indexed mutations

in maize using the UniformMu transposon-tagging population. **BMC Genomics**. v. 9, n.8, p. 1-12. 2007.

SMIDANSKY, E.D., MARTIN, J.M., HANNAH, L.C., FISCHER, A.M.; GIROUX, M.J. Seed yield and plant biomass increases in rice are conferred by deregulation of endosperm ADP-glucose pyrophosphorylase. **Planta**. v. 216, p. 656-664. 2003.

SOUTHWORTH, J., RANDOLPH, J., HABECK, M., DOERING, O., PFEIFER, R., RAO, D.; JOHNSTON, J. Consequences of future climate change and changing climate variability on maize yields in the midwestern United States. **Agric. Ecosys. Environ.** v. 82, p. 139-158. 2000.

### 3. REVISÃO DE LITERATURA

#### 3.1. Importância e origem do milho

O milho é uma gramínea da família Poaceae, tribo Maydeae, gênero *Zea* e espécie *Zea mays* L. Taxonomicamente identificado como *Zea mays* L. spp *mays*, para distinguir do seu parente silvestre mais próximo, o teosinto (PATERNIANI; CAMPOS, 1999).

É um dos alimentos mais nutritivos que existem, contendo quase todos os aminoácidos conhecidos, com exceção de lisina e o triptofano, sendo extensivamente utilizado como alimento humano ou ração animal, e cultivado em boa parte do globo (SOUZA; BRAGA, 2004).

Possui alto potencial produtivo e responde bem à tecnologia. Seu cultivo geralmente é mecanizado, se beneficiando muito de técnicas modernas de plantio e colheita. Na agricultura brasileira, em termos de área plantada, a cultura do milho é a segunda mais cultivada (12,8 milhões de hectares), após a cultura da soja, sendo o Brasil o terceiro produtor mundial, com 56,06 milhões de toneladas de grãos colhidos (FAO, 2011).

O milho tem sua origem remontada por divesas hipóteses. Galinat (1995) relata que o milho descende do teosinte, que é uma gramínea com várias espigas sem sabugo, espécie sexualmente compatível com o milho. Segundo Filho (2007), o milho teve origem na América, e é largamente cultivado no mundo inteiro, sendo reputada como a espécie mais cultivada e produtiva e, se considerada individualmente, sua taxa de multiplicação é de 1:600 a 1:1000.

Estudos arqueológicos fornecem elementos que permitem afirmar que o milho já existia como cultura, ou seja, domesticado, há cerca de quatro mil anos e que apresentava as principais características morfológicas que o definem botanicamente na atualidade (GUIMARÃES, 2007).

### **3.2. Metabolismo de carboidratos**

O metabolismo central do carbono é a parte do metabolismo que fornece os precursores para todas as outras rotas metabólicas. A geração de energia, a redução da coenzima  $\text{NAD}^+$ , oxidação das moléculas de  $\text{CO}_2$  e os precursores necessários para as funções celulares estão relacionados a esse metabolismo (CAMPBELL, 2008).

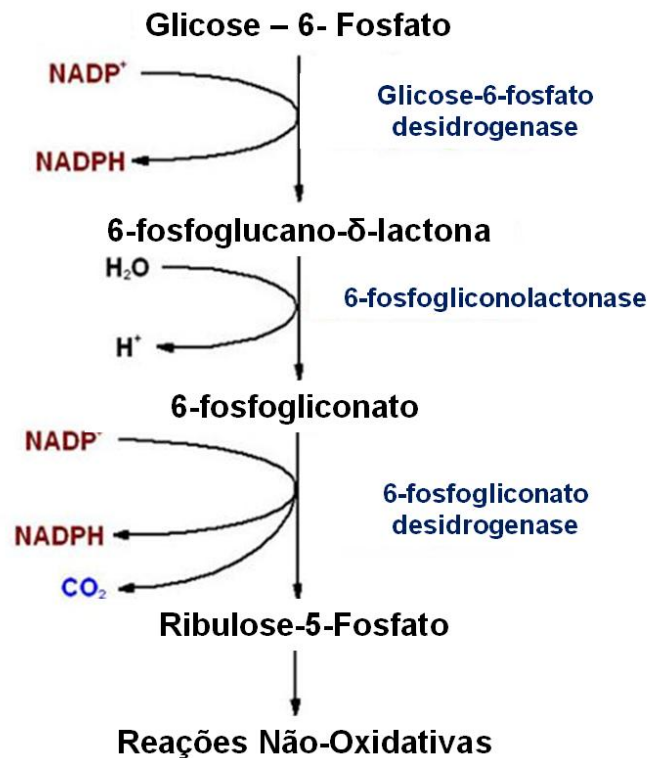
Nesse contexto, a glicose ocupa uma posição central no metabolismo de muitos seres vivos, apresentando um nível relativamente alto de energia potencial, o que a torna um bom combustível para as reações que ocorrem no ambiente intracelular. Este fato é potencializado pela possibilidade de armazenamento celular em formas poliméricas de elevado peso molecular (amido, glicogênio, etc.) que são compatíveis homeostaticamente (não desregulam os níveis de glicose no ambiente interno). A glicose, em situações de exigência energética, vai ser libertada destas formas poliméricas de armazenamento, ficando disponível para entrar em processos de oxidação e consequente síntese de ATP. É importante ressaltar, que a glicose é também usada como precursor de inúmeros intermediários metabólicos em reações de biossíntese (LEHNINGER, 2006).

De uma forma geral, podemos apontar três vias metabólicas principais para a glicose: o seu armazenamento (como polissacarídeo); a sua oxidação pela via das pentoses-fosfato, originando ribose-5-fosfato para a síntese de ácidos nucleicos, e a forma reduzida de nicotinamida adenina dinucleotídeo fosfato NADPH para processos de redução; a oxidação via glicólise originando piruvato e providenciando ATP e intermediários metabólicos para outras vias (CAMPBELL, 2008).

### **3.3. A via das pentoses fosfato**

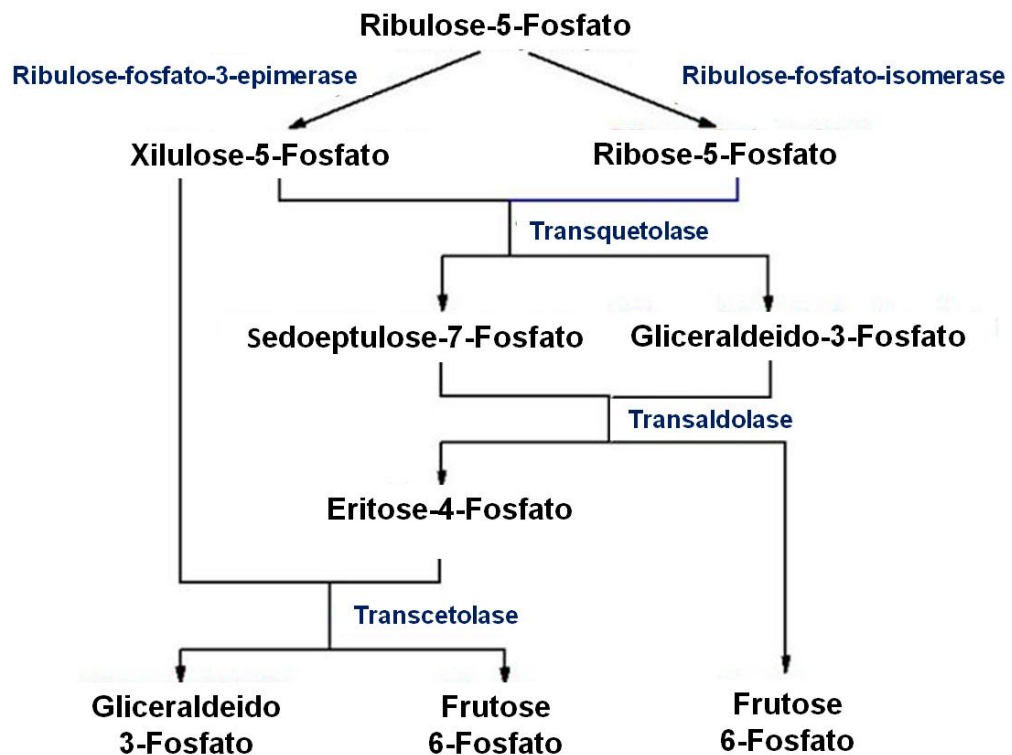
Na maioria das células, o maior destino catabólico da glicose-6-fosfato é a glicólise. No entanto, parte dessa molécula será também degradada pela via das fosfopentoses. A via das pentoses fosfato (PPP) é uma via alternativa de oxidação da glicose, que leva à produção de dois compostos importantes: ribose-5-fosfato e NADPH. Esta via é constituída por duas fases; uma oxidativa e outra não oxidativa (BERG, 2011).

A via das pentoses fosfato compreende uma etapa inicial, oxidativa (oxPPP), em que a glicose 6-fosfato sofre descarboxilação, originando uma pentose fosfato (ribulose 5-fosfato) e NADPH, por meio de duas reações de oxidação catalisadas por desidrogenases intercalada por uma reação de hidrólise (Figura 1).



**Figura 1:** Esquema representativo das reações e enzimas da via Oxidativa das pentoses fosfato. Adaptado de LEHNINGER, 2006.

A etapa subsequente, não-oxidativa (noxPPP), constitui um sistema de rearranjos moleculares, formando açúcares fosforilados onde a ribulose 5-fosfato é transformada em ribose 5-fosfato e/ou em xilulose 5-fosfato por ação de uma isomerase e uma epimerase, respectivamente. Essas pentoses sofrem, a seguir, uma série de conversões, catalisadas por duas enzimas: a transcetolase, que transfere grupos de dois carbonos e tem tiamina pirofosfato como grupo prostético, e a transaldolase que transfere grupo de três carbonos. Todas as reações da etapa não-oxidativa são reversíveis, permitindo a livre interconversão de açúcares (Figura 2).

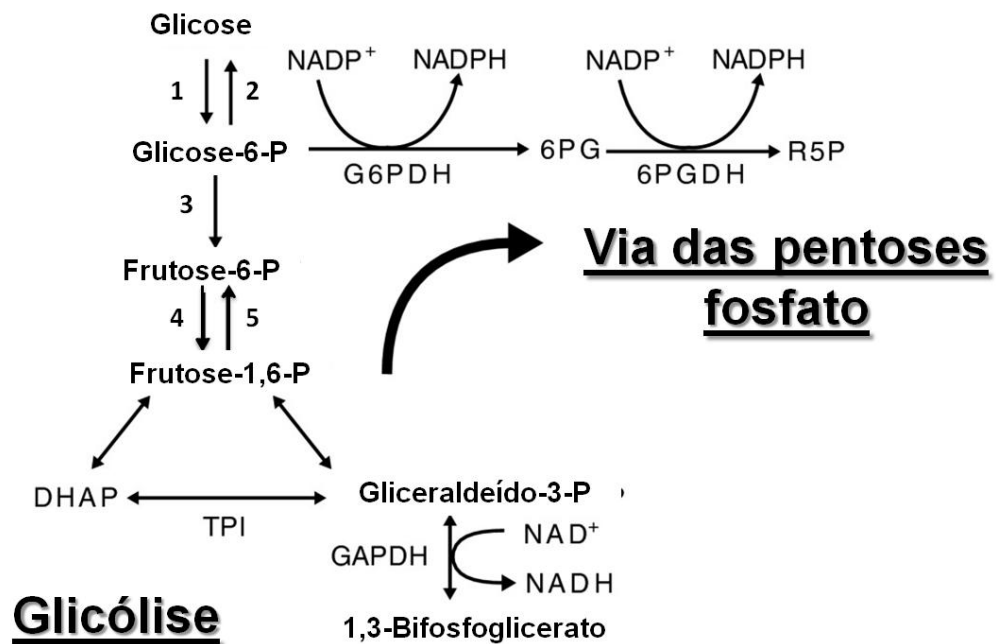


**Figura 2:** Esquema representativo das reações e enzimas da via não-oxidativa das pentoses fosfato. Adaptado de LEHNINGER, 2006..

De maneira geral, a via das pentoses fosfato apresenta cinco principais funções. Permite a combustão total da glicose em uma série de reações independentes do ciclo de Krebs e serve como fonte de pentoses para a síntese dos ácidos nucleicos; forma o NADPH extramitocondrial necessário para a síntese dos lipídeos; converte hexoses em pentoses e, provoca a degradação oxidativa de pentoses pela conversão a hexoses, que podem entrar para a via glicolítica (LEHNINGER, 2006).

A via das pentoses fosfato e a glicólise, apesar de terem funções diferentes, são intimamente relacionadas por apresentarem compostos intermediários comuns: glicose 6-fosfato, frutose 6-fosfato e gliceraldeído-3-fosfato. O compartilhamento de compostos e a reversibilidade das reações da etapa não-oxidativa da via das pentoses fosfato tornam possível o direcionamento dos açúcares fosforilados desta via para a glicólise ou vice-

versa. Além disso, as duas etapas podem ser acionadas em separado, sendo que, o caminho a ser seguido pela glicose 6-fosfato na via das pentoses fosfato é determinado pelas demandas celulares de NADPH ou ribose 5-fosfato (KRUGER; SCHAEWEN, 2003) (Figura 3).



**Figura 3:** Diagrama simplificado da ligação entre a glicólise e a Via da Pentose Fosfato. **1:** Hexoquinase; **2:** Glicose-6-fosfatase; **3:** Fosfoglicoisomerase; **4:** fosfofrutoquinase; **5:** Frutose-1,6 bifosfatase. Adaptado de Grant, *Journal of Biology* 2008.

As trioses-P são os principais produtos da assimilação fotossintética de carbono nos cloroplastos. Essas moléculas podem ser exportadas para o citosol, sendo utilizadas na síntese de açúcares solúveis, açúcares alcoóis, oligossacarídeos solúveis (ex. frutanos). Alternativamente, a triose-P pode ser mantida dentro do cloroplasto para utilização na síntese de amido ou para regeneração do acceptor inicial do CO<sub>2</sub> no Ciclo de Calvin. A maior parte das trioses-P formadas (gliceraldeído 3- fosfato e dihidroxiacetona fosfato) é dirigida para a regeneração da Rubisco (RuBP), enquanto 1/6 delas podem ser desviadas para a síntese de amido e/ou sacarose.

Após a formação das trioses-P, dois importantes *pools* alimentam o metabolismo de carboidratos das plantas: um é composto pelos intermediários da via das pentoses fosfato, juntamente com as trioses-P; e o outro é constituído pelas hexoses fosfato.

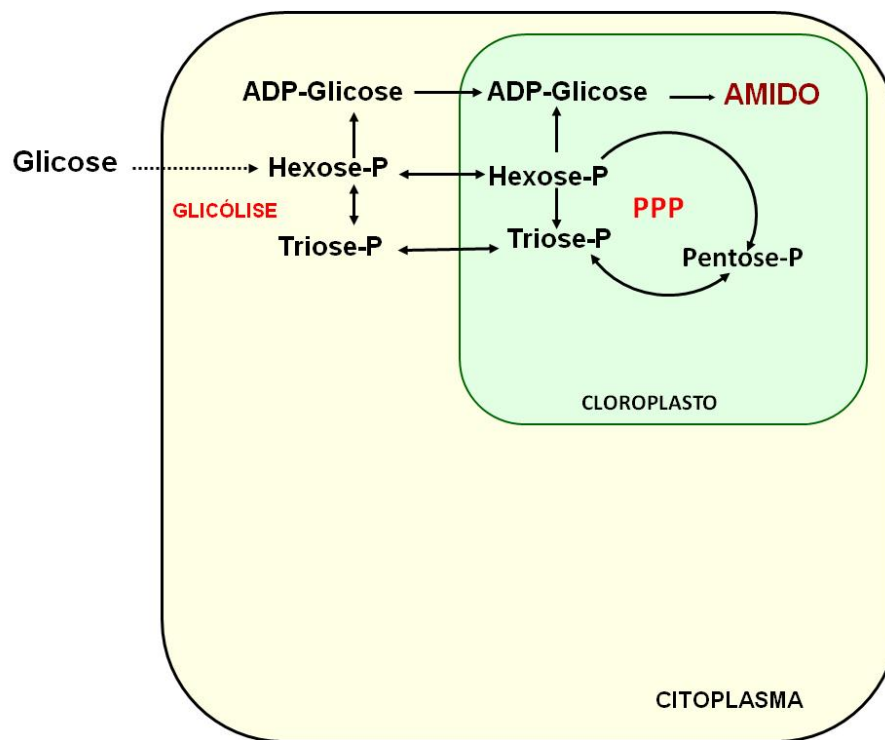
Os metabólitos do *pool* de hexoses fosfato (glicose 1-fosfato, glicose 6-fosfato e frutose 6-fosfato) estão em equilíbrio devido à ação de duas enzimas: a fosfoglucomutase e a glicose 6-fosfato isomerase, responsáveis pela interconversão destes três metabólitos. A entrada neste *pool* pode ser feita por meio de metabólitos provenientes da gliconeogênese e da fosforilação de hexoses livres e a saída ocorre em ocasiões onde há a síntese de amido a partir de glicose 1-fosfato, síntese de sacarose utilizando glicose 1-fosfato e frutose 6-fosfato formação de parede celular a partir de glicose 1-fosfato e por meio de reações oxidativas da via das pentoses fosfato (glicose 6-fosfato). É importante ressaltar que o maior dreno desses metabólitos é a utilização da frutose 6-fosfato como precursora da via glicolítica (CAMPBELL, 2008).

Embora a glicose seja normalmente representada como produto final da fotossíntese em equações resumidas, na realidade pouca glicose livre é formada. Na medida em que a glicose (ou outros monossacarídeos como manose, galactose, xilose, etc) é produzida, vai sendo convertida em seus polímeros, como o amido, que é o principal carboidrato de reserva das plantas, ou sacarose, um dissacarídeo que é o principal açúcar de transporte das plantas.

A reação catalisada pela 6-fosfogluconato desidrogenase (G6PDH) é considerada a etapa limitante de velocidade e de controle de passos do fluxo da OxPPP devido à inibição por *feedback* da enzima pelo NADPH. A enzima G6PDH tem sido associada com a assimilação de nitrogênio, estresse oxidativo e respostas de defesa. Surpreendentemente, mutantes G6PDH de *Arabidopsis* interferem com enzimas citosolicas e plastídicas mostrando fenótipos diferentes do esperado. Em contraste, mutantes de 6PGL em *Arabidopsis* mostram sementes com fenótipos letais ou com plantas de tamanho reduzido quando a enzima é bloqueada ou tem sua expressão reduzida, respectivamente (BERG, 2011).

### 3.4. A via das pentoses fosfato e a produção de amido em milho

Classicamente, a biossíntese de amido é conhecida por exigir ATP e pirofosfato, mas não NADPH ou NADH. O modelo convencional indica que as hexoses fosfato são incorporadas diretamente no amido em sementes de milho (HANNAH; JAMES, 2008). Entretanto, estudos com  $^{13}\text{C}$  indicam que a maior parte das hexoses sintetizadas são recicladas à partir do metabolismo central de carbono antes da incorporação do amido e que a via das pentoses fosfato é importante para a direcionar as hexoses-trioes para a produção e acúmulo de amido nas sementes (SPIELBAUER et al., 2006) (Figura 4).



**Figura 4.** Esquema geral da biossíntese de amido em grãos de milho, utilizando o metabolismo de carbono central. O fluxo convencional da biossíntese de amido é: sacarose/glicose é transportado para dentro da célula e convertido em ADP-glicose no citosol e, depois, ADP-glicose é transportada para os cloroplastos e incorporada em amido (SPIELBAUER et al., 2006).

Estudos indicam que a atividade da enzima da via das pentoses fosfato 6PGDH está associada a tecidos mais profundos e é especialmente importante no endosperma para promover o crescimento do embrião e do endosperma basal em milho. Experiências histológicas sugerem uma participação ativa de 6PGDH plastídica na biossíntese de amido. Mutações nesse gene causam reduções graves no enchimento de grãos e prejudicam o desenvolvimento do endosperma (SPIELBAUER et al., 2006).

Recentes pesquisas demonstram que uma variação estável de calor potencializa a atividade da ADP-glicose pirofosforilase aumentando, portanto, a produção de amido no milho (GREENE; HANNAH, 1998; SMIDANSKY et al., 2003). O aumento da fotossíntese líquida, em linhagens de milho, é observado em temperaturas altas próximas a 35°C, assimilando açúcares em quantidade normal ou mesmo em taxas elevadas (CHELKH; JONES, 1995), no entanto, o peso da semente e teor de amido são dramaticamente reduzidos (JONES et al., 1984). A produção de carboidratos pela fotossíntese e o transporte para a semente não são prejudicados pelo calor, entretanto, as alterações nas vias metabólicas, devido ao aumento da temperatura, modificam o fluxo do carbono central a partir da sacarose, reduzindo o enchimento dos grãos de milho pelo amido.

#### 4. REFERÊNCIAS BIBLIOGRÁFICAS

BERG, J.M. **Bioquímica**. 6. ed. Rio de Janeiro: Guanabara Koogan, 2008. 114 p.

CAMPBELL, Mary K.; Shawn O. Farrell. **Biochemistry**. 7th edition. 2008.

CAMPBELL, MARY; SHAWN, K. O. Farrell. **Biochemistry**. 7th edition. 2008.

CHELKH, N.; JONES, R.J. Heat stress effects on sink activity of developing maize kernels grown in vitro. **Physiol. Plant**. v. 95, p. 59-66. 1995.

FAO. **FAOSTAT**. Statistics Division 2011. Disponível em: <http://faostat.fao.org/>. Acesso em: 05 jan. 2013.

FILHO, F. D. **Manual da cultura do milho**. Jaboticabal: FUNEP, 576 p. 2007.

GALINAT, W. C. The origin of maize: grain of humanity. New York: **New York Botanical Garden Journal**, v.44, p.3-12, 1995.

GREENE, T.W.; HANNAH, L.C. Enhanced stability of maize endosperm ADP-glucose pyrophosphorylase is gained through mutants that alter subunit interactions. **Proc. Natl. Acad. Sci**. 1998.

GUIMARÃES, P. S. Desempenho de híbridos simples de Milho (*Zea mays* L.) E correlação entre heterose e divergência genética entre as linhagens parentais cultivares e população de plantas na produção de milho-verde. 2007. 132f. Dissertação (Mestrado em Agricultura Tropical e Subtropical) – Instituto Agronômico de Campinas, Campinas, 2007.

HANNAH, L.C.; JAMES, M.G. The complexities of starch biosynthesis in cereal endosperms. **Curr. Opin. Biotechnol**. v. 19, p. 160-165. 2008.

KRUGER, N.J, VON SCHAEWEN, A. The oxidative pentose phosphate pathway: structure and organisation». **Curr. Opin. Plant Biol.** v. 6 (3). p. 236–46. 2003.

LEHNINGER, A. L. **Princípios de bioquímica.** 4. ed. São Paulo: Sarvier, 2006.

PATERNIANI, E.; CAMPOS, M.S. **Melhoramento do milho.** In: BORÉM, A. Melhoramento de espécies cultivadas. Viçosa. p.429-485. 1999.

SMIDANSKY, E.D., MARTIN, J.M., HANNAH, L.C., FISCHER, A.M.; GIROUX, M.J. Seed yield and plant biomass increases in rice are conferred by deregulation of endosperm ADP-glucose pyrophosphorylase. **Planta.** v. 216, p. 656-664. 2003.

SOUZA, P.M. de; BRAGA , M.J. **Aspéctos econômicos da produção e comercialização do milho no Brasil.** In: GALVÃO, J.C.C.; MIRANDA, G.V. (Eds) Tecnologia de produção do milho. 20.ed. Viçosa: Universidade Federal de Viçosa, v.1, p.13-53, 2004.

SPIELBAUER, G., MARGL, L., HANNAH, L.C., ROMISCH, W., ETTENHUBER, C., BACHER, A., GIERL, A., EISENREICH, W.; GENSCHEL, U. Robustness of central carbohydrate metabolism in developing maize kernels. **Phytochemistry** v. 67, p. 1460-1475. 2006.

## **5. ARTIGOS**

**5.1. IDENTIFICAÇÃO E CARACTERIZAÇÃO *IN SILICO* DOS GENES E  
ENZIMAS DA VIA DAS PENTOSE FOSFATO EM MILHO**

## RESUMO

Os avanços no sequenciamento em larga escala dos genomas mudaram dramaticamente o âmbito da investigação biológica e a biologia computacional. Assim, o conhecimento da organização e funcionamento de genes em células vegetais torna-se extremamente importante para a identificação de vias e processos metabólicos. Nesse sentido, o objetivo deste trabalho foi promover a caracterização *in silico* dos genes e enzimas da via das pentoses fosfato em milho (*Zea mays*), por meio da utilização de sequências homólogas de *Arabidopsis thaliana* dessa via. Foram identificados e caracterizados 25 genes em *Arabidopsis* e 23 em milho que codificam sete enzimas da via das pentoses fosfato, por análises dos bancos de dados *NCBI* e *Maizesequence*. Foi verificado que, cada uma das enzimas dessa via é codificada por pelo menos dois genes, em ambas as espécies avaliadas. Além disso, apesar de *Arabidopsis* possuir maior número de genes, a espécie possui também, maior número de duplicações gênicas conservadas entre os genes parálogos. Nas avaliações de distância entre as sequências nucleotídicas e análises filogenéticas, graus distintos de similaridade entre os genes foram identificados, apresentando, em geral, maior similaridade entre os genes parálogos. Entretanto, o gene AT5G44520 foi pouco similar aos demais que codificam a enzima Ribose-5-fosfato isomerase e os genes codificadores da enzima transcetolase não se agrupou entre genes parálogos. Assim, foi possível caracterizar os genes e enzimas da via das pentose fosfato de milho.

**Palavras chave:** Bioinformática, alinhamento de sequências, similaridade, homologia.

## ABSTRACT

Advances in the large scale sequencing of the genomes has changed dramatically the scope of biological research and biology of computing. Thus, knowledge of the organization and genes function in plant cells is extremely important for identifying pathways and metabolic processes. In this sense, the aim of this work was to promote in silico characterization of genes and enzymes of the pentose phosphate pathway in maize (*Zea mays*). It was identified and characterized 25 genes in *Arabidopsis* and 23 in maize encoding seven pentose phosphate pathway enzymes, by analysis in the databases of the NCBI and Maizesequence. It was verified that each one of the enzymes is encoded by, at least, two genes in both species evaluated. Furthermore, although *Arabidopsis* have bigger number of genes, It has also more gene duplications conserved among paralogs. In the evaluations of the distance between the nucleotide sequences and phylogenetic analyzes, different degrees of homology among genes were identified, with generally high similarity between paralogous genes. However, AT5G44520 gene was less similar to other genes encoding the enzyme Ribose-5-phosphate isomerase and the gene encoding the enzyme transketolase not clumped together between paralogous genes. Thus, it was possible to characterize the genes and enzymes of the pentose phosphate pathway in maize.

**Keywords:** Bioinformatics, sequences alignment, similarity, homology.

### 5.1.2. INTRODUÇÃO

A via das pentoses fosfato atende a necessidade de todos os organismos de uma fonte de NADPH para ser usado na biossíntese redutora. Esta via é constituída por duas fases: a formação de NADPH na fase oxidativa e a interconversão não oxidativa de açúcares. Na fase de oxidação, NADPH é gerado quando a glicose-6-fosfato é oxidada a ribose 5-fosfato (BERG, 2011).

A seção oxidativa da via das pentoses fosfato (oxPPP) converte a glicose-6-P em ribulose-5-P. Duas desidrogenases, glicose-6-P desidrogenase e 6-fosfogliconato desidrogenase, catalisam a oxidação irreversível da glicose-6-P, gerando duas moléculas de NADPH e uma molécula de CO<sub>2</sub>. O produto da primeira reação oxidativa, 6-phosphoglicono- $\delta$ -lactona, pode hidrolisar a 6 - fosfogliconato espontaneamente ou por catálise pela 6-fosfogliconolactonase. A parte não oxidativa do PPP (noxPPP) regenera a glicose-6-P a partir de ribulose-5-P por meio de reações reversíveis catalisadas por ribulose-5-P-epimerase, ribose-5-P-isomerase, transaldolase e transcetolase.

O interesse pela oxPPP aumentou nos últimos anos, provocado por uma apreciação crescente do papel central da via no metabolismo. Em células não fotossintéticas, o oxPPP é uma fonte coenzimas reduzidas importante em processos biossintéticos, tais como a síntese de ácidos graxos e a assimilação de nitrogênio inorgânico (NEUHAUS; EMES, 2000) e na participação de manutenção da homeostase celular (KRUGER; SCHAEWEN, 2003).

Para a obtenção do conhecimento sobre os genes e enzimas que compõem essa via, métodos computacionais têm sido cada vez mais utilizados nos últimos anos. Como o ritmo dos sequenciamentos genômicos tem aumentado, a necessidade de métodos rápidos de descoberta de genes torna-se cada vez maior. A sequência do genoma é apenas o começo de um esforço maior para entender as funções de um organismo, e um dos passos mais críticos nesse processo é a identificação precisa de todos os genes e suas proteínas associadas (LIEW et al., 2005).

A forma mais confiável para identificar um gene em um genoma recentemente sequenciado é por meio da utilização de uma sequência altamente similar conhecida em outro organismo. Para que o gene de interesse possa ser encontrado por similaridade, uma base de dados é utilizada para detectar sequências nucleotídicas ou protéicas semelhantes, devido a ancestralidade comum. A semelhança entre as sequências de uma proteína em diferentes organismos evidencia que a sequência de DNA correspondente pode ter origem semelhante. Se a relação evolutiva é distante, a proteína homóloga possui informação limitada sobre a função do novo gene. Para as proteínas intimamente relacionadas, a similaridade pode proporcionar informação funcional detalhada se a proteína relacionada tiver sido experimentalmente caracterizada (PERTEA; SALZBERG, 2002).

Além da caracterização estrutural e funcional dos genes, a filogenia ou estudo de relações evolutivas permite um estudo mais amplo do que características morfológicas e fisiológicas. Li (1997) cita três principais motivos para explicar a grande importância da utilização de sequências de DNA para estudos evolutivos. Inicialmente, o DNA (e as proteínas) evoluem de maneira mais regular do que caracteres morfológicos e fisiológicos. Em segundo lugar, estes dados permitem melhor tratamento quantitativo do que dados morfológicos, e, por último, os dados de sequências são mais abundantes e informativos.

Nesse sentido, o conhecimento da organização, relação e funcionamento dos genes e enzimas que compõem as vias metabólicas nas células vegetais torna-se extremamente importante para a compreensão de todos os processos com os quais cada uma das enzimas interage. Assim, o objetivo deste trabalho foi caracterizar os genes e enzimas da via das pentoses fosfato, em milho, por meio da utilização de sequências protéicas conhecidas de *Arabidopsis thaliana* para as mesmas enzimas em bancos de dados disponíveis na internet.

### 5.1.3. MATERIAL E MÉTODOS

#### 5.1.3.1. Identificação de sequências da via das pentoses fosfato em milho

Para a identificação dos genes da via das pentoses fosfato em milho, foi realizada a busca inicial dos genes correspondentes em *Arabidopsis thaliana*, haja vista que, a espécie possui todos os genes e enzimas dessa via caracterizados. Os genes de *A. thaliana* foram selecionados no banco de dados disponíveis em “The Arabidopsis Information Resource” (TAIR) (<http://www.arabidopsis.org/>) utilizando-se para tal, o nome de cada enzima da via: (glicose-6-fosfato desidrogenase, 6-fosfogluconolactonase, 6-fosfogluconato desidrogenase, ribose-5-fosfato isomerase, transaldolase, ribulose-fosfato 3-epimerase e transcetolase). Após a mineração, as sequências foram conferidas no banco de dados “Plant Metabolic Network” (PMN) (<http://www.plantcyc.org/>).

As sequências de mRNA, bem como das proteínas correspondentes de cada gene da via das pentoses fosfato na linhagem de milho B73 (*Zea mays L.*) foram identificadas por meio do *software* BLAST, no banco de dados do *National Center for Biotechnology Information* (NCBI; <http://blast.ncbi.nlm.nih.gov/Blast.cgi>) e as sequências incompletas foram confirmadas no banco de dados do milho *Maizesequence* (<http://www.maizesequence.org/blast>) para sua completa anotação, bem como para verificação de sua função correlata.

Para confirmar se as sequências encontradas correspondem a ORFs (*open reading frame*), foi usado o *software* ORF FINDER (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>), que analisa as ORFs, alinhando-as e comparando-as com sequências conhecidas de outras espécies, depositadas em bancos de dados como o *GenBank*.

#### 5.1.3.2. Alinhamento das sequências

As sequências de nucleotídeos obtidas, no formato FASTA, também conhecido como formato Pearson, foram alinhadas utilizando o *software* CLUSTAL W (CHENNA et

al. 2003), por meio do alinhamento par a par e múltiplo (JEANMOUGIN et al., 1998). As correções manuais foram realizadas com o auxílio do *software* GeneDoc, versão 2007 (NICHOLAS; NICHOLAS 2000).

#### **5.1.3.3. Estimativa da distância das sequências**

A distância genética entre as sequências de milho obtidas foi estimada pela diferença na composição de nucleotídeos, com completa deleção de *gaps*, para cada par de sequências analisadas. A medida representa a metade da diferença da soma de quadrados nas contagens de bases (ou resíduos). Usando o *software* MEGA 5.1 (TAMURA et al., 2011) essa distância foi calculada por meio da diferença entre composição total, entre duas sequências, dividido pelo número de posições comparadas, excluindo as lacunas e os dados em falta.

#### **5.1.3.4. Análise filogenética**

As sequências de aminoácidos das enzimas da via das pentoses fosfato de *Arabidopsis thaliana* e milho foram alinhadas com a utilização do *software* Clustal W (THOMPSON et al., 1994). A árvore filogenética foi reconstruída pelo *software* MEGA 5.1 por meio dos métodos vizinho mais próximo para a análise de similaridade entre as sequências nucleotídicas dos genes em milho, e pelo método de máxima verossimilhança para a reconstrução filogenética entre as sequências de aminoácidos das enzimas de milho e *Arabidopsis*, usando a distância p e a distância Kimura dois parâmetros, respectivamente.

Todos os alinhamentos foram inspecionados e corrigidos manualmente, quando necessário. As análises de confiança dos agrupamentos foram realizadas através de *bootstrap* com 1.000 réplicas, sendo considerados confiáveis aqueles com no mínimo de 50% das réplicas. O valor de *bootstrap* representa as porcentagens de casos nos quais cada grupo apareceu nas árvores resultantes.

### 5.1.4. RESULTADOS E DISCUSSÃO

A estratégia adotada para caracterizar os genes que codificam as enzimas da via das pentoses fosfato em milho, por meio da análise de sequências similares em *Arabidopsis thaliana*, gerou a identificação e localização cromossômica de 23 genes (Tabela 1). Dados referentes às características gênicas e proteicas foram coletados e estão demonstrados nas Tabelas 1 e 2.

**Tabela 1:** Resumo dos genes de *Arabidopsis thaliana* e milho e suas respectivas enzimas codificadas da via das pentoses fosfato, bem como sua localização cromossômica em cada espécie.

Enzima	Locus em <i>Arabidopsis</i>	Cromossomo	Entrada da Enzima	Gene Model em milho	Cromossomo			
Glicose-6-fosfato desidrogenase	AT1G09420	1	EC 1.1.1.49	GRMZM2G179521	1			
	AT1G24280	1		GRMZM2G130230	2			
	AT3G27300	3		GRMZM2G031107	5			
	AT5G35790	5		GRMZM2G426964	9			
	AT5G13110	5		GRMZM2G177077	10			
6-fosfogliconolactonase	AT5G40760	5	EC 3.1.1.31	GRMZM2G122126	1			
	AT1G13700	1				GRMZM2G074946	1	
	AT3G49360	3				GRMZM2G148769	2	
	AT5G24400	5				GRMZM2G136918	7	
	AT5G24420	5				GRMZM2G127798	6	
AT5G24410	5	EC 1.1.1.44	GRMZM2G145715	3				
AT3G02360	3				GRMZM2G440208			4
AT1G64190	1				EC 5.3.1.6			GRMZM2G104070
AT5G41670	2	GRMZM2G035599	2					
AT2G01290	2	GRMZM5G891282	4					
AT1G71100	1	GRMZM2G456086	7					
Ribose-5-fosfato isomerase	AT5G61410	5	EC 5.1.3.1	GRMZM2G026807	1			
	AT3G01850	3				GRMZM2G083102	7	
	AT1G63290	1				GRMZM2G178960	9	
Transcetolase	AT3G60750	3	EC 2.2.1.1	GRMZM2G010494	3			
	AT2G45290	2				GRMZM2G033208	9	
Transaldolase	AT5G13420	5	EC 2.2.1.2	GRMZM2G139550	4			
	AT1G12230	1				GRMZM2G134256	8	

**Tabela 2:** Caracterização *in silico* dos genes e enzimas da via das pentoses fosfato em milho.

Enzima	Abreviação	Gene Model	Nome do gene	Início	Fim	Sentido	Tamanho do gene (bp)	Tamanho enzima (aa)	Massa molecular (Da)	Acess Genbank
Glicose-6-fosfato desidrogenase	G6PDH	GRMZM2G179521	LOC100284317	70948034	70951991	forward	1967	605	68,35	ACN35127.1.
Glicose-6-fosfato desidrogenase	G6PDH	GRMZM2G130230	LOC100284317	36942405	36950819	reverse	1877, 1843, 2115	507	57,629	ACG39996.1.
Glicose-6-fosfato desidrogenase	G6PDH	GRMZM2G031107	LOC100383421	180606107	180611732	forward	1947, 1956	517	58,624	ACN34086.1.
Glicose-6-fosfato desidrogenase	G6PDH	GRMZM2G426964	LOC100304292	133929494	133932052	reverse	1405	384	44,113	ACN25403.1.
Glicose-6-fosfato desidrogenase	G6PDH	GRMZM2G177077	gpm175	99232380	99238980	forward	2703	598	67,57	ACG29334.1.
6-fosfogliconolactonase	6PGL	GRMZM2G122126	LOC100285593	207021423	207023452	reverse	1476	322	34,354	ACR34195.1.
6-fosfogliconolactonase	6PGL	GRMZM2G074946	IDP1953	73245938	73,251,175	reverse	1931, 1890, 1662, 1647	288	32	ACF82627.1.
6-fosfogliconolactonase	6PGL	GRMZM2G148769	Umc2374	217787761	217791673	reverse	1503, 2135	324	34,782	ACG46383.1.
6-fosfogliconolactonase	6PGL	GRMZM2G136918	Pco143240a	158381739	158387917	reverse	1579, 1512, 1316	281	31,311	ACN25195.1.
6-fosfogliconato desidrogenase	6PGDH	GRMZM2G127798	Pgd2	57980361	57983057	reverse	2594, 2033, 2039	482	52,851	ACF85199.1.
6-fosfogliconato desidrogenase	6PGDH	GRMZM2G145715	Pdh1	150098854	150102739	reverse	2069	480	52,675	ACN35899.1.
6-fosfogliconato desidrogenase	6PGDH	GRMZM2G440208	AF061839	18436118	18438040	reverse	1923	507	54,167	ACG41643.1.
Ribose 5-fosfato isomerase	R5PI	GRMZM2G104070	LOC100285676	152550931	152552032	forward	2046	580	63,272	ACN34846.1.
Ribose 5-fosfato isomerase	R5PI	GRMZM2G035599	LOC100283465	142093854	142094627	forward	1433	257	26,803	ACG36951.1.
Ribose 5-fosfato isomerase	R5PI	GRMZM5G891282	LOC100284292	187449297	187450591	reverse	1295	263	27,425	ACN31522.1.
Ribose 5-fosfato isomerase	R5PI	GRMZM2G456086	Pco087551	9255171	9256567	reverse	2030	274	28,778	ACG36789.1.
Ribulose-fosfato 3-epimerase	RPE	GRMZM2G026807	LOC100274365	14563890	14579039	forward	1766	274	28,933	ACR37495.1.
Ribulose-fosfato 3-epimerase	RPE	GRMZM2G083102	LOC100283478	136,742,676	136746945	forward	1145, 1241, 1166, 704, 1171	225	24,141	ACG36997.1.
Ribulose-fosfato 3-epimerase	RPE	GRMZM2G178960	LOC100282201	146967883	146971758	reverse	1412	274	29,003	ACN37044.1.
Transcetolase	TKa	GRMZM2G010494	Zm.124451	154398189	154401921	forward	2361	633	68,974	ACN3446.1.
Transcetolase	TKa	GRMZM2G033208	LOC100279513	22732320	22736865	forward	2418	675	72,933	AAAN65341.1.
Transadolase	TA	GRMZM2G139550	LOC100192890	33512825	33514077	forward	1739	394	42,905	ACF79996.1.
Transaldolase	TA	GRMZM2G134256	LOC100284307	159099276	159102258	forward	1922	429	46,246	ACF84672.1.

Foram identificados pelo menos dois genes homólogos para cada uma das enzimas da via das pentoses fosfato, em ambas as espécies analisadas. Apesar de o milho ser diploide, sua história remonta eventos distintos de poliploidia e, devido a esse fato, não é surpresa que haja grande número de duplicações gênicas e divergência funcional nessa espécie, haja vista que, isso é tido como o grande sucesso dos poliplóides (OHNO, 1970; LEVIN, 1983; HUGHES, 1994).

Apesar da *A. thaliana* ser haplóide e conter o segundo menor genoma entre as plantas fanerógamas, ela possui um genoma complexo com alto grau de duplicação dos genes que codificam enzimas de processos metabólicos centrais, tais como a via das pentoses fosfato (WOLFE, 2001). Isso explica o alto grau de redundância gênica da via das pentoses fosfato encontrado nessa espécie. Entretanto, pelo milho ter um genoma bem maior que a *Arabidopsis*, esperava-se que o número de genes redundantes fosse maior que o encontrado.

Kruger e Schaewen (2003) também relataram que o genoma de *A. thaliana* contém múltiplas cópias de genes que estão preditos para codificar cada uma das enzimas da via das pentoses fosfato e segundo pesquisadores do *Genome Initiative Arabidopsis*, parte desta aparente redundância pode ser devido a um requerimento para as diferentes isoenzimas funcionarem no citosol e nos cloroplastos. No entanto, esta hipótese não pode explicar a multiplicidade de genes que codificam algumas enzimas da via não oxidativa das Pentoses Fosfato, uma vez que, se supõe que estejam restritas ao cloroplasto (SCHNARRENBERGER et al., 1995; DEBNAM; EMES; 1999).

O *software* ORF FINDER foi utilizado para identificar se as sequências de milho encontradas possuem uma sequência aberta de leitura (ORF), e nos casos positivos, indicar a quantidade e a disposição dessas estruturas nos genes. Na Tabela 3, estão demonstrados os dados do tamanho real da sequência obtidos pelo banco de dados do *MaizeSequence*, o qual incluindo regiões codificadoras e não codificadoras, como os promotores, bem como o número de ORFs e a maior região codificadora identificada pelo ORF FINDER, para cada um dos genes analisados. Foram encontradas de 5 a 20 sequências abertas de leitura para as sequências analisadas, com tamanhos e localizações diferentes nos genes em questão (Apêndice A).

**Tabela 3:** Quantidade e tamanhos das maiores ORFs dos genes da via das pentoses fosfato identificados pelo *software* ORF Finder. O tamanho real das sequências depositadas no banco de dados foi indicado como parâmetro de comparação. Os tamanhos estão representados em pares de bases (pb). A sequência da ORF representa o intervalo entre o primeiro códon (metionina) e o último códon (*stop* códon).

Gene Model	Tamanho real	Maior ORF	Número de ORFs
GRMZM2G179521	1967	1818	9
GRMZM2G130230	1877	1524	17
GRMZM2G031107	1947	1554	12
GRMZM2G426964	1405	1155	6
GRMZM2G177077	2703	1797	13
GRMZM2G122126	1476	1077	9
GRMZM2G074946	1931	867	14
GRMZM2G148769	2135	975	17
GRMZM2G136918	1579	846	8
GRMZM2G127798	2594	1449	11
GRMZM2G145715	2069	1443	17
GRMZM2G440208	1923	1524	13
GRMZM2G104070	2046	1743	15
GRMZM2G035599	1433	774	8
GRMZM5G891282	1295	843	5
GRMZM2G456086	2030	825	11
GRMZM2G026807	1766	825	12
GRMZM2G083102	1145	678	9
GRMZM2G178960	1412	825	9
GRMZM2G010494	2361	1902	14
GRMZM2G033208	2418	1866	20
GRMZM2G139550	1739	1185	12
GRMZM2G134256	1922	1290	13

Alguns trabalhos relataram a identificação de genes contendo dois ou mais ATGs no quadro de leitura da extremidade 5', qualquer um dos quais pode corresponder ao códon usado para iniciar a tradução de uma enzima cataliticamente ativa (SMALL et al., 1998, CHRISTENSEN et al., 2005). No presente estudo, todas as enzimas analisadas possuíam um quadro de leitura funcional, com ATGs na extremidade 5' e *stop* códon na extremidade 3'.

Se a sequência entre dois ATGs contém uma sequência que direciona a localização subcelular (muitos dos quais são N-terminal), então a proteína a ser iniciada a partir da sequência a montante conterá esta informação, ao passo que a proteína iniciada à jusante

não conterà. Isto, naturalmente, conduzirá a segmentação diferencial das duas proteínas com provável localização e função diferentes (KOCHETOV; SARAI, 2004, KOCHETOV, 2005).

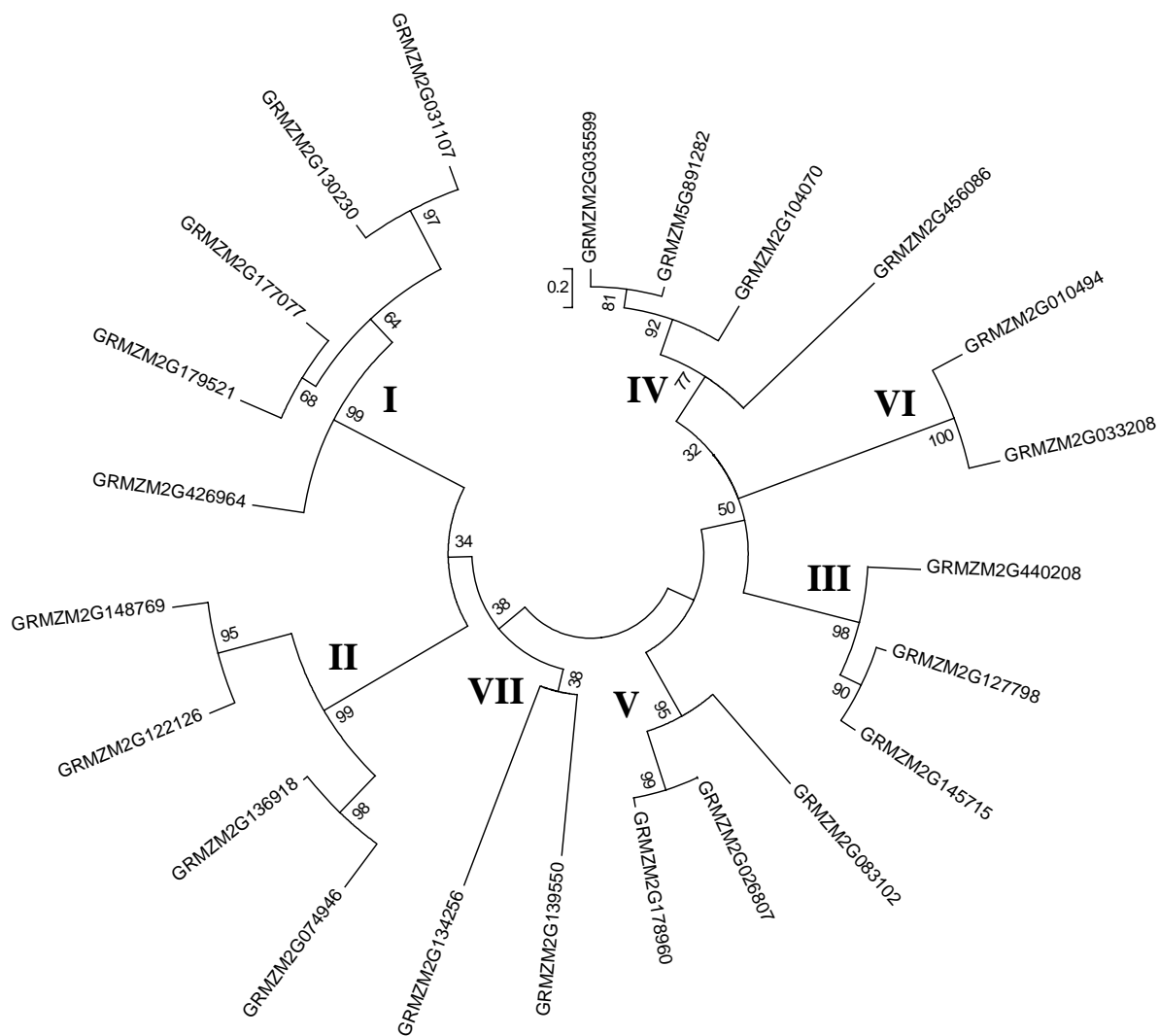
Ao analisar o alinhamento das sequências nucleotídicas, a distância genética (Tabela 4) e a árvore filogenética (Figura 5) entre os genes que codificam as mesmas enzimas da via das pentoses fosfato em milho percebeu-se graus diferentes de similaridade entre os genes, entretanto, não houve a identificação de sequências totalmente iguais. Além disso, ao comparar a estrutura dos genes, foi observado que, apesar de alto grau de similaridade encontrada, a estrutura gênica apresenta muitas diferenças em relação a composição, quantidade e tamanho de éxons e íntrons (Figura 6).

O agrupamento produzido pelos genes da via das pentoses fosfato em milho compôs sete grupos, um de cada enzima, o que indica alta similaridade entre as duplicações gênicas com mesma função, tratando-se assim, de genes parálogos (Figura 5).

Os sete agrupamentos foram classificados em algarismos romanos, e seguiu-se a ordem da via metabólica, sendo **I** formado pelos genes homólogos codificadores da enzima Glicose-6-fosfato desidrogenase; **II**: 6-fosfato gluconolactonase; **III**: 6-fosfogluconato desidrogenase; **IV**: Ribose 5-fosfoisomerase; **V**: Ribulose-fosfato 3-epimerase; **VI**: Transcetolase e **VII**: Transaldolase.

**Tabela 4:** Estimativas da diferença da composição de bases entre as sequências analisadas. A análise envolveu 23 sequências de nucleotídeos. Todas as posições que contêm lacunas e dados faltantes foram eliminadas. Houve um total de 454 posições no conjunto de dados finais. As análises foram realizadas no MEGA 5.1.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1. GRMZM2G179521																							
2. GRMZM2G130230	3.23																						
3. GRMZM2G031107	5.17	0.23																					
4. GRMZM2G426964	7.71	0.97	0.25																				
5. GRMZM2G177077	5.66	0.36	0.02	0.16																			
6. GRMZM2G074946	3.20	0.38	0.68	1.54	0.77																		
7. GRMZM2G122126	0.34	4.76	6.96	9.81	7.52	5.04																	
8. GRMZM2G148769	0.39	3.93	5.89	8.47	6.39	4.35	0.08																
9. GRMZM2G136918	4.73	0.43	0.41	0.87	0.46	0.28	6.99	6.17															
10. GRMZM2G127798	0.68	4.55	6.52	9.22	6.94	3.69	0.98	1.09	5.80														
11. GRMZM2G145715	0.77	3.54	5.21	7.60	5.56	2.72	1.24	1.18	4.64	0.10													
12. GRMZM2G440208	5.25	16.69	20.74	25.56	21.69	16.11	3.94	5.07	19.65	4.90	6.35												
13. GRMZM2G104070	0.62	1.50	2.89	4.78	3.31	2.05	1.34	1.09	2.81	2.24	1.98	8.97											
14. GRMZM2G035599	1.67	5.14	7.28	9.89	7.94	6.59	1.41	1.42	7.69	4.27	4.51	6.81	1.31										
15. GRMZM5G891282	1.13	4.09	6.13	8.62	6.75	5.14	1.37	1.42	6.05	3.53	3.70	6.98	0.72	0.21									
16. GRMZM2G456086	4.75	3.76	4.85	6.23	5.34	5.99	5.37	4.83	5.69	8.75	8.32	15.74	2.24	1.91	1.85								
17. GRMZM2G026807	2.13	0.94	1.60	2.85	1.80	1.59	2.56	1.79	2.49	3.01	2.26	12.72	1.27	3.62	3.33	4.04							
18. GRMZM2G083102	0.55	1.42	2.79	4.68	3.20	1.88	1.28	1.03	2.68	2.02	1.75	8.89	0.01	1.47	0.85	2.51	1.17						
19. GRMZM2G178960	2.78	0.64	1.05	2.04	1.18	1.22	3.46	2.57	1.92	3.61	2.70	14.56	1.65	4.57	4.13	4.42	0.08	1.53					
20. GRMZM2G010494	3.68	0.31	0.43	1.12	0.48	0.07	5.41	4.58	0.34	4.14	3.06	17.20	2.37	7.00	5.68	6.14	1.33	2.19	0.90				
21. GRMZM2G033208	0.48	2.20	3.80	5.94	4.22	1.89	1.62	1.63	2.85	1.34	1.23	7.83	0.54	2.65	1.54	4.51	2.43	0.46	2.78	2.49			
22. GRMZM2G139550	2.15	0.14	0.69	1.77	0.87	0.26	3.57	2.94	0.60	3.19	2.39	14.07	0.95	4.36	3.28	3.91	0.84	0.86	0.69	0.33	1.28		
23. GRMZM2G134256	1.88	0.28	0.99	2.20	1.22	0.43	3.35	2.83	0.72	3.13	2.44	13.35	0.69	3.81	2.69	3.42	1.14	0.62	1.06	0.62	0.93	0.06	



**Figura 5:** Árvore filogenética dos genes que codificam as enzimas da via das pentoses fosfato em milho utilizando a composição de nucleotídeos, por meio do método estatístico da máxima verossimilhança, pelo modelo de Tamura-Nei. Os valores em cada nó dos ramos indicam os números de réplicas de *bootstrap* (total = 1000). Os sete agrupamentos foram classificados em algarismos romanos, e seguiu-se a ordem da via metabólica, sendo **I** formado pelos genes homólogos codificadores da enzima Glicose-6-fosfato desidrogenase; **II**: 6-fosfogluconolactonase; **III**: 6-fosfogluconato desidrogenase; **IV**: Ribose 5-fosfato isomerase; **V**: Ribulose-fosfato 3-epimerase; **VI**: Transcetolase e **VII**: Transaldolase.



**Figura 6:** Revisão esquemática da estrutura dos genes identificados no genoma do milho que codificam as enzimas da via das pentoses fosfato. Os éxons estão representados em caixas. Os íntrons estão indicados em linhas contínuas.

Durante a evolução, genes e suas regiões regulatórias sofrem uma variedade de modificações, incluindo substituições nucleotídicas, duplicações, recombinações e eventos de reparo (LIBERLES, 2001). As duplicações gênicas produzem duas cópias idênticas e estas sequências podem evoluir de diferentes formas. A preservação pode levar à persistência de duas cópias com sequências similares com funções análogas ou com padrões de diversificação (GONZÁLEZ-DUARTE; ALBALAT, 2005). Nesse trabalho, apesar da identificação de diferentes padrões de distância genética e similaridade entre os genes que codificam as mesmas enzimas (Tabela 4 e Figura 5), as funções permaneceram análogas.

Análises filogenéticas pelos métodos da máxima verossimilhança com o alinhamento de 25 sequências completas dos genes da via das pentoses fosfato em *Arabidopsis* e 23 sequências em milho, agrupou a maioria das sequências em grupos que representam as enzimas da via das pentoses fosfato, entretanto, não houve uma linearidade na formação de ramos entre as vias oxidativa e não oxidativa, como se esperava. Foram formados basicamente três grupos: (II, V, VI e VII); (I) e (III, IV, VII) (Figura 7). Além disso, o agrupamento das sequências interespecíficas e intraespecíficas ocorreram aleatoriamente, o que sugere laços filogenéticos entre os genes das duas espécies analisadas.

Por se tratar de uma árvore que agrupou apenas sequências proteicas de duas espécies, não existe a possibilidade de se predizer a origem evolutiva dessas enzimas e afirmar se as primeiras duplicações gênicas ocorreram antes ou depois da especiação, ou seja, se os genes são ortólogos (homólogos que divergiram devido à especiação) ou parálogos (homólogos que divergiram depois de duplicação gênica). Entretanto, pode-se identificar genes homólogos e proteínas análogas, resultante de eventos evolutivos independentes. Afinal, como sabido, essas proteínas possuem a mesma função, embora possam apresentar diferentes estruturas primárias, secundárias e terciárias, bem como os genes que as codificam possuem estruturas e sequências diferentes (DOYLE; GAUT, 1999; GONZÁLEZ-DUARTE; ALBALAT, 2005).

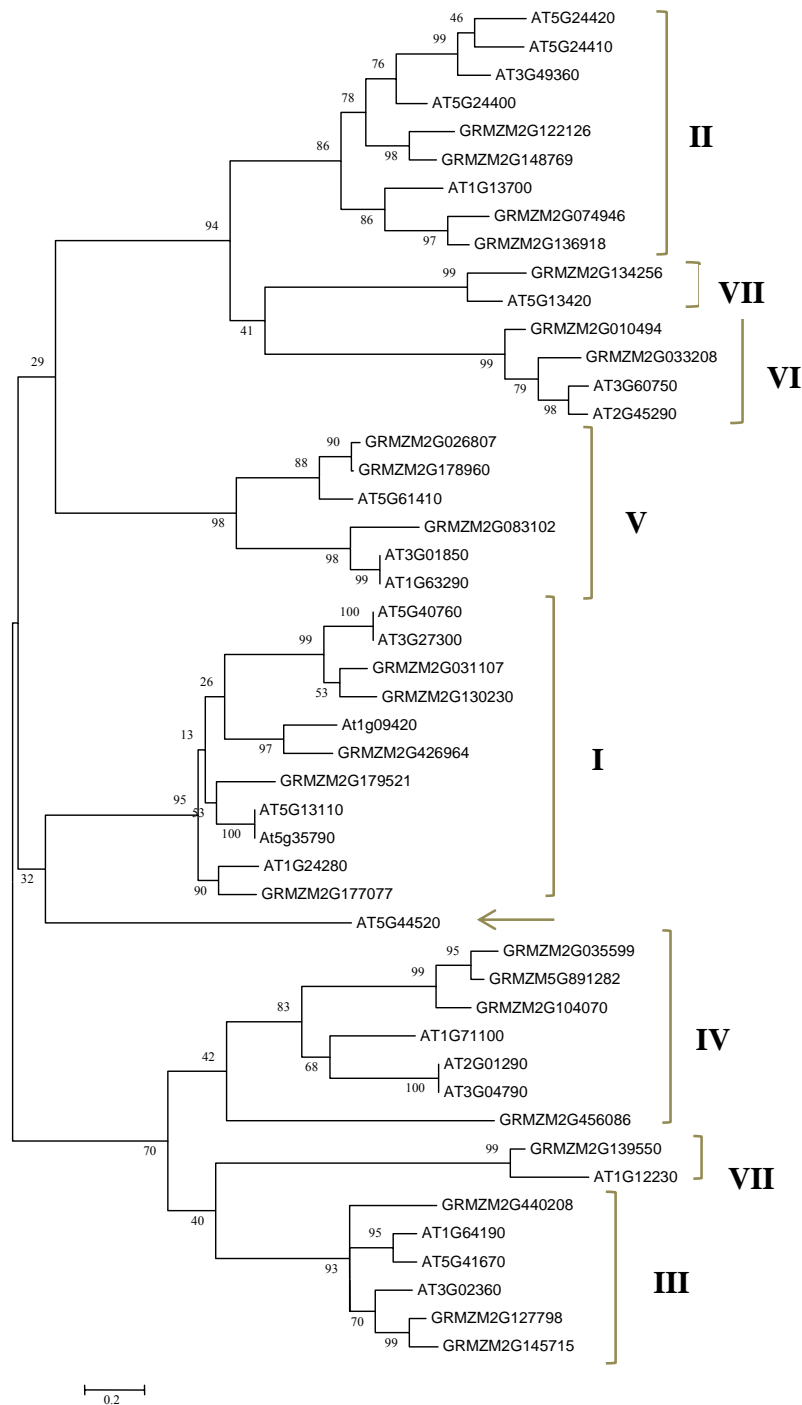
Vale salientar também que uma árvore filogenética construída à partir de sequências de DNA ou proteínas não concorda, necessariamente, com a árvore que

representa a verdadeira história evolutiva das espécies ou famílias gênicas envolvidas (PAGE, et., 2007).

Analisando o grupo I (Figura 7), pode-se verificar sequências com alto grau de similaridade em *Arabidopsis*, formado pelos genes (AT5G35790 e AT5G13110) e (AT3G27300 e AT5G40760). Tratam-se de duas duplicações gênicas na qual, os dois genes se mantiveram conservados ao longo da evolução, tanto ao nível de sequência, quando em suas funções.

No agrupamento II (Figura 7), os genes intraespecíficos tiveram maior padrão de similaridade, e assim, apresentaram menor distância genética. Entretanto, o gene AT5G24400 foi mais similar aos genes de milho GRMZM2G074946 e GRMZM2G136918, que os genes da mesma espécie.

No grupo IV (Figura 7), formado por quatro genes homólogos em *Arabidopsis* e 4 genes em milho, o agrupamento formado foi subdividido em dois subgrupos, formados por sequências intraespecíficas. Entretanto, o gene de milho GRMZM2G456086 se separou dos dois agrupamentos. Analisando a árvore, sugerimos que, durante a evolução, uma duplicação gênica deu origem a esse gene e que, ao longo do tempo permaneceu pouco alterado, diferentemente do ocorrido com os demais genes. Fato interessante ocorreu, também, no agrupamento do gene AT5G44520. Este apresentou sequências mais semelhantes as dos genes que codificam a enzima Glicose-6-fosfato desidrogenase, que aos demais genes que codificam a Ribose-5-fosfato isomerase. Supõe-se que, durante o período de evolução, esse gene tenha se duplicado e sofrido mutações, acarretando nova função.



**Figura 7:** Árvore filogenética dos genes que codificam as enzimas da via das pentoses fosfato em *Arabidopsis thaliana* e milho utilizando a composição de aminoácidos, por meio do método estatístico da máxima verossimilhança, utilizando o modelo de Tamura-Nei para quaisquer correções. Os valores em cada nó dos ramos indicam os números de réplicas de bootstrap (total = 1000). Os sete agrupamentos foram classificados em algarismos romanos, e seguiu-se a ordem da via metabólica, sendo **I** formado pelos genes homólogos codificadores da enzima Glicose-6-fosfato desidrogenase; **II**: 6-fosfogliconolactonase; **III**: 6-fosfogliconato desidrogenase ; **IV**: Ribose 5-fosfato isomerase; **V**: Ribulose-fosfato 3-epimerase; **VI**: Trancetolase e **VII**: Transaldolase.

No grupo V (Figura 7), foram formados dois subgrupos, um contemplando os genes de milho (GRMZM2G026807 e GRMZM2G178960) e o gene de *Arabidopsis* AT5G61410. O outro grupo foi formado pelos genes AT3G01850 e AT1G63290 (*Arabidopsis*) e GRMZM2G083102 (milho). Os genes de *Arabidopsis* AT3G01850 e AT1G63290 mostraram-se bastante similares, com uma probabilidade de ramo de 99%, tratando-se de uma redundância gênica com função análoga.

Os genes de milho e *Arabidopsis*, que codificam a transcetolase, foram agrupados em um único grupo. Para esses genes, a similaridade foi maior entre os genes da mesma espécie.

Os dois subgrupos formados pelos genes que codificam a enzima transaldolase não apresentaram alta similaridade entre si e se agruparam a diferentes enzimas da via das pentoses fosfato. Um dos grupos formados pelos genes GRMZM2G134256 e AT5G13420 está mais relacionado, por similaridade, aos genes que codificam a enzima transcetolase, outra enzima da via não oxidativa, entretanto, o subgrupo formado pelos genes GRMZM2G139550 e AT1G12230, possui maior similaridade aos genes do ramo oxidativo, cofatores da 6-fosfogliconato desidrogenase.

Em outra análise, verificou-se que, apesar de *Arabidopsis* possuir maior número de genes que codificam as enzimas da via das pentoses fosfato (25), comparado ao milho (23), a espécie possui maior número de genes redundantes. Foram identificados quatro pares de genes com sequências altamente homólogas (>99%), enquanto que em milho, não foram verificadas sequências com esse grau de similaridade, possivelmente porque os genes parálogos provenientes das duplicações na espécie foram se divergindo ao longo do tempo devido ao acúmulo de mutações.

A detecção da variabilidade existente entre os genes verificados nesse trabalho foi importante no estudo da divergência funcional das famílias gênicas apresentadas. Se a função da proteína mudou, alguns resíduos de aminoácidos podem ter sido alterados durante a evolução. Isto sugere que taxas evolutivas nesses sítios variaram em genes homólogos diferentes de uma mesma família gênica (GU; VELDEN, 2002).

### 5.1.5. CONCLUSÕES

Os resultados obtidos neste trabalho permitiram as seguintes conclusões:

- Vinte e três genes e as enzimas da via das pentoses fosfato em milho foram identificados e caracterizados.
- Existem pelo menos dois genes homólogos que codificam cada uma das enzimas da via das pentoses fosfato, em milho.
- O espécie *Zea mays* possui menor número de genes que codificam a via das pentoses fosfato e maior divergências entre os genes que codificam as mesmas enzimas quando comparado a *Arabidopsis thaliana*.
- Não houve uma linearidade na formação de ramos entre as vias oxidativa e não oxidativa nas análises de filogenia entre as duas espécies analisadas.
- Foi possível caracterizar os genes e enzimas da via das pentose fosfato em milho.

### 5.1.6. REFERÊNCIAS BIBLIOGRÁFICAS

BERG I.A. Ecological aspects of the distribution of different autotrophic CO<sub>2</sub> fixation pathways. **Applied Environmental Microbiology**, v. 77, p. 1925–1936. 2011.

CHENNA, R.; SUGAWARA, H.; KOIKE, T.; LOPEZ, R.; GIBSON, T.J. Multiple sequence alignment with the Clustal series of programs. **Nucleic Acids Res**, v. 31: p. 3497-3500. 2013.

CHRISTENSEN, B.; NIELSEN, M.S.; HASELMANN, K.F.; PETERSEN, T.E.; SORENSEN, E.S. Post-translationally modified residues of native human osteopontin are located in clusters: Identification of 36 phosphorylation and five O-glycosylation sites and their biological implications. **Biochem J**, v. 390(Pt 1), p. 285–292. 2005.

DEBNAM, P.M.; EMES, M.J. Subcellular distribution of enzymes of the oxidative pentose phosphate pathway in root and in leaf tissues. **J Exp Bot**. v. 50, p. 1653–1661. 1999.

DOYLE, J.J.; GAUT, B.S. Evolution of genes and taxa: a primer. **Plant Mol Biol** 42: 1-23. p. 1653–1661. 1999.

GONZALEZ-DUARTE R AND ALBALAT R. Merging protein, gene and genomic data: the evolution of the MDR-ADH family. **Heredity**, v. 95(3), p. 184-97. 2005.

GU, X.; VENDEL. K. DIVERGE: phylogeny-based analysis for functional-structural divergence of a protein family. **Bioinformatics**. v.18, n.3, p. 500-501. 2002.

HUGHES, A.L. The evolution of functionally novel proteins after gene duplication. **Proceedings of the Royal Society of London**, v. 256, p. 119–124. 1994.

JEANMOUGIN, F.; THOMPSON, J.D.; GOUY, M.; HIGGINS, D.G.; GIBSON, T.J. Multiple sequence alignment with Clustal X. **Trends Biochem Sci.**, v. 23, p. 403–405. 1998.

KOCHETOV, A.V. AUG codons at the beginning of protein coding sequences are frequent in eukaryotic mRNAs with a suboptimal start codon context. **Bioinformatics**, v. 21(7): 837-840. 2005.

KOCHETOV, A.V.; SARAI, A. Translational polymorphism as a potential source of plant proteins variety in Arabidopsis thaliana. **Bioinformatics**, v. 20(4), p. 445-447. 2004.

KRUGER, N.J, SCHAEWEN, A. V. The oxidative pentose phosphate pathway: structure and organisation». **Curr. Opin. Plant Biol.** v. 6 (3). p. 236–46. 2003.

LEVIN, D.A. Polyploidy and novelty in flowering plants. **American Naturalist**, v. 122, p. 1–25. 1983.

LI, W.H. **Molecular evolution**. Massachusetts: Sinauer, 487p. 1997.

LIBERLES, D.A. Evaluation of methods for determination of a reconstructed history of gene sequence evolution. **Mol. Biol. Evol.** v. 18, p. 2040–2047. 2001.

LIEW, A.; YAN, W.C. H.; Yang, M. Pattern recognition techniques for the emerging field of bioinformatics: A review. **Pattern Recognition**, v. 38(11), p. 2055–2073, 2005.

NEI, M.; KUMAR, S. **Molecular evolution and phylogenetics**. Oxford: Oxford University Press, 333 p. 2000.

NEUHAUS, H.E., M.J. EMES. Nonphotosynthetic Metabolism in Plastids, **Annual Review of Plant Physiology and Plant Molecular Biology**, v. 51, p. 111-40. 2000.

NICHOLAS, K.K.; NICHOLAS, H.B.J.R. GeneDoc: a tool for editing and annotating multiple sequence alignments. GeneDoc, Multiple sequence alignment editor & shading utility (version 2.6.001). **Copyright** 2000. 1997.

OHNO S. **Evolution by gene duplication**. New York: Springer-Verlag. 1970.

PAGE, R.D.M.; CHARLESTON, M.A. From gene to organismal phylogeny: Reconciled trees and the gene tree/species tree problem. **Mol. Phyl. Evol** 7 (2): 231-240. 2007.

PERTEA, M.; S. L. SALZBERG . Computational gene finding in plants. **Plant Molecular Biology**, v. 48(1-2), p. 39-48. 2002.

SCHNARRENBERGER, C.; FLECHNER, A.; MARTIN, W. Enzymatic evidence for a complete oxidative pentose phosphate pathway in chloroplasts and an incomplete pathway in the cytosol of spinach leaves. **Plant Physiol**. v. 108, p. 609–614. 1995.

SMALL, D. J., VAYDA, M. E. AND SIDELL, B. D. A novel vertebrate myoglobin gene containing three A+T-rich introns is conserved among Antarctic teleost species, which differ in myoglobin expression. **J. Mol. Evol**. v. 47, p. 156-166. 1998.

TAMURA, K.; PETERSON, D.; PETERSON, N.; STECHER, G.; NE, M.; KUMAR, S. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. **Molecular Biology and Evolution**, v. 28, p. 2731-2739. 2011.

THOMPSON, J.D.; HIGGINS D.G.; GIBSON T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucleic Acids Research**, v.22, p.4673-4680, 1994.

WOLFE, K. H. Yesterday's polyploids and the mystery of diploidization. **Nature Reviews Genetics**, New York, v.2, p. 333-341. 2001.

**5.2. ANÁLISES COMPARATIVAS SOBRE MÉTODOS EXPERIMENTAIS E *IN SILICO* PARA A DETERMINAÇÃO DA LOCALIZAÇÃO SUBCELULAR DAS ENZIMAS DA VIA DAS PENTOSE FOSFATO DE MILHO**

## RESUMO

O conhecimento da localização subcelular de uma proteína pode fornecer informações importantes sobre a sua função e os mecanismos regulatórios na célula. Assim, identificar e caracterizar o meio em que as proteínas se expressam é um dos desafios para o entendimento global do funcionamento celular. Neste estudo, buscou-se determinar a localização subcelular das enzimas da via das pentoses fosfato em milho por meio da determinação experimental e da predição realizada por *softwares* disponíveis na internet. Todas as dez enzimas experimentalmente caracterizadas dessa via foram expressas no cloroplasto. Estes resultados foram confrontados aos dados gerados por 13 preditores de localização subcelular. A comparação revelou diferenças marcantes entre os *softwares* e na precisão de suas previsões, com acurácias que variaram de 40% a 90%, na qual revelou serem os *softwares* SLPFA e Euk-mPloc os mais acurados. Além disso, analisando individualmente cada uma das enzimas, a acurácia das predições variou de 7% a 84%. Neste trabalho, os preditores que utilizam métodos de previsões múltiplas proporcionaram um desempenho significativamente melhor quando comparados aos que se baseiam apenas em sequências ou peptídeos sinais, revelando assim, os *softwares* mais recomendados para previsões de localização subcelular em milho.

**Palavras chave:** Sequências proteicas, preditores, compartimentalização celular, citoplasma, cloroplasto, acurácia.

## ABSTRACT

The knowledge about the protein subcellular localization can provide important information about its function and regulatory mechanisms in the cell. Thus, identifying and characterizing the environment in which the proteins are expressed is one of the challenges for understanding the overall operation of the cell. In this study, It was sought to determine the subcellular localization of the pentose phosphate pathway enzymes in maize through the experimental determination and predictions by software available on the internet. Ten enzymes were experimentally characterized, and all enzymes analyzed were expressed in the chloroplast. These results were compared to data generated by 13 subcellular localization predictors. The comparison revealed striking differences between the softwares and the accuracy of their predictions, with accuracies ranging from 40% to 90%, which proved to be the software SLPFA and EUK-mPloc the most accurate. Moreover, analyzing individually each enzymes, the accuracy of the predictions ranged from 7% to 84%. In this work, the predictors that use multiple methods provided a significantly better performance compared to those based only on sequences or peptide signals, thus revealing the most recommended software for predict subcellular localization in maize.

**Keywords:** Protein sequences, predictors, cellular compartmentalization, cytoplasm, plastid, accuracy.

### 5.2.1. INTRODUÇÃO

Uma célula eucariótica, por processo de evolução, desenvolveu um sistema de endomembranas que são especializados em diferentes funções celulares. A constituição e localização desses microambientes subcelulares, denominados organelas, capacitou o desempenho de processos bioquímicos extremamente específicos. O perfeito funcionamento de uma organela depende de seu conteúdo, que é definido de acordo com as suas funções (SHEN; BURGER, 2007).

As células de plantas são altamente organizadas e muitos processos biológicos estão associados com estruturas subcelulares específicas. As proteínas podem existir ou mover-se, simultaneamente, entre duas ou mais diferentes localizações subcelulares. Proteínas com vários locais ou características de localização dinâmicas são particularmente interessantes, porque podem ter funções biológicas especiais. A localização subcelular é uma característica chave das proteínas, visto que está relacionada com a função biológica e sua interação com o meio (PUJOL et al., 2007).

Atualmente, a identificação da localização subcelular de uma proteína utiliza técnicas eficientes, principalmente no que se refere ao sistema de clonagem e expressão, pois permitem gerar rapidamente fusões com o gene *gfp* (GFP, *green-fluorescent protein*) tanto na extremidade amino quanto na carboxi-terminal. Por essa metodologia cerca de 80% das construções gênicas permite detectar claramente a localização intracelular (SIMPSON *et al.*, 2000).

O primeiro passo para determinar experimentalmente a localização subcelular de uma enzima consiste na análise das informações genômicas e a identificação da sequência codificante completa (CDS, *complete coding sequence*) de cada gene. O CDS é a sequência de nucleotídeo que corresponde à sequência de aminoácido da proteína, começando tipicamente com o ATG e terminando com o *stop codon*.

Apesar dos recentes avanços nas técnicas, a determinação experimental da localização subcelular de proteínas continua moroso e caro. Dada a complexidade, sistemas de predição de localização são interessantes para identificar e examinar possíveis candidatos para posterior análise experimental (XIE et al., 2005). Nesse sentido, as

predições *in silico* tem recebido considerável atenção nos últimos anos e muito progresso foi alcançado, entretanto, ainda não apresentam uma resolução satisfatória.

Os atuais preditores disponíveis para a localização subcelular das proteínas podem ser agrupados em duas categorias principais. Um deles baseia-se na existência de sinais nas sequências N-terminais, incluindo peptídeos sinais, peptídeos mitocondriais e peptídeos cloroplásticos (TargetP, iPSORT, ), ou na composição de aminoácidos das sequências de proteína em diferentes localizações subcelulares (WoLF PSORT, CeLLO, SubLoc). A outra categoria se baseia na multiplicidade de ferramentas para a execução da predição da localização subcelular, na qual utilizam-se parâmetros como de redes neurais e cadeias de Markov (PredSL), ontologia do gene, filogenia e domínio funcional da proteína (Plant-mPLOC, SherLoc2, SubCellProt, SLPFA, Euk-mPLOC 2.0 e MultiLoc2).

Como a acurácia de predição dos vários *softwares* disponíveis tem se mostrado abaixo do esperado, tem-se buscado desenvolver *softwares* preditores na tentativa de aumentar a confiabilidade da predição (HEAZLEWOOD et al., 2004). Nesse sentido, o presente estudo buscou caracterizar, experimentalmente, a localização subcelular das enzimas da via das pentoses fosfato e, a partir disso, identificar os preditores mais acurados, por meio da comparação dos resultados de ensaios experimentais e de predição.

## 5.2.2. MATERIAL E MÉTODOS

### 5.2.2.1. Predição da localização subcelular das enzimas da via das pentoses fosfato em milho

As sequências dos genes que codificam as enzimas da via das pentoses fosfato foram mineradas a partir da versão atual da anotação do genoma da linhagem de milho B73 (<http://www.maizesequence.org>). Cada sequência prevista de mRNA ou cDNA foi cuidadosamente examinada na extremidade 5' e 3' para garantir que o códon de iniciação e parada estivessem corretos.

Duas localizações subcelulares, cloroplasto e citoplasma, foram classificadas para as proteínas de milho. Treze ferramentas de predição foram selecionadas para a localização subcelular: TargetP (EMANUELSSON et al, 2007), Wolf PSORT (HORTON et al., 2007), iPSORT (BANNAI et al, 2002), CeLLO (YU et al., 2006), SubLoc (CHEN, et al., 2001), PredSL (PETSALAKI et. al., 2006), Plant-mPLoc (CHOU; SHEN, 2010B), SherLoc2 (BRIESEMEISTER, et al., 2009), Multiloc2 (BLUM, et al., 2009), SubCellProt (GARG et al., 2009), SLPFA (TAMURA; AKUTSU, 2007), SLP-Local (MATSUDA, et al., 2005) e Euk-mPLoc 2.0 (CHOU e SHEN, 2010A). A seleção dos *softwares* foi baseada na diversidade dos algoritmos, banco de dados utilizados e na apresentação das características empregadas.

### 5.2.2.2. Avaliação experimental da localização subcelular das enzimas da via das pentoses fosfato em milho

#### 5.2.2.2.1. Material Genético

Plantas de milho (*Zea mays* L.) provenientes de linhagens puras de W22 foram cultivadas em casas de vegetação e em campos experimentais da Universidade da Flórida, durante o verão de 2011, em Gainesville, Flórida, nos EUA. Folhas jovens foram

coletadas, identificadas em envelope de papel, imersas imediatamente em nitrogênio líquido e mantidos em *ultrafreezer* a  $-80^{\circ}\text{C}$ , para posterior extração de RNA.

#### **5.2.2.2.2. Extração de RNA**

O RNA total foi extraído a partir de 0,1 g de folhas de plantas de milho, as quais foram maceradas em nitrogênio líquido ( $\text{N}_2$ ), transferidas para um tubo de 2 mL, homogeneizadas em 1,0 mL do reagente Trizol (Life Technologies) e mantidas por 5 minutos a temperatura ambiente. Em seguida, foram adicionados 0,6 mL de clorofórmio, agitadas vigorosamente e mantidas por 5 min a temperatura ambiente. Após centrifugação a 12.000 g por 15 min a  $4^{\circ}\text{C}$ , o sobrenadante foi coletado e transferido para um tubo novo de 2 mL. Adicionou-se ao sobrenadante, 0,4 mL de álcool isopropílico e a mistura foi mantida à temperatura ambiente por 10 min e novamente centrifugada a 12.000 g por 15 min, a  $4^{\circ}\text{C}$ . O sobrenadante foi descartado e o precipitado lavado com 1 mL de etanol 75 %. As amostras foram agitadas vigorosamente e centrifugadas a 8.000 g por 5 min. Logo em seguida, foi feita a secagem do RNA precipitado em bomba de vácuo por 10 min. O precipitado de RNA foi dissolvido em 50  $\mu\text{L}$  de água livre de RNase (tratada com DEPC, 0.1% v/v) e armazenado à  $-80^{\circ}\text{C}$ . O RNA total extraído foi quantificado em espectrofotômetro, na diluição de 1:10 v/v. As medidas de absorvância foram realizadas nos comprimentos de onda de 260 e 280 nm, cuja relação entre as leituras fornece uma estimativa de pureza do ácido nucléico (SAMBROOK et al., 1989). Utilizou-se água-DEPC para o ajuste inicial do aparelho (Biophotometer/8,5 mm/Eppendorf).

#### **5.2.2.2.3. Obtenção de cDNA**

As amostras de RNA foram tratadas com DNase (Deoxiribonuclease I, Invitrogen) para remoção de possíveis resíduos de DNA genômico. A reação de transcrição reversa foi realizada utilizando-se 2  $\mu\text{g}$  de RNA total, transcriptase reversa (Superscript III, Invitrogen) e Oligo dT(12-18) *primer* (Invitrogen) em presença de inibidores de RNase

(RNase OUT, Invitrogen) em um volume final de 20  $\mu$ L. A ordem e as etapas de incubação foram realizadas de acordo com as recomendações de tempo e temperaturas do fabricante dos reagentes.

#### **5.2.2.2.4. Amplificação dos genes da PPP**

Uma vez obtido o cDNA, a amplificação visando à obtenção da sequência completa do gene, sequência *full-length* (incluindo região codificante e não codificante), bem como amplificações dos CDSs (sequência codificante dos genes) a serem utilizados nas clonagens, foi realizada. Dois oligonucleotídeos foram produzidos para cada amplificação específica, totalizando-se, assim, 44 oligonucleotídeos para cada uma das etapas. O gene GRMZM2G440208, codificador da enzima 6-fosfogluconato desidrogenase, já foi caracterizado anteriormente (SPIELBAUER et al., 2006), por isso, não foi amplificado.

Os oligonucleotídeos para a amplificação *full-length* foram desenhados para se anelar nas regiões 5' UTR e 3' UTR de cada gene (Tabela 5). Para a amplificação dos CDSs, os oligonucleotídeos *forward* foram desenhados para permitir a amplificação a partir do ATG, codificador do aminoácido metionina no início da síntese protéica, já os oligonucleotídeos *reverse* foram desenhados com a remoção do *stop codon* dos CDSs para a subsequente fusão na extremidade C-terminal da proteína GFP (Tabela 6). Outra característica dos oligos *forward* é a presença dos nucleotídeos 5' CACC 3', na extremidade 5', necessários para a realização da clonagem direcional no vetor pENTR/D-TOPO.

As reações de PCR foram realizadas em placa de 96 amostras, com um gradiente de temperatura de  $60 \pm 5$  °C, em termociclador 5331 (Eppendorf, USA). Obtido os CDSs, foram realizadas as clonagens no vetor de entrada pENTR/D-TOPO (Invitrogen, USA) compatíveis com o sistema *Gateway* (Invitrogen, USA). Estas construções foram então recombinadas com o vetor pB7FWG2 (KARIMI et al., 2002), obtendo-se assim construções que permitem a fusão da proteína GFP na extremidade carboxi das proteínas de interesse.

**Tabela 5:** Lista dos *Primers* para a amplificação dos *full-length* dos genes da via das pentoses fosfato em milho.

Enzima	Gene Model	Primers ORF (5' - 3')	
		Forward	Reverse
Glicose-6-fosfato desidrogenase	GRMZM2G179521	AAGCAATCGCCAAGAACATC	CCCTTCCACAATAAGGCGTA
Glicose-6-fosfato desidrogenase	GRMZM2G130230	AGAGCTCTTCCCTCCCACAC	TGAATGGAACACGCAATTTT
Glicose-6-fosfato desidrogenase	GRMZM2G031107	GTTCTCCGCTCTTGGTCTTG	CCGTAAGGGTCACTCAAACA
Glicose-6-fosfato desidrogenase	GRMZM2G426964	TGTAGATGCCGACATGACAAC	TCCTTGAAGGCATGTCTCC
Glicose-6-fosfato desidrogenase	GRMZM2G177077	CCAAAGTAAGAAAGCCACACG	CGAGGGCGTGCAAACCTAC
6-fosfogluconolactonase	GRMZM2G122126	CCGTAAGGGTCACTCAAACA	CCCTTCCACAATAAGGCGTA
6-fosfogluconolactonase	GRMZM2G074946	AATCCACCCTAAACCATAG	GGATAAAACCGCAACACAC
6-fosfogluconolactonase	GRMZM2G148769	ACGGAGCATCCCCCTTAATA	GAAATCCTGGAGGACGAATG
6-fosfogluconolactonase	GRMZM2G136918	GCAGGAAGTGCAGAGGAGAA	AATTCCGGTGTCTTTGACG
6-fosfogluconato desidrogenase	GRMZM2G127798	TAAGTGGTGGGGGAGGGAAG	GGAGCATGGTGCATTGATTA
6-fosfogluconato desidrogenase	GRMZM2G145715	GCAGGTAGTAACGTAGGCTG	ATGTTCCGATCTTGAGTTGCC
Ribose 5-fosfato isomerase	GRMZM2G104070	AGCCGCTCTGGGAATCTGAC	TGGAACCTGCACATCCAGAA
Ribose 5-fosfato isomerase	GRMZM2G035599	GATTCGCGGATTTGATTCC	CATACTCCATAGTAATCAGCAGACA
Ribose 5-fosfato isomerase	GRMZM5G891282	CATTCACTTGCCCCCTTTT	TGATGATTAACATACTCCTGCAGAC
Ribose 5-fosfato isomerase	GRMZM2G456086	GACCTCACCTCTCCGAACCTG	TGATAGTCAATCCCCACATTG
Ribulose-fosfato 3-epimerase	GRMZM2G026807	ACAAACCAATCGTTACCG	CTCCGTCTCCTTTTCATCTCC
Ribulose-fosfato 3-epimerase	GRMZM2G083102	CCTCACCACTCCCCTACAA	TTTCTGTTTACAGTCATGCCATT
Ribulose-fosfato 3-epimerase	GRMZM2G178960	CAACAAGCAAACAACCCAAAG	GGGTGCCTTTACAATGGAGA
Transketolase	GRMZM2G010494	ATGTTCCGATCTTGAGTTGCG	CCAAAGTAAGAAGCCACACG
Transketolase	GRMZM2G033208	TGGAACCTGCACATCCAGAA	GGGTGCCTTTACAATGGAGA
Transadolase	GRMZM2G139550	CCCCTCTGCAGTCTTCAGTC	TAAGTGCAGCTCAGGTGGTG
Transaldolase	GRMZM2G134256	CTACTTATGGATGCCGACGA	TGGCGACATTGAGTTCAAGA

**Tabela 6:** Lista dos *Primers* para a amplificação dos CDS dos genes da via das pentoses fosfato em milho.

Enzima	Gene Model	Primers ORF (5' - 3')	
		Forward	Reverse
Glicose-6-fosfato desidrogenase	GRMZM2G179521	CACCATGTCCATTTCGACCAT	TAGAGCTTG AAGGCGCCGT
Glicose-6-fosfato desidrogenase	GRMZM2G130230	CACCATGTCCAGGAGGATCTTCA	CTCTTCTTTGCAAGGGTCCG
Glicose-6-fosfato desidrogenase	GRMZM2G031107	CACCATGTCCAGGAGGGTCGTC	CAGCTGAACCTTAGAAAGGGTC
Glicose-6-fosfato desidrogenase	GRMZM2G426964	CACCATGGCTAGATTAAGTTCTAGG	TCGACTCTCCAGTCGTCATCC
Glicose-6-fosfato desidrogenase	GRMZM2G177077	CACCATGGCGCTCTCCTGCAT	CGTTTCGTGCCTTGTAGTGCTC
6-fosfogluconolactonase	GRMZM2G122126	CACCATGATCTCGACGGCTA	GGGCTTCTGATGCTCGACAAC
6-fosfogluconolactonase	GRMZM2G074946	CACCATGGAGAAGTGAATCATG	CCAGTCTGTTCTCGGTACC
6-fosfogluconolactonase	GRMZM2G148769	CACCATGTCCACCTCCGTCTC	GAGACTCTCCTTGTCTGCAGCA
6-fosfogluconolactonase	GRMZM2G136918	CACCATGGAAAAGGAATCATC	GTTAGTCGCTCGTGACGATC
6-fosfogluconato desidrogenase	GRMZM2G127798	CACCATGGCGCTACAAGAATC	GCTCATCGGTTGTTGGAGATG
6-fosfogluconato desidrogenase	GRMZM2G145715	CACCATGGCTCTCACGCAATT	AGTTCCTCCCTGCAAGCAGT
Ribose 5-fosfato isomerase	GRMZM2G104070	CACCATGGTCAGTCCGCCCGC	CAGCACTCCATCTTTGAGCCCTC
Ribose 5-fosfato isomerase	GRMZM2G035599	CACCATGGCCGCCACCTCGAT	CTAACTCCTTCCATGACTGCG
Ribose 5-fosfato isomerase	GRMZM5G891282	CACCATGGGCGAGCCCG	TACTACTGCTCCTTCTCTGCTC
Ribose 5-fosfato isomerase	GRMZM2G456086	CACCATGGCCGCCACCG	AACTCGTCTCGTTTGCACACTG
Ribulose-fosfato 3-epimerase	GRMZM2G026807	CACCATGGCGACGCCGTC	AGCTCTTCTGCGCTACAGC
Ribulose-fosfato 3-epimerase	GRMZM2G083102	CACCATGGCGCGCGGAAGATA	AATCCGTTTTTGTCTGAGAGCCC
Ribulose-fosfato 3-epimerase	GRMZM2G178960	CACCATGGTCATCGCCGCCG	CCTTGTGCGTGTACAGCTGCAGG
Transketolase	GRMZM2G010494	CACCATGGGCTGCGCG	GACCATTTCCAGGGCCTCG
Transketolase	GRMZM2G033208	CACCATGGGCTGCGCG	GCTCTTGAAGCTCTTGGCAGC
Transadolase	GRMZM2G139550	CACCATGTCTGTTACCTCACCC	CGTTGTTGAACATTGGGAGGTG
Transaldolase	GRMZM2G134256	CACCATGACCGGCACGGTG	CTCAGCTTTCACACTGCCAT

### 5.2.2.2.5. Clonagens dos genes de interesse visando fusão com o gene GFP

O protocolo padronizado para a clonagem no vetor de entrada pENTR/D-TOPO foi estabelecido como uma reação contendo 1 µl do produto de PCR, 1 µl da solução de sais, 3 µl de H<sub>2</sub>O esterilizada e 1 µl do vetor TOPO, sendo iniciada a transformação das bactérias competentes após 10 minutos em temperatura ambiente. Para a clonagem no vetor pENTR, o mesmo protocolo foi utilizado. A recombinação do vetor de entrada pENTR/D-TOPO com o vetor de destino pB7FWG2, foi estabelecida da seguinte forma: 1 µl do DNA da clonagem no TOPO, 1,5 µl do DNA do vetor de destino, 5,5 µl de T.E. (pH 8.0) e 2 µl da enzima clonase.

Incubou-se esta reação a 25 °C, por duas horas e meia, sendo que após este período adicionou-se 1 µl de Proteinase K e incubou-se, novamente, a 37 °C por 10 minutos antes de iniciar a transformação das bactérias competentes.

A confirmação e verificação das sequências clonadas foi realizada por meio da PCR, utilizando *primers* específicos (Tabela 6), *primers* internos (Tabela 7) e por sequenciamento (dados não apresentados).

**Tabela 7:** Lista dos pares de *primers* internos para a confirmação das sequências introduzidas nos vetores utilizados.

Enzima	Gene Model	Internal Primers (5' - 3')	
		Forward	Reverse
Glicose-6-fosfato desidrogenase	GRMZM2G179521	GGAACATCCGGAACGTG	CTTCAGCTGCTCATGGAC
Glicose-6-fosfato desidrogenase	GRMZM2G130230	AAGGGCGTGGAGGATATTTT	ACCCGAAGAACAACAGATGC
Glicose-6-fosfato desidrogenase	GRMZM2G031107	TGCAGAGGAACTGAGTGCTG	GCCACCAACAGAAGACCTG
Glicose-6-fosfato desidrogenase	GRMZM2G426964	AGAACCTTGGATGGCCCTAT	CCTGGGCTCTCTGGAGTGTA
Glicose-6-fosfato desidrogenase	GRMZM2G177077	GAGCCTCTCTGGTCAAGACAAT	TGGTACCACCTTTTGTGTGACT
6-fosfogluconolactonase	GRMZM2G122126	TGCCATCAATGACGAACTGT	TGACCCACTTCTGGTTTTCA
6-fosfogluconolactonase	GRMZM2G074946	TTGCAAAGAACCATGCAGAG	CAGAAGGACCAGGTCAATC
6-fosfogluconolactonase	GRMZM2G148769	GATTCAGTGGACTGGAGCAA	CCCAGAAGCTGGAGATCAAA
6-fosfogluconolactonase	GRMZM2G136918	AAGAAGGATTCTCGCCAAG	AGGGTGGTTCGGGAAAAG
6-fosfogluconato desidrogenase	GRMZM2G127798	TCGAGTATGGCGATATGCAG	CTGCTGCACCGTCCACTT
6-fosfogluconato desidrogenase	GRMZM2G145715	GCTTACCCACCGACTTGAGA	GCTCCTTCGAGGCATACAAGT
Ribose 5-fosfato isomerase	GRMZM2G104070	GAAGCATTGTTGAAGGCACTC	CCTCAATTGCACCATACTTTG
Ribose 5-fosfato isomerase	GRMZM2G035599	CTCCGCGAGAAGATGGC	GACGGTGTTGGTCTGAG
Ribose 5-fosfato isomerase	GRMZM5G891282	GACCCGGACCTCAACCTC	ATGACCTCGACGGGGATG
Ribose 5-fosfato isomerase	GRMZM2G456086	TAGCCTCCACCACCAACTTC	CGCTCTTACCTGCTCGAC
Ribulose-fosfato 3-epimerase	GRMZM2G026807	ACTTTGCGAAGCTTGGTGAT	ACTGTTCCGGTCAAATGGAT
Ribulose-fosfato 3-epimerase	GRMZM2G083102	CGCTTCTGGATTACATTCC	AACCGGCTCGACTGCATC
Ribulose-fosfato 3-epimerase	GRMZM2G178960	CAAGGCAGGTGCTGATATTGT	CTTCTCAATTCTGCAATTTTCC
Transketolase	GRMZM2G010494	ATCGATGAGGGTGCTTCACT	CCCTCTGGAAGTCACCAAAC
Transketolase	GRMZM2G033208	GGGAGGTACTTTGGAAATTATGG	CAAGGTTCACTTTCCGAATTG
Transadolase	GRMZM2G139550	CCAGGAAGTTATCGTAATGG	GCCTCAGGTGTTCCAATCTT
Transaldolase	GRMZM2G134256	ACATAAAGATCCCAGCTACTGC	GCCTCAGGTGTTCCAATCTT

#### 5.2.2.2.6. Sequenciamento

Todos os clones que apresentaram resultados positivos para os vetores TOPO e p-ENTR foram sequenciados. Ao todo, 19 amostras foram sequenciadas para cada vetor.

Para a reação de sequenciamento, os reagentes e quantidades utilizadas foram: BigDye terminator Mix (Applied Biosystem®), 1µL; tampão do BigDye terminator Mix, 2µL; primer (Invitrogen®) 4pM, 1µL; água deionizada, 4µL e do produto de PCR purificado, 2µL. Os *primers* utilizados foram o M13 (Life Technologies®) *forward* e *reverse*, bem como os *primers* internos (Tabela 6). As condições da reação foram: 96°C por 10 segundos, 50°C por 5 segundos e 60°C por 4 minutos, por 40 ciclos. Depois da reação de sequenciamento, as amostras foram purificadas, utilizando-se isopropanol 75% e etanol 70%. Assim, sequenciadas pelo método Sanger (1974).

Todo o material proveniente do sequenciamento da linhagem W22 foi alinhado com sequências provenientes do genoma da linhagem B73, encontradas no banco de dados do NCBI, por meio das ferramentas de alinhamento local (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) e do *software* Spidey (<http://www.ncbi.nlm.nih.gov/spidey/>). Além disso, análises dos cromatogramas foram realizadas com a ajuda do *software* Chromas Lite 2.1, para verificação da qualidade do sequenciamento.

#### 5.2.2.2.7. Preparo das plantas de *Nicotiana benthamiana*

Sementes de *Nicotiana benthamiana* foram colocadas em vasos de 0,5L contendo substrato. Vinte e cinco dias após a semeadura, as mudas foram individualmente transplantadas para novos vasos contendo o mesmo meio.

As plantas foram mantidas em câmaras de crescimento vegetal, com temperatura, umidade e fotoperíodo controlados, para seu crescimento. Após 15 dias, as plantas foram utilizadas nos ensaios de expressão transiente dos genes que codificam as enzimas da via das pentoses fosfato em milho, por meio da transformação mediada por *Agrobacterium tumefaciens*, que contém o vetor binário pBIN19 e a cópia do transgene GFP.

#### **5.2.2.2.8. Transformação de *Agrobacterium tumefaciens* pelo método de choque térmico**

Alíquotas de *Agrobacterium* competentes, conservadas em -80 °C foram imediatamente imersas em gelo. Em cada alíquota, foi adicionado 1µg de DNA plasmidial purificado do vetor pB7FWG2, para cada um dos genes clonados. Após o descongelamento e homogeneização da solução, as alíquotas foram novamente congeladas em nitrogênio líquido por 5 min. Posteriormente, as amostras foram deixadas em temperatura ambiente por 10 minutos e logo transferidas para um tubo contendo 2 ml de meio YEP (10 g extrato levedura, 10 g Bacto-peptona, 5 g NaCl), sendo, incubadas a 28 °C, por 4 horas.

As células foram plaqueadas em meio YEP com meio seletivo (estreptomicina e gentamicina) e deixadas na estufa a 37 °C até o aparecimento de colônias isoladas.

#### **5.2.2.2.9. Transformação das plantas via Agroinfiltração**

Uma colônia de *Agrobacterium*, proveniente das placas transformadas com os genes de interesse, foi transferida para 2 mL de meio YEB com os antibióticos específicos e 20 µL de acetosiringona. A cultura de bactérias foi incubada por 16 horas em mesa agitadora orbital (150 rpm), à temperatura 26°C. Uma alíquota de 1 mL foi transferida para meio fresco, permanecendo nas mesmas condições por seis horas, até atingir a densidade ótica (D.O.<sub>600</sub>) de 0,8-1,0 a 660 nm (MATHIS e HINCHEE, 1994). Este meio de cultura foi então centrifugado a 4.000 g, por 8 min. O precipitado foi ressuspensão em uma solução de infiltração contendo 50mM MES pH 5,6, 2% sacarose e 200µM de acetosiringona, ajustando a concentração final para a mesma concentração aferida pela D.O. A suspensão foi coletada em seringa esterilizada e usada para infiltrar a solução na parte inferior das folhas de tabaco. Após a infiltração, as plantas foram recolocadas nas câmaras de crescimento por vinte e quatro horas até as análises microscópicas.

Foram utilizadas duas colônias diferentes provenientes de uma mesma placa transformada. A infiltração foi realizada em duas folhas de duas plantas diferentes, totalizando-se assim, duas repetições biológicas e duas repetições técnicas.

As plantas controle foram infiltradas com o gene pCAMBIA-SERcont-GFP, usado como controle. Esse gene codifica uma proteína que é expressa em todos os compartimentos celulares, com exceção do cloroplasto.

#### **5.2.2.2.10. Análises microscópicas**

As avaliações da localização subcelular foram realizadas em microscópio de fluorescência modelo Leica DMI4000 (Leica, Alemanha), visando assim a detecção (com filtro de 505 a 550 nm) da emissão da GFP (excitada em 488 nm). A captura digital das imagens foi obtida utilizando-se o *software* Leica IM50 (Leica, Alemanha). Os padrões de localização subcelular foram comparados aos controles testados.

As estruturas coradas em verde fluorescente nas figuras indicam que a construção gênica fusionada ao GFP foi expressa no determinado compartimento celular visualizado. O espectro vermelho foi utilizado para destacar auto-fluorescências nas amostras para a certificação que tratam-se de um sinal verdadeiro.

### 5.2.3. RESULTADOS E DISCUSSÃO

As sequências protéicas dos genes que codificam as enzimas da via das pentoses fosfato em milho, identificadas por meio do banco de dados do *Maizesequence*, foram utilizadas para predizer a localização subcelular de cada uma das enzimas. Foram utilizados 13 preditores sendo observada discordância entre os métodos para as mesmas sequências protéicas (Tabela 8).

Pelo fato da grande maioria dos *softwares* existentes para a estimativa da localização subcelular de proteínas serem limitados em relação ao método empregado ou devido ao número de acessos em seu banco de dados, foram adotados vários preditores na tentativa de ampliar a acurácia da real localização enzimática. Em geral, os *softwares* não se restringem a localização de cloroplasto e citoplasma, entretanto, neste trabalho, apenas as duas localizações foram analisadas.

Os resultados representaram uma visão mais realista da distribuição das enzimas na célula de milho, principalmente devido ao número e da exploração adequada das forças combinadas desses *softwares* de previsão. Na Tabela 9 estão agrupadas as localizações subcelulares por meio da soma geral das predições, sendo a predição final aquela que apareceu na maioria das observações. Dos 23 genes que codificam as enzimas que compõem a via das pentoses fosfato, 13 genes (57%) foram preditos como cloroplastídicos; 8 genes (35%) para ambas as localizações; 1 gene (4%) predito para o citoplasma e 1 gene (4%) não teve localização determinada.

**Tabela 8:** Predição da localização subcelular das enzimas da via das pentoses fosfato a partir de 13 diferentes *softwares* preditores.

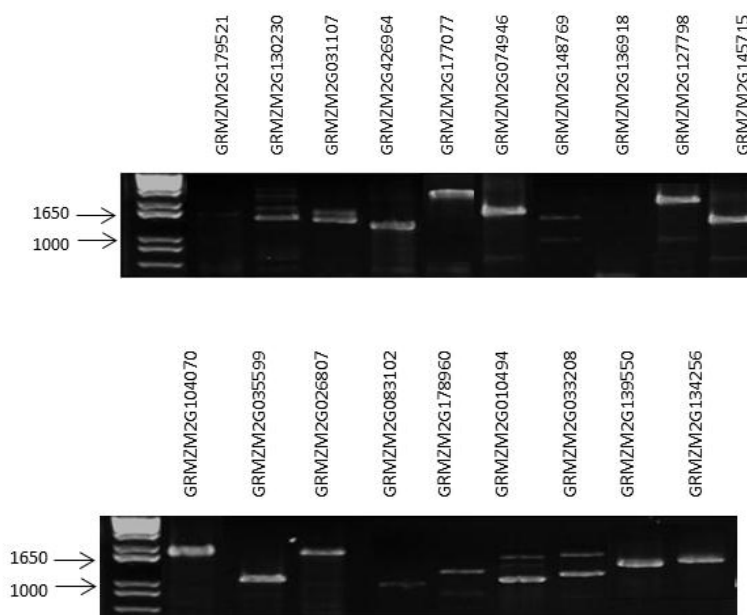
Enzima	Gene Model	Softwares													
		Target P	Wolf Psort	iPSORT	CeLLO	SUBLOC	SLP-Local	PredSL	Plant-mPLoc	SherLoc2	MultiLoc2	SubCellProt	SLPFA	Euk-mPLoc	
Glicose-6-fosfato desidrogenase	GRMZM2G179521	Cloroplasto	Cloroplasto	-	Cloroplasto	Citoplasma	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Citoplasma	Cloroplasto	-
Glicose-6-fosfato desidrogenase	GRMZM2G130230	-	Cloroplasto	-	Citoplasma	Citoplasma	Citoplasma	Cloroplasto	Ambos	Citoplasma	Citoplasma	Citoplasma	Cloroplasto	Cloroplasto	
Glicose-6-fosfato desidrogenase	GRMZM2G031107	Cloroplasto	Citoplasma	-	Citoplasma	Citoplasma	Cloroplasto	Cloroplasto	Ambos	Citoplasma	Cloroplasto	Ambos	Cloroplasto	Cloroplasto	
Glicose-6-fosfato desidrogenase	GRMZM2G177077	Ambos	Cloroplasto	-	Ambos	Citoplasma	Cloroplasto	Citoplasma	Cloroplasto	Ambos	Cloroplasto	-	Cloroplasto	Cloroplasto	
Glicose-6-fosfato desidrogenase	GRMZM2G426964	Cloroplasto	Cloroplasto	Cloroplasto	Ambos	Citoplasma	Cloroplasto	Citoplasma	Citoplasma	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Citoplasma
6-fosfogliconolactonase	GRMZM2G148769	Cloroplasto	Cloroplasto	Cloroplasto	Ambos	Citoplasma	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto
6-fosfogliconolactonase	GRMZM2G122126	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Citoplasma	Cloroplasto	Cloroplasto	-	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	-
6-fosfogliconolactonase	GRMZM2G074946	Cloroplasto	Cloroplasto	Cloroplasto	Ambos	Citoplasma	Cloroplasto	-	Citoplasma	Citoplasma	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto
6-fosfogliconolactonase	GRMZM2G136918	Ambos	Citoplasma	-	Citoplasma	Citoplasma	Citoplasma	-	Ambos	Cloroplasto	Citoplasma	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto
6-fosfogliconato desidrogenase	GRMZM2G127798	-	Citoplasma	-	-	Citoplasma	-	-	-	Cloroplasto	Citoplasma	-	Cloroplasto	Citoplasma	
6-fosfogliconato desidrogenase	GRMZM2G145715	-	Citoplasma	Cloroplasto	Cloroplasto	Citoplasma	-	-	-	Citoplasma	Cloroplasto	Citoplasma	Cloroplasto	Cloroplasto	
6-fosfogliconato desidrogenase	GRMZM2G440208	Cloroplasto	Cloroplasto	Cloroplasto	Ambos	Citoplasma	Cloroplasto	Cloroplasto	-	Citoplasma	Citoplasma	-	Cloroplasto	Cloroplasto	
Ribose-5-fosfato isomerase	GRMZM2G456086	Citoplasma	Cloroplasto	Cloroplasto	Cloroplasto	Citoplasma	Cloroplasto	Cloroplasto	Cloroplasto	Ambos	Ambos	Cloroplasto	Cloroplasto	Cloroplasto	Citoplasma
Ribose-5-fosfato isomerase	GRMZM2G035599	Cloroplasto	Citoplasma	-	Citoplasma	Citoplasma	Ambos	Cloroplasto	Cloroplasto	Citoplasma	Citoplasma	Cloroplasto	Cloroplasto	Cloroplasto	-
Ribose-5-fosfato isomerase	GRMZM5G891282	Citoplasma	Ambos	-	Citoplasma	Citoplasma	-	-	Cloroplasto	Citoplasma	Citoplasma	Cloroplasto	Citoplasma	Cloroplasto	
Ribose-5-fosfato isomerase	GRMZM2G104070	Ambos	Citoplasma	-	Ambos	Citoplasma	Cloroplasto	-	Cloroplasto	Citoplasma	-	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto
Ribulose-fosfato-3-epimerase	GRMZM2G178960	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Citoplasma	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto
Ribulose-fosfato-3-epimerase	GRMZM2G083102	Citoplasma	Cloroplasto	-	Citoplasma	Citoplasma	Citoplasma	Cloroplasto	Ambos	Cloroplasto	Cloroplasto	Cloroplasto	Citoplasma	Cloroplasto	
Ribulose-fosfato-3-epimerase	GRMZM2G026807	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Citoplasma	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto
Transquetolase	GRMZM2G010494	Citoplasma	Citoplasma	-	Citoplasma	Citoplasma	-	-	Cloroplasto	Citoplasma	Citoplasma	Cloroplasto	Cloroplasto	Cloroplasto	-
Transquetolase	GRMZM2G033208	Citoplasma	Cloroplasto	-	Ambos	Citoplasma	Citoplasma	-	Cloroplasto	Citoplasma	Citoplasma	Cloroplasto	Cloroplasto	Cloroplasto	-
Transadolase	GRMZM2G139550	Cloroplasto	Cloroplasto	Cloroplasto	Ambos	Citoplasma	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto
Transaldolase	GRMZM2G134256	Cloroplasto	Cloroplasto	-	Cloroplasto	Citoplasma	Cloroplasto	Ambos	Ambos	Citoplasma	Ambos	Cloroplasto	Cloroplasto	Cloroplasto	

**Tabela 9:** Predição da localização subcelular das enzimas da via das pentoses fosfato a partir do consenso dos preditores utilizados.

<b>Enzima</b>	<b>Gene Model</b>	<b>Localização Subcelular</b>
Glicose-6-fosfato desidrogenase	GRMZM2G179521	Cloroplasto
Glicose-6-fosfato desidrogenase	GRMZM2G130230	Cloroplasto e <b>Citoplasma</b>
Glicose-6-fosfato desidrogenase	GRMZM2G031107	Cloroplasto e <b>Citoplasma</b>
Glicose-6-fosfato desidrogenase	GRMZM2G177077	Cloroplasto
Glicose-6-fosfato desidrogenase	GRMZM2G426964	Cloroplasto
6-fosfogliconolactonase	GRMZM2G148769	Cloroplasto
6-fosfogliconolactonase	GRMZM2G122126	Cloroplasto
6-fosfogliconolactonase	GRMZM2G074946	Cloroplasto
6-fosfogliconolactonase	GRMZM2G136918	Cloroplasto e <b>Citoplasma</b>
6-fosfogliconato desidrogenase	GRMZM2G127798	Indeterminado
6-fosfogliconato desidrogenase	GRMZM2G145715	Cloroplasto e <b>Citoplasma</b>
6-fosfogliconato desidrogenase	GRMZM2G440208	Cloroplasto
Ribose-5-fosfato isomerase	GRMZM2G456086	Cloroplasto
Ribose-5-fosfato isomerase	GRMZM2G035599	Cloroplasto e <b>Citoplasma</b>
Ribose-5-fosfato isomerase	GRMZM5G891282	Cloroplasto e <b>Citoplasma</b>
Ribose-5-fosfato isomerase	GRMZM2G104070	Cloroplasto
Ribulose-fosfato-3-epimerase	GRMZM2G178960	Cloroplasto
Ribulose-fosfato-3-epimerase	GRMZM2G083102	Cloroplasto e <b>Citoplasma</b>
Ribulose-fosfato-3-epimerase	GRMZM2G026807	Cloroplasto
Transquetolase	GRMZM2G010494	<b>Citoplasma</b>
Transquetolase	GRMZM2G033208	Cloroplasto e <b>Citoplasma</b>
Transadolase	GRMZM2G139550	Cloroplasto
Transaldolase	GRMZM2G134256	Cloroplasto

Em relação aos ensaios experimentais para determinar a localização subcelular das enzimas da via das pentoses fosfato, foram amplificados 19 dos 22 genes selecionados visando à obtenção da sequência *full length* do gene (Figura 8). Não foi possível amplificar nenhuma sequência a partir dos *primers* desenhados para o gene GRMZM2G122126 que codifica a enzima 6-fosfogluconolactonase, bem como os dois genes que codificam a enzima transcetolase, GRMZM2G010494, GRMZM2G033208.

O gene GRMZM2G440208, codificador da enzima 6-fosfogluconato desidrogenase fez parte das análises de comparação, entretanto, foi caracterizado anteriormente e possui localização subcelular cloroplastídica definida (SPIELBAUER et al., 2006).



**Figura 8:** Visualização em gel de agarose 0,8% da reação em cadeia da polimerase dos *full lengths* clonados com sucesso no vetor *p2FGW7*.

Apesar da confecção de três pares de *primers* para os genes GRMZM2G122126, GRMZM2G010494, GRMZM2G033208, a partir de suas anotações no banco de dados da linhagem B73, não foi possível amplificar nenhuma sequência. A não amplificação pode ter ocorrido devido a anotação de entrada desses genes conterem falhas ou erros nas sequências, sendo necessários, estudos posteriores para a correta identificação dos mesmos.

Os CDSs dos genes de interesse foram obtidos a partir da amplificação do DNA dos vetores bacterianos purificados, utilizando-se os oligonucleotídeos para a amplificação da ORF. Dessa forma, novas construções no vetor de entrada pENTR / D-TOPO, sem o códon de terminação, puderam ser obtidas. Procedeu-se assim, novas avaliações para confirmação das sequências, a partir do uso dos *primers* internos e novo sequenciamento.

Os dados de sequenciamento dos CDSs para a linhagem de milho utilizada nesse trabalho (W22), indicou que 10 dos 19 genes sequenciados, possuem polimorfismo nucleotídico em relação às sequências disponíveis nos bancos de dados do *NCBI* e *MaizeSequence* para a linhagem B73. Dentre esses polimorfismos, foram identificadas

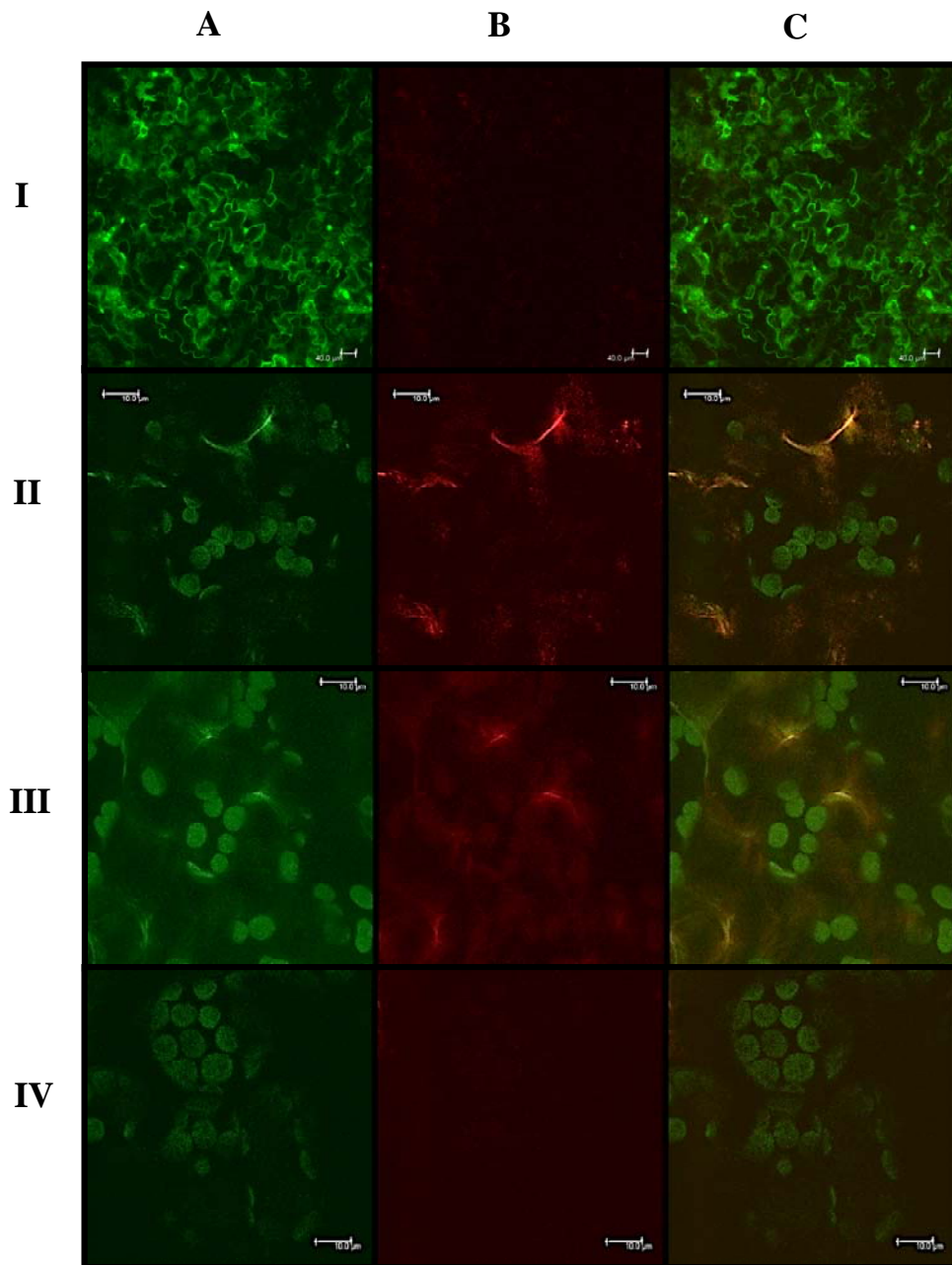
substituições e transversões de bases. Destacam-se os genes GRMZM2G104070 e GRMZM2G136918, que apresentaram sete substituições nucleotídicas (Apêndice 2).

Os CDSs obtidos foram transferidos para o vetor binário *gateway* (pB7FWG2) que fusiona o GFP na região C-terminal da proteína, por meio de reação com a Gateway® LR Clonase™ II Enzyme Mix. Os clones positivos foram transformados em *Agrobacterium* e os ensaios de expressão transiente foram conduzidos em tabaco, por meio da visualização da fluorescência do gene de interesse fusionado ao *gfp* para a identificação da localização. Os resultados destes ensaios são apresentados nas Figuras 9 a 11.

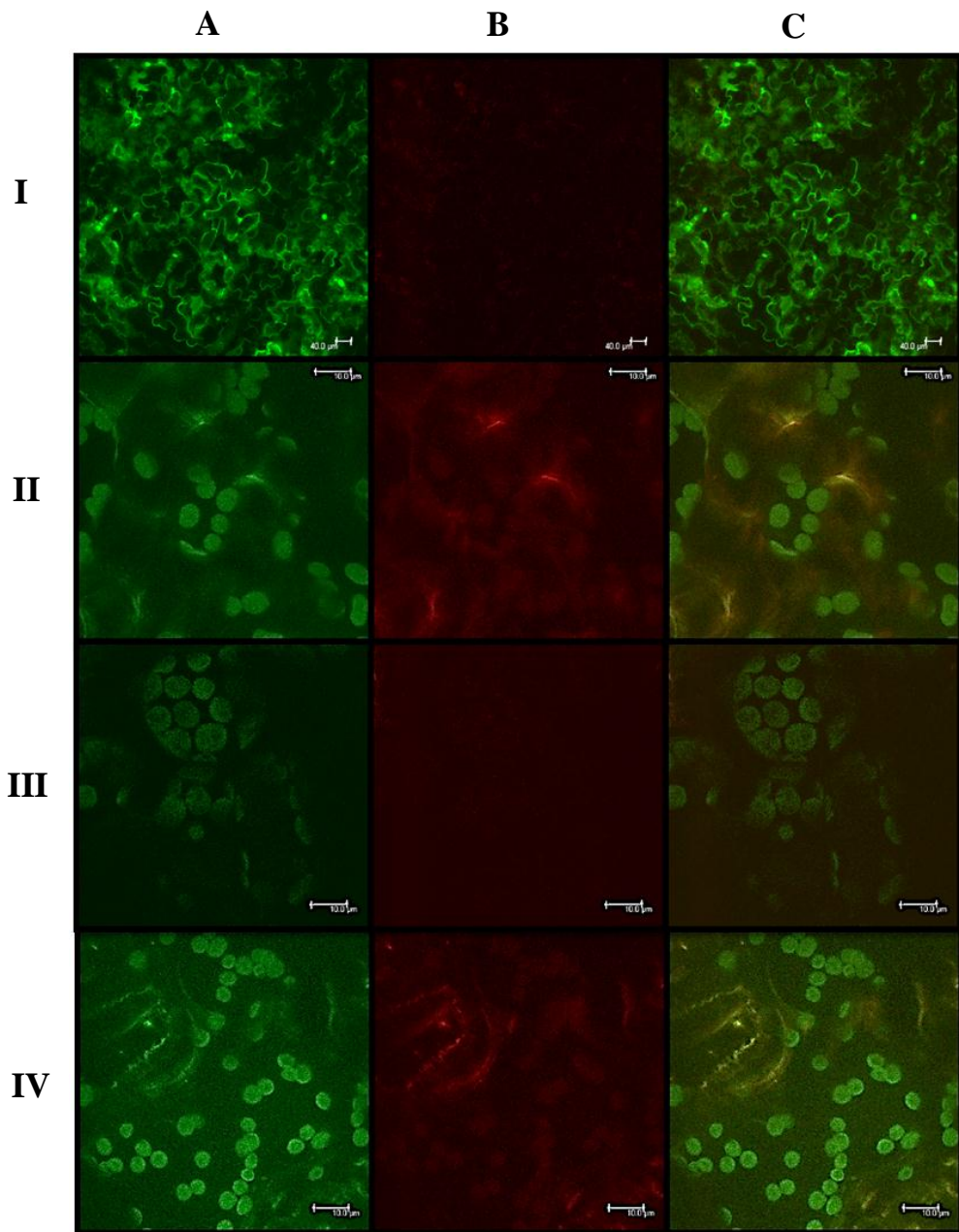
Seis das sete enzimas da via das pentoses fosfato tiveram a localização celular determinadas. Devido à impossibilidade de não amplificação dos dois genes que codificam a enzima transcetolase, o processo de localização subcelular para essa enzima não pode ser finalizado.

Ao todo, dez enzimas foram caracterizadas, nove neste trabalho e uma em estudo anterior (SPIELBAUER et al., 2006). Todas as enzimas avaliadas estavam localizadas no cloroplasto. Foi obtida a localização subcelular de um gene que codifica a enzima Glicose-6-fosfato desidrogenase (Figura 9-II); dois genes que codificam a 6-fosfogluconolactonase (Figura 9-IV e 10-II); dois genes da 6-fosfogluconato desidrogenase (Figura 10-III e 10-IV); um gene da Ribose 5-fosfato isomerase (Figura 11-II); dois genes para a ribulose-fosfato 3 epimerase (Figura 9-III e Figura 11-III) e, finalmente, um gene que codifica a enzima transaldolase (Figura 11-IV).

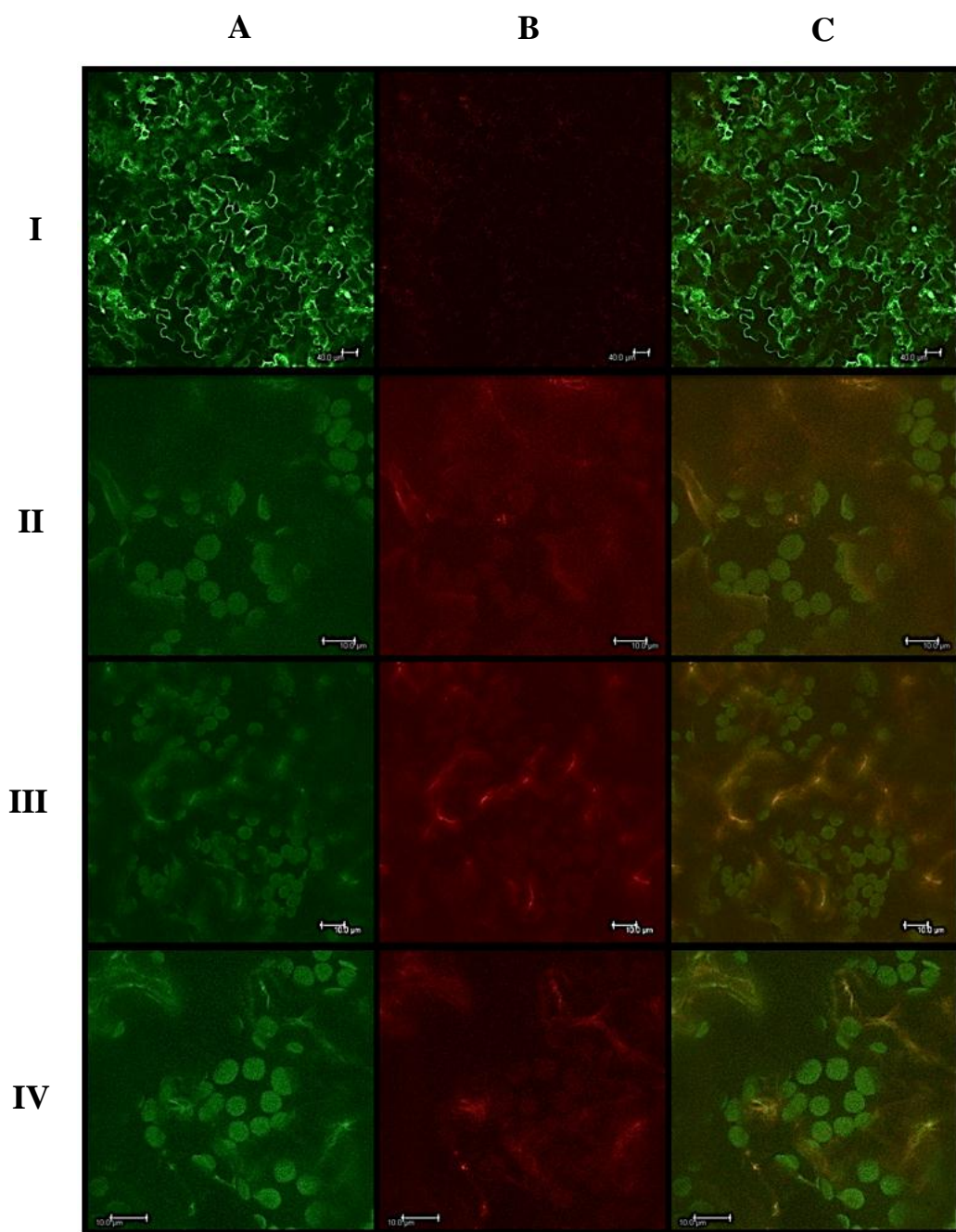
A via das pentoses fosfato inclui uma série de reações catalisadas por sete enzimas diferentes e, até o momento, a distribuição subcelular destas enzimas, em células vegetais, não foi completamente decifrada (KRUGER; SCHAEWEN 2003). Acredita-se que as enzimas que catalisam as reações oxidativas do PPP existam nos cloroplastos e citoplasma, enquanto que aquelas que catalisam a fase não-oxidativa são restritas aos cloroplastos (SCHNARRENBARGER et al. 1995; DEBNAM; EMES, 1999; KRUGER; SCHAEWEN, 2003).



**Figura 9:** Ensaios de expressão transiente para avaliação da localização subcelular das enzimas da via das pentoses fosfato em milho. Células de tabaco expressando as construções gênicas fusionadas com GFP, sob incidência de fluorescência confocal (coluna A), sinal auto-fluorescência sobre o mesmo campo (coluna B) e a imagem resultante da junção das duas anteriores (coluna C). Aumento de 20X. Escalas de tamanho de 10 $\mu$ m e 40  $\mu$ m. **I:** Controle; **II:** GRMZM2G031107; **III:** GRMZM2G083102; **IV:** GRMZM2G148769.



**Figura 10:** Ensaios de expressão transitente dos controles para avaliação da localização subcelular das enzimas da via das pentoses fosfato em milho. Células de tabaco expressando as construções gênicas fusionadas com GFP, sob incidência de fluorescência confocal (coluna A), sinal autofluorescência sobre o mesmo campo (coluna B) e a imagem resultante da junção das duas anteriores (coluna C). Aumento de 20X. Escalas de tamanho de 10 $\mu$ m e 40  $\mu$ m. **I:** Controle; **II:** GRMZM2G136918; **III:** GRMZM2G127798; **IV:** GRMZM2G145715.



**Figura 11:** Ensaios de expressão transiente dos controles para avaliação da localização subcelular das enzimas da via das pentoses fosfato em milho. Células de tabaco expressando as construções gênicas fusionada com GFP, sob incidência de fluorescência confocal (coluna A), sinal autofluorescência sobre o mesmo campo (coluna B) e a imagem resultante da junção das duas anteriores (coluna C). Aumento de 20X. Escalas de tamanho de 10 $\mu$ m e 40  $\mu$ m. **I:** Controle; **II:** GRMZM2G035599; **III:** GRMZM2G178960; **IV:** GRMZM2G139550.

Anderson (1971), encontrou formas citosólicas e plastídicas de ribose 5-fosfato-isomerase em folhas de ervilha, embora a concentração cloroplastídica era muito maior. Atividades da transaldolase e transcetolase foram demonstradas no citosol, em embriões de colza (NISHIMURA; BEEVERS, 1979).

Utilizando protoplastos de sementes de mamona (*Ricinus communis* L.), Simcox e colaboradores (1977) verificaram que todas as enzimas da via das pentoses fosfato se expressavam em ambos os compartimentos, entretanto, a predominância foi o dobro no cloroplasto. Em algas da espécie *Chlamydomonas reinhardtii*, Klein (1986) verificou que mais de 70% de todas as enzimas da via oxidativa das pentoses fosfato estão localizadas no cloroplasto.

Hong e Copeland (1990) demonstraram que duas enzimas da fase não oxidativa da via das fosfopentoses, ribulose-fosfato 3-epimerase e ribose 5-fosfato isomerase, estavam presentes tanto no citoplasma, quanto no cloroplasto, em nódulos de raízes de soja. No entanto, um estudo em espinafre sugeriu que as enzimas da via não-oxidativa das pentoses fosfato estão restritas ao cloroplasto (SCHNARRENBARGER et al., 1995) e que apenas as duas primeiras enzimas da via oxidativa (G6PDH e 6PGDH) são encontradas em ambos os locais. Em outro estudo, Phillip e Michael (1999) não encontraram significativa expressão citoplasmática de transcetolase, transaldolase, ribulose-fosfato 3-epimerase e ribose 5-fosfato isomerase em tecidos de raízes e folhas de ervilha.

Neste estudo não foi a localização subcelular de todas as enzimas da via das pentoses fosfato, entretanto, diante de alguns trabalhos, podemos supor que, a distribuição das enzimas, principalmente as da fase não-oxidativa, pode variar devido ao tecido específico analisado, das fases de desenvolvimento do indivíduo, da espécie estudada e das condições ambientais. Um trabalho realizado em milho, desenvolvido por Debnam e Emes (1999), demonstrou que em raízes de milho, as enzimas da via oxidativa das pentoses fosfato, foram encontradas em ambos os compartimentos (citoplasma e cloroplasto), entretanto em folhas, as enzimas foram encontradas somente em cloroplastos.

Comparando os resultados dos dados experimentais em relação aos preditos, pode-se perceber algumas diferenças (Tabela 10). Quatro enzimas, codificadas pelos genes GRMZM2G148769, GRMZM2G440208 GRMZM2G178960, GRMZM2G139550 e analisadas pelos *softwares* preditores, concordaram com os dados experimentais; cinco

foram parcialmente semelhantes, pois a previsão foi para ambos os locais (GRMZM2G031107; GRMZM2G136918; GRMZM2G145715; GRMZM2G035599 e GRMZM2G083102); um gene não teve previsão conclusiva (GRMZM2G127798).

Devido à utilização de apenas de folhas de plântulas de milho para esse estudo, inferimos que, as localizações celulares para ambos os compartimentos poderá ser incluído como predição positiva.

**Tabela 10:** Comparação dos resultados da localização subcelular provenientes da análise experimental e por preditores

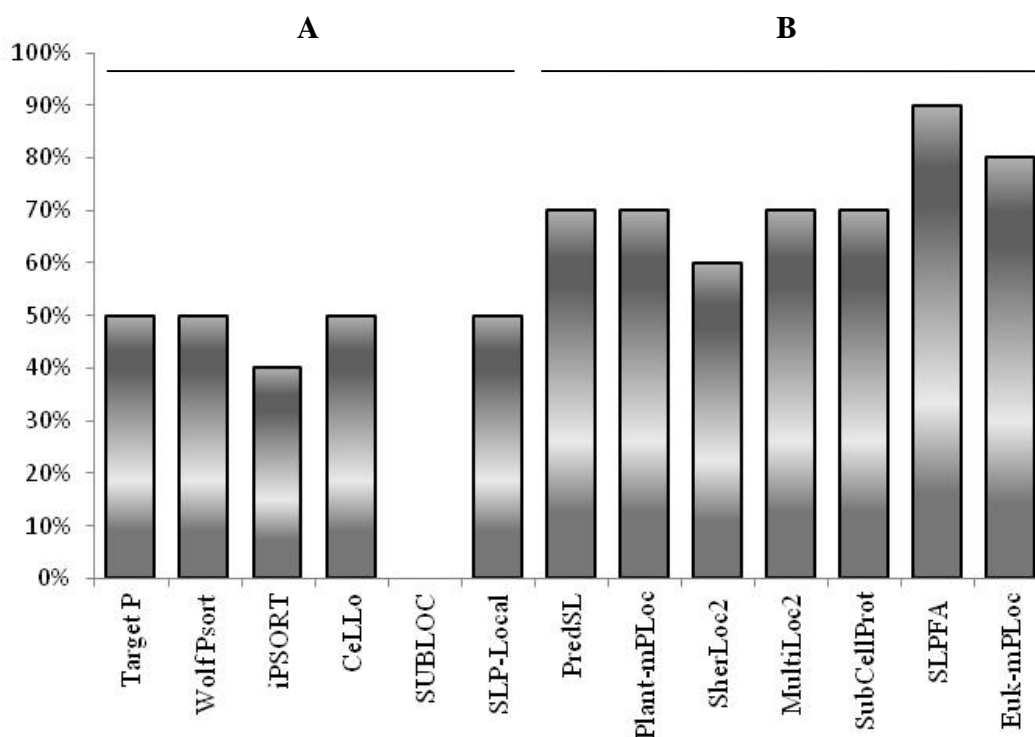
Enzima	Gene Model	Preditores	Experimental
Glicose-6-fosfato desidrogenase	GRMZM2G031107	Cloroplasto e Citoplasma	Cloroplasto
6-fosfogliconolactonase	GRMZM2G148769	Cloroplasto	Cloroplasto
6-fosfogliconolactonase	GRMZM2G136918	Cloroplasto e Citoplasma	Cloroplasto
6-fosfogliconato desidrogenase	GRMZM2G127798	Indeterminado	Cloroplasto
6-fosfogliconato desidrogenase	GRMZM2G145715	Cloroplasto e Citoplasma	Cloroplasto
6-fosfogliconato desidrogenase	GRMZM2G440208	Cloroplasto	Cloroplasto
Ribose-5-fosfato isomerase	GRMZM2G035599	Cloroplasto e Citoplasma	Cloroplasto
Ribulose-fosfato-3-epimerase	GRMZM2G178960	Cloroplasto	Cloroplasto
Ribulose-fosfato-3-epimerase	GRMZM2G083102	Cloroplasto e Citoplasma	Cloroplasto
Transadolase	GRMZM2G139550	Cloroplasto	Cloroplasto

Analisando individualmente os resultados dos preditores, podemos perceber variações nas acurácias em relação à localização subcelular experimental. Foi realizada a soma dos resultados dos preditores em: exclusivamente cloroplasto ou cloroplasto, citoplasma e sem predição definida. Dos 13 preditores analisados, apenas o SUBLOC (HU; SUN, 2001) foi totalmente discordante em relação às localizações experimentalmente definidas (Tabela 11). Hu e colaboradores (2009) relataram baixa resolução para esse *software*, com baixa porcentagem de previsão positiva, além disso, apresentou maior discordância em relação aos dados experimentais em relação a outros preditores analisados para genes humanos.

**Tabela 11:** Comparação dos resultados da localização subcelular provenientes da análise experimental e por preditores. A acurácia média de cada preditor, segundo os desenvolvedores dos *softwares* está representada em porcentagem.

<i>Gene Model</i>	Experimental	<i>Softwares</i>												
		Target P	Wolf Psort	iPSORT	CeLLO	SUBLOC	SLP-Local	PredSL	Plant-mPLoc	SherLoc2	MultiLoc2	SubCellProt	SLPFA	Euk-mPLoc
		85%	80%	83%	80-90%	79%	86%	87%	86%	93%	78-89%	85%	90.96%	90%
GRMZM2G031107	Cloroplasto	Cloroplasto	Citoplasma	-	Citoplasma	Citoplasma	Cloroplasto	Cloroplasto	Ambos	Citoplasma	Cloroplasto	Ambos	Cloroplasto	Cloroplasto
GRMZM2G148769	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Ambos	Citoplasma	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto
GRMZM2G136918	Cloroplasto	Ambos	Citoplasma	-	Citoplasma	Citoplasma	Citoplasma	-	Ambos	Cloroplasto	Citoplasma	Cloroplasto	Cloroplasto	Cloroplasto
GRMZM2G127798	Cloroplasto	-	Citoplasma	-	-	Citoplasma	-	-	-	Cloroplasto	Citoplasma	-	Cloroplasto	Citoplasma
GRMZM2G145715	Cloroplasto	-	Cloroplasto	-	Cloroplasto	Citoplasma	-	-	-	Citoplasma	Cloroplasto	Citoplasma	Cloroplasto	Cloroplasto
GRMZM2G440208	Cloroplasto	-	Citoplasma	Cloroplasto	Cloroplasto	Citoplasma	-	Cloroplasto	-	Citoplasma	Cloroplasto	Citoplasma	Cloroplasto	Cloroplasto
GRMZM2G035599	Cloroplasto	Cloroplasto	Citoplasma	-	Citoplasma	Citoplasma	Ambos	Cloroplasto	Cloroplasto	Citoplasma	Citoplasma	Cloroplasto	Cloroplasto	-
GRMZM2G178960	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Citoplasma	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto
GRMZM2G083102	Cloroplasto	Citoplasma	Cloroplasto	-	Citoplasma	Citoplasma	Citoplasma	Cloroplasto	Ambos	Cloroplasto	Cloroplasto	Cloroplasto	Citoplasma	Cloroplasto
GRMZM2G139550	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Citoplasma	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto	Cloroplasto
<b>Número</b>	Cloroplasto	5	5	4	4	0	4	7	4	6	7	6	9	8
<b>de</b>	Ambos	0	0	0	1	0	1	0	3	0	0	1	0	0
<b>Localizações</b>	-	2	0	6	1	0	2	3	2	1	1	3	0	1

O preditor mais acurado foi o SLPFA com nove coincidências das 10 avaliações (90% de acerto). Em seguida, o Euk-mPLoc com 8 (80%); PredSL, Plant-mPLoc, MultiLoc2 e SubCellProt com eficiência de 70%; SherLoc2, 60%; Target P, Wolf Psort, CeLLO e SLP-Local com 50% e, por último, o iPSORT com 4 padrões semelhantes ao encontrado na localização subcelular experimental (40%). Estas variações de desempenho podem ser provenientes dos diferentes métodos de previsão que cada *software* utiliza (Figura 12).



**Figura 12:** Comparação do desempenho da acurácia computacional de 13 preditores, por meio da utilização das dez sequências protéicas das enzimas da via das pentoses fosfato, em milho. As porcentagens denotam o número de coincidências dos preditores, em relação aos dados experimentais. **A:** Preditores que utilizam apenas similaridade de sequências; **B:** Preditores com múltiplas plataformas de análise.

Além de apresentar a maior acurácia neste trabalho, o *software* SLPFA apresentou eficiência similar a encontrada nos testes desenvolvidos para validar o método (90%) (TAMURA; AKUTSU, 2007). Entretanto, para a maioria dos testes aqui avaliados, a

acurácia identificada foi significativamente menor das observadas nos manuais dos diferentes métodos (Tabela 11). Entretanto, vale ressaltar novamente que, neste estudo utilizamos apenas um tipo de tecido e um estágio de maturação. Análises com diferentes perfis podem apresentar resultados distintos em relação aos observados aqui.

Hu e colaboradores (2009), trabalhando com genes humanos, encontraram acurácia diferente das encontradas neste trabalho. Eles verificaram porcentagens de acerto de 83% para SherLoc2 e 75% para WoLF-PSORT (75%), contra 60% e 50%, respectivamente, obtidos aqui. Em relação ao preditor MultiLoc2, os pesquisadores encontraram uma acurácia semelhante (aproximadamente 70%).

Os preditores com menores acurácias (Target P, Wolf Psort, iPSORT, CeLLO, SUBLOC e SLP-Local) utilizam basicamente a análise de similaridade de peptídeos sinais e sequência de aminoácidos (HOEGLUND, 2006). O principal problema na predição baseada apenas em similaridade de sequências é que os resultados podem ser ambíguos se proteínas homólogas caracterizadas não estiverem disponíveis no banco de dados do *software* (HU et al., 2009).

A resolução básica do preditor SLPFA é de 90,96% e, segundo os autores, pode ser até 40% maior que TargetP. Neste trabalho, as eficiências foram muito semelhantes como relatado pelo método. A eficiência do primeiro foi de 90% enquanto que o segundo foi e 50%, o que representa 45% de diferença na acurácia dos dois testes.

Com exceção do SUBLOC, as acurácias dos preditores variaram de 40% a 90% de coincidência da localização subcelular, em relação à caracterização experimental. Apesar do número relativamente pequeno de proteínas analisadas neste estudo (10), os nossos resultados apresentaram acurácias inferiores as indicada pelos preditores, por exemplo, Wolf Psort foi reivindicado com uma precisão de mais de 80% (HORTON et al., 2006) e MultiLoc 2 com acurácia de 75% (HOEGLUND, 2006). Hu e colaboradores (2009), comparando a localização subcelular predita por *softwares* e a caracterização experimental, demonstraram que a acurácia também foi menor que os preditores indicavam (54–83%).

Analisando individualmente a predição dada para cada um dos genes que codificam as enzimas da via em questão, diferenças significativas foram verificadas. Dos treze preditores avaliados, foram identificadas discrepâncias que variaram de 7% (genes

GRMZM2G178960 e GRMZM2G139550) até 84% como encontrado no GRMZM2G127798 (Tabela 11).

O gene que codifica a enzima 6-fosfogliconato desidrogenase (GRMZM2G127798) não apresentou localização definida. Para esse gene 53% dos preditores não conseguiram definir a localização subcelular dessa proteína (Tabela 11).

Comparando-se os resultados dos preditores que utilizam apenas similaridade de sequências (A) e os preditores com multiplataformas (B), observou-se que o segundo grupo possui as maiores acurácias (>72%), enquanto que o primeiro grupo, excluindo o preditor SubLoc, apresentou média de 48% de acurácia.

Os resultados demonstraram que o desempenho da predição varia consideravelmente entre os diversos *softwares*. Tomando-se como base os preditores mais eficientes aqui relatados, percebeu-se que a maioria utiliza outros suportes que não somente a composição de aminoácidos ou peptídeos sinais, enquanto que, os *softwares* com menor eficácia baseiam-se apenas na composição de aminoácidos. Isso indica que, a combinação de sequências, filogenia e métodos estatísticos podem aumentar a acurácia da predição da localização subcelular das enzimas.

Em alguns casos, as previsões podem estar incorretas, mesmo que a maioria dos preditores reporte as mesmas localizações. Neste estudo, boa parte das previsões estavam em desacordo com as localizações subcelulares experimentalmente analisadas, haja vista que, todos os preditores reportam níveis de eficiência maiores que 80% e, neste trabalho, apenas dois preditores apresentaram acurácia dentro dessa margem.

Com base nas diferenças entre as acurácias dos preditores e os dados experimentais obtidos neste trabalho, sugere-se cautela na utilização desses *softwares*, buscando sempre fazer comparações entre diferentes plataformas, bem como, dar prioridade a aqueles que fornecem variados algoritmos para a predição. Além disso, a localização subcelular deverá abranger mais dados experimentais e ser realizada em diferentes tecidos e estágios de maturação, para a identificação do local de expressão da enzima e suas interações com o meio.

### **3.2.4. CONCLUSÕES**

Os resultados obtidos permitem as seguintes conclusões:

- Todas as enzimas caracterizadas experimentalmente foram expressas somente no cloroplasto.
- Diferentes tecidos e estágios de maturação da planta deve ser analisado em experientos de localização subcelular de proteínas.
- Preditores multiplataformas são mais recomendados para previsões de localização subcelular.

### 5.2.5. REFERÊNCIAS BIBLIOGRÁFICAS

ANDERSON, L.E. Chloroplast and cytoplasmic enzymes. **Cell Physiology**, v. 17, p. 757–764. 1971.

BANNAI, H., TAMADA, Y., MARUYAMA, O., NAKAI, K., AND MIYANO, S. Extensive feature detection of N-terminal protein sorting signals", **Bioinformatics**, v. 18(2), p.298-305, 2002.

BERG I.A. Ecological aspects of the distribution of different autotrophic CO<sub>2</sub> fixation pathways. **Applied Environmental Microbiology**, v. 77, p. 1925–1936. 2011.

BLUM, T.; BRIESEMEISTER, S.; KOHLBACHER, O. MultiLoc2: integrating phylogeny and Gene Ontology terms improves subcellular protein localization prediction. **BMC Bioinformatics**, v. 10, p. 274. 2009.

BRIESEMEISTER, S.; BLUM, T.; BRADY, S.; LAM, Y.; KOHLBACHER, O.; SHATKAY, H. SherLoc2: a high-accuracy hybrid method for predicting subcellular localization of proteins. *J. Proteome Res.*, v. 8(11), p. 5363–5366. 2009.

CHELKH, N.; JONES, R.J. Heat stress effects on sink activity of developing maize kernels grown in vitro. **Physiol. Plant.** v. 95, p. 59-66. 1995.

CHEN, H.; HUANG, N.; SUN, Z. SubLoc: a server/client suite for protein subcellular location based on SOAP. **Bioinformatics**, v. 1;22(3), p. 376-377. 2006.

CHOU, K.C.; SHEN, H. B. A New Method for Predicting the Subcellular Localization of Eukaryotic Proteins with Both Single and Multiple Sites: Euk-mPLoc 2.0, **PLoS ONE**, v. 5. 2010.

CHOU, K.C.; SHEN, H. B. Plant-mPLoc: a top-down strategy to augment the power for predicting plant protein subcellular localization, **PLoS ONE**, v. 5. 2010.

CONAB. **Acompanhamento de safra brasileira: grãos, décimo primeiro levantamento.** Agosto de 2011.

EMANUELSSON, O.; BRUNAK, S.; HEIJNE, G. V.; NIELSEN, H. Locating proteins in the cell using TargetP, SignalP, and related tools. **Nature Protocols**, v. 2, p. 953-971. 2007.

GARG, P.; SHARMA, V.; CHAUDHARI, P.; ROY, N. SubCellProt: predicting protein subcellular localization using machine learning approaches. **In Silico Biol.** v. 9(1-2), p. 35-44. 2009.

HEAZLEWOOD, J.L., TONTI-FILIPPINI, J.S., GOUT, A.M., DAY, D.A., WHELAN, J.; MILLAR, A. Experimental Analysis of the Arabidopsis Mitochondrial Proteome Highlights Signaling and Regulatory Components, Provides Assessment of Targeting Prediction Programs, and Indicates Plant-Specific Mitochondrial Proteins. **Plant Cell**, v. 16 (1), p. 241–256. 2004.

HONG, Z.Q.; COPELAND, L. Pentose phosphate pathway enzymes in nitrogen-fixing leguminous root nodules. **Phytochemistry**, v.29, p. 2437–2440. 1990.

HORTON, P.; PARK, K-J.; OBAYASHI, T.; FUJITA, N.; HARADA, H.C.J. COLLIER, A.; NAKAI, K. WoLF PSORT: Protein Localization Predictor. **Nucleic Acids Research**, 2006.

HU, Y.; LEHRACH, H.; JANITZ, M. Comparative analysis of an experimental subcellular protein localization assay and in silico prediction methods. **J Mol Hist**, v. 40. P. 343–352. 2009.

KARIMI, M.; INZE, D.; DEPICKER, A. GATEWAY™ vectors for Agrobacterium-mediated plant transformation. **Trends Plant Sci**, v. 7, p. 193–195. 2002.

LIEW, A.; YAN, W.C. H.; Yang, M. Pattern recognition techniques for the emerging field of bioinformatics: A review. **Pattern Recognition**, v. 38(11), p. 2055–2073, 2005.

MATHIS, N.L.; HINCHEE, M.A.W. Agrobacterium inoculation technics for plant tissues. In: GELVIN, S.B.; SCHILPEROORT, R.A. (Eds.). *Plant molecular biology manual*. 2.ed. Dordrecht : Kluwer Academic, 1994.

MATSUDA, S., VERT, J.-P., SAIGO, H., UEDA, N., TOH, H., AND AKUTSU, T. A novel representation of protein sequences for prediction of subcellular location using support vector machines. **Protein Science**, v. 14, p. 2804-2813. 2005.

NISHIMURA, M.; BEEVERS, H. Subcellular distribution of gluconeogenic enzymes in germinating castor bean endo- sperm. **Plant Physiology**, v.64, p. 31–37. 1979.

PETSALAKI, E.I., BAGOS, P.G., LITOU, Z.I., HAMODRAKAS, S.J. PredSL: A Tool for the N-terminal Sequence-based Prediction of Protein Subcellular Localization. **Geno. Prot. Bioinfo**. V. 4(1), P. 48-55. 2006.

PHILLIP, M. D.; MICHAEL J.E. Subcellular distribution of enzymes of the oxidative pentose phosphate pathway in root and leaf tissues. **Journal of Experimental Botany**, v. 50,

PRYER, K.M.; SCHNEIDER, J.; ZIMMERJ, E.A.; BANKS, A. Deciding among green plants for whole genome studies. **Trends in Plant Science**, v.7, p. 550-554, 2002.

PUJOL, C., MARÉCHAL-DROUARD, L. DUCHÊNE, A. How can organellar protein Nterminal sequences be dual targeting signals? In silico analysis and mutagenesis approach. **Journal of Molecular Biology**, v. 369, p. 356–367. 2007.

SAMBROOK, J.; FRITSCHI, E.F.; MANIATIS, T. **Molecular cloning: a laboratory manual**. Cold Spring Harbor Laboratory Press, New York. 1989.

SANGER, F.; COULSON, A.R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. **J Mol Biol.** v. 94, p. 441–448. 1974.

SHEN, Y.Q.; BURGER, G. Unite and conquer: enhanced prediction of protein subcellular localization by integrating multiple specialized tools. **BMC Bioinformatics**, v. 8, p. 420-430. 2007.

SIMPSON, J., WELLENREUTHER, R., POUSTKA, A., PEPPERKOK, R. WIEMANN, S. Systematic subcellular localization of novel proteins identified by large-scale cDNA sequencing. **EMBO Reports**, v. 1 (3), p. 287–292. 2000.

SPIELBAUER, G., MARGL, L., HANNAH, L.C., ROMISCH, W., ETTENHUBER, C., BACHER, A., GIERL, A., EISENREICH, W.; GENSCHEL, U. Robustness of central carbohydrate metabolism in developing maize kernels. **Phytochemistry** v. 67, p. 1460-1475. 2006.

TAMURA, T; AKUTSU, T. Subcellular location prediction of proteins using support vector machines with alignment of block sequences utilizing amino acid composition, **BMC Bioinformatics**, v. 8, p. 466, 2007.

XIE, D., LI, A., WANG, M., FAN, Z.; FENG, H. LOCSVMPSI: a web server for subcellular localization of eukaryotic proteins using SVM and profile of PSI-BLAST. *Nucleic Acids Research*, 33, v. 105, p. 110. 2005.

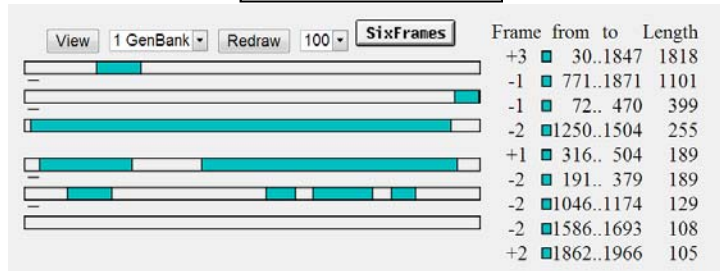
YU, C.S.; CHEN, Y.C.; LU, C.H.; HWANG, J.K. Prediction of protein subcellular localization. **Proteins: Structure, Function and Bioinformatics**, v. 64, p. 643-651. 2006.

## **5. APÊNDICE**

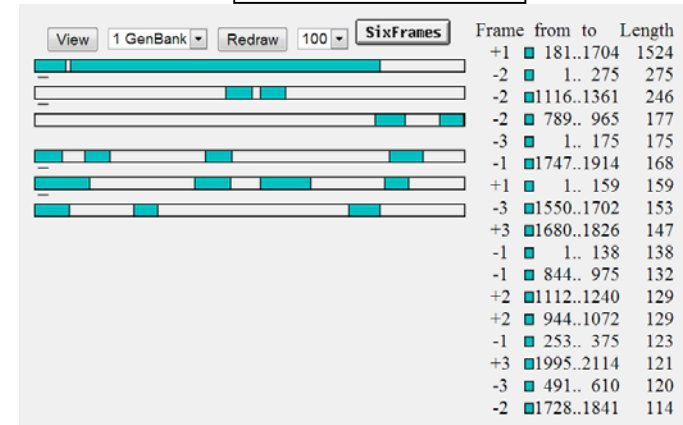
## Resultados do Software ORF FINDER

### Glicose-6-fosfato desidrogenase

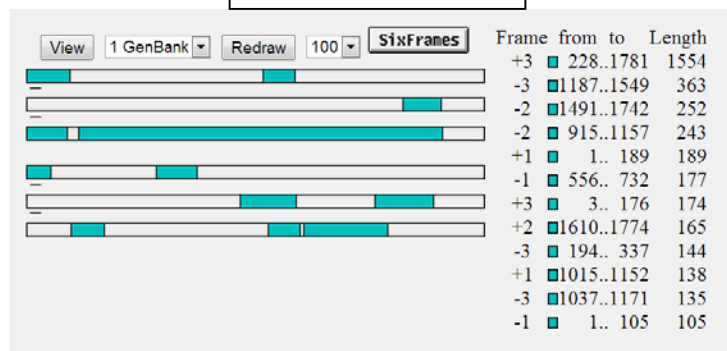
GRMZM2G179521



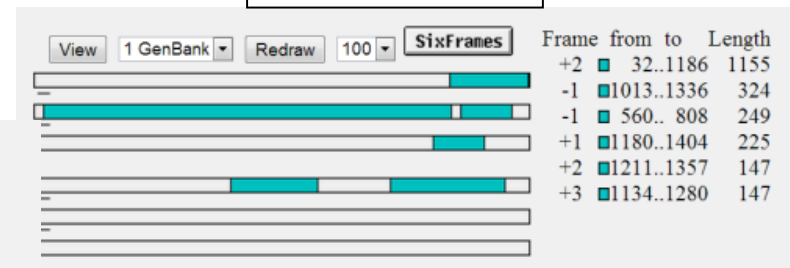
GRMZM2G130230



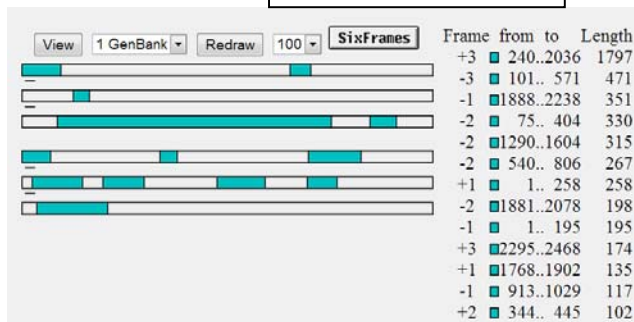
GRMZM2G031107



GRMZM2G426964

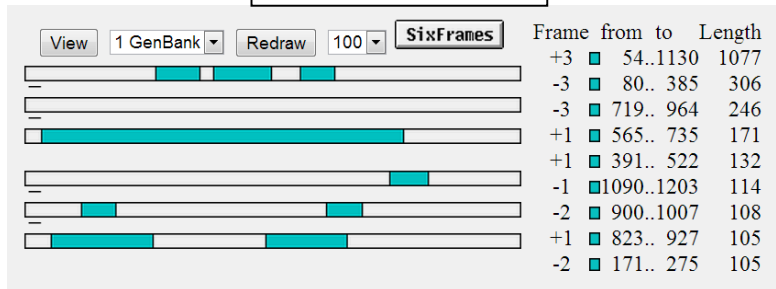


GRMZM2G177077

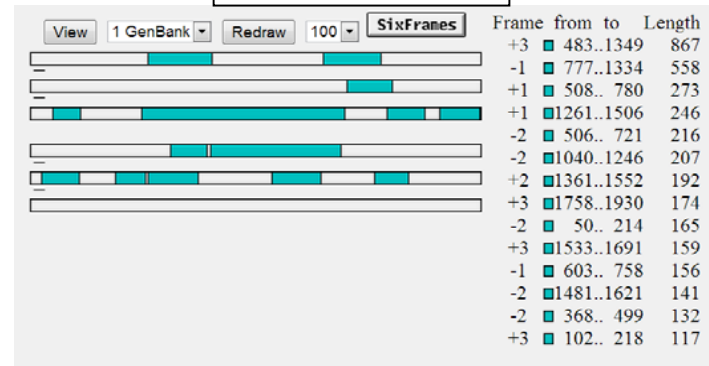


## 6-phosphogluconolactonase

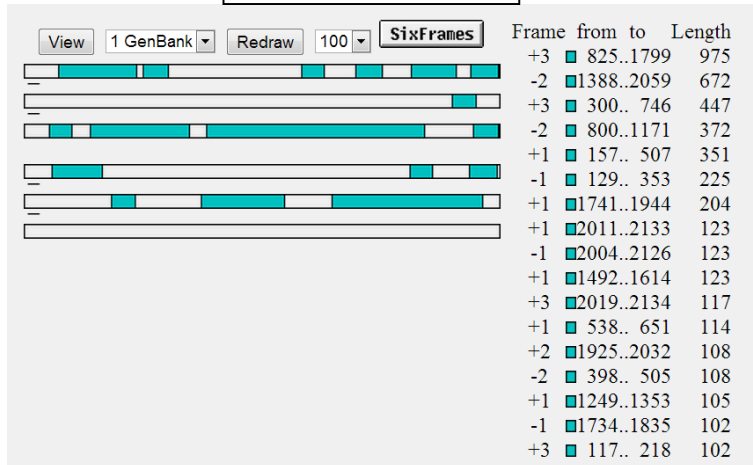
GRMZM2G122126



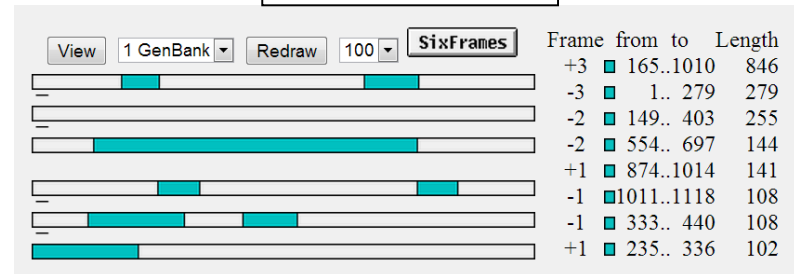
GRMZM2G074946



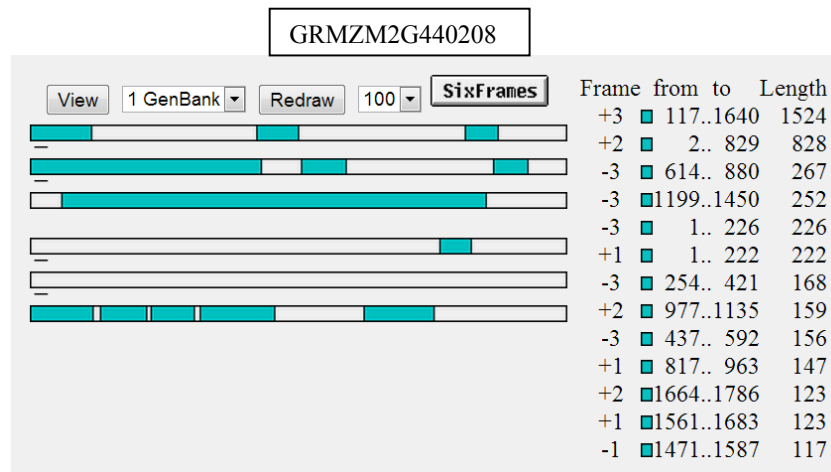
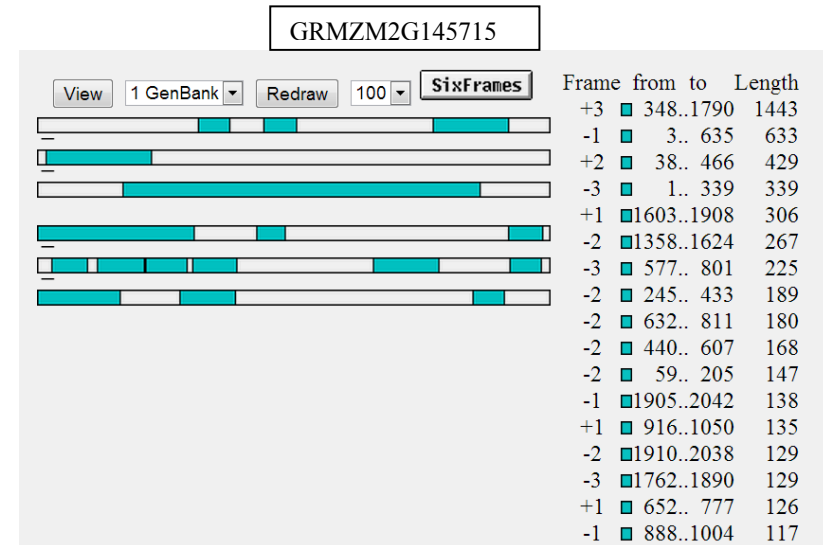
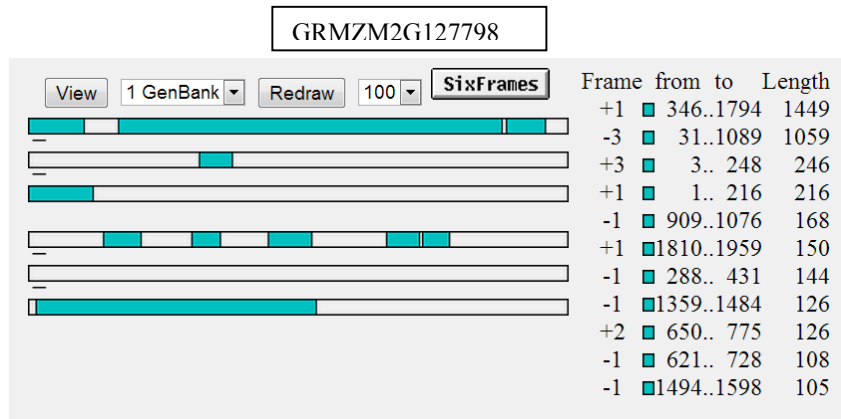
GRMZM2G148769



GRMZM2G136918

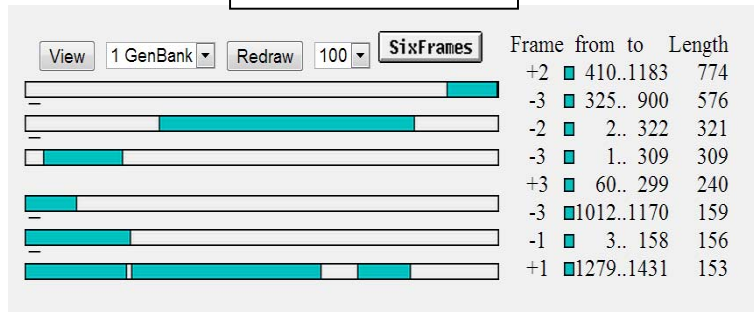


## 6-fosfoluconato desidrogenase

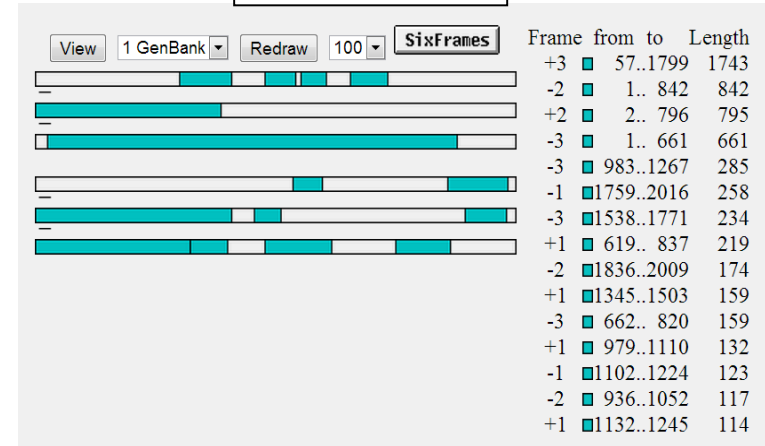


## Ribose 5-phosphate isomerase

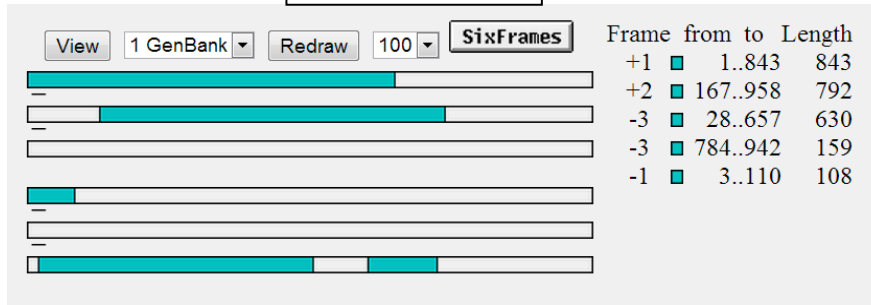
GRMZM2G104070



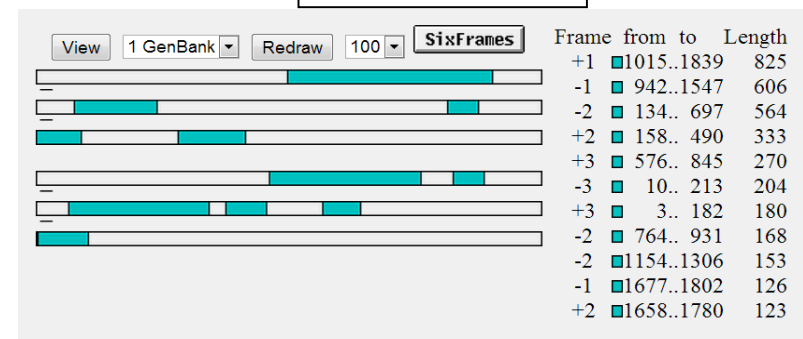
GRMZM2G035599



GRMZM2G891282

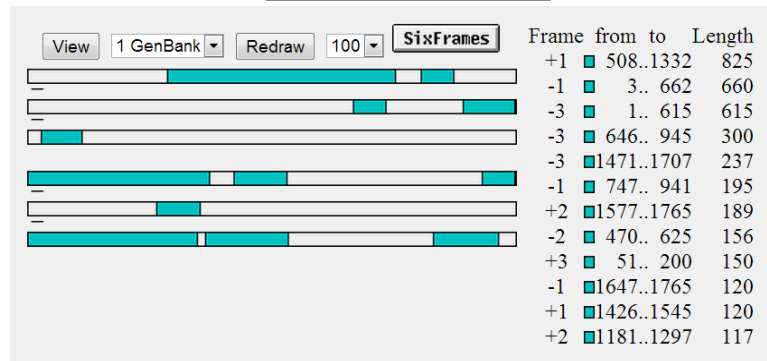


GRMZM2G456086

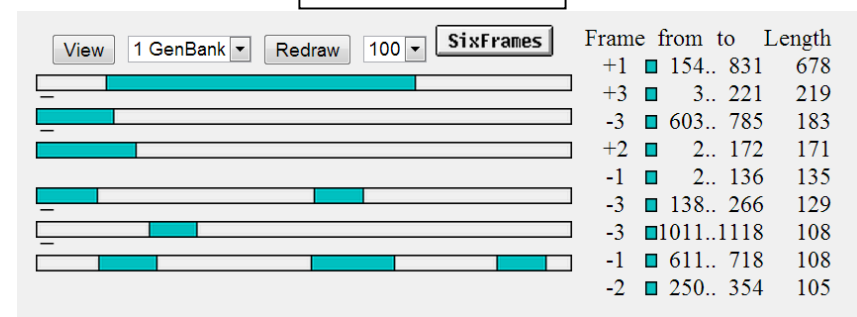


## Ribulose-phosphate 3-epimerase

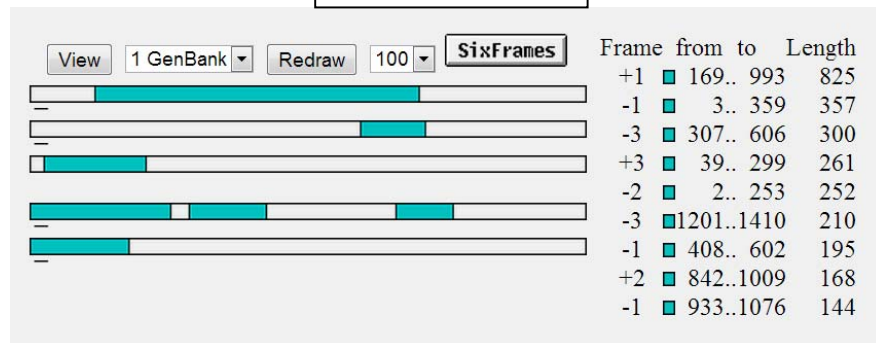
GRMZM2G026807



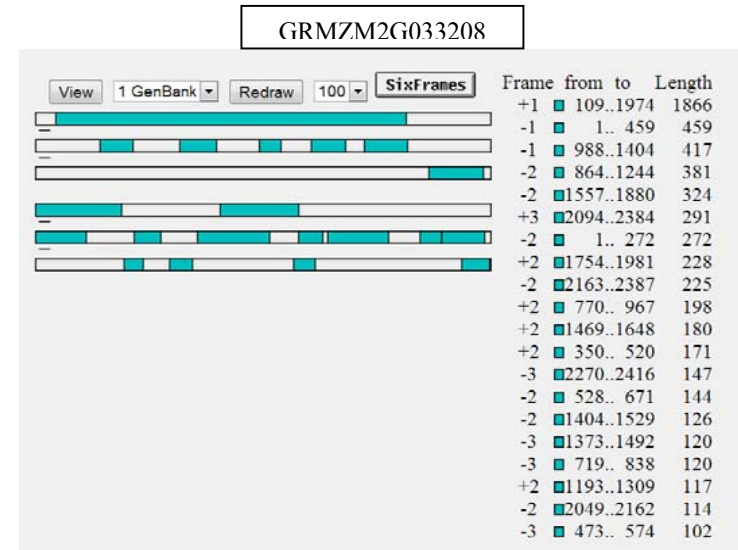
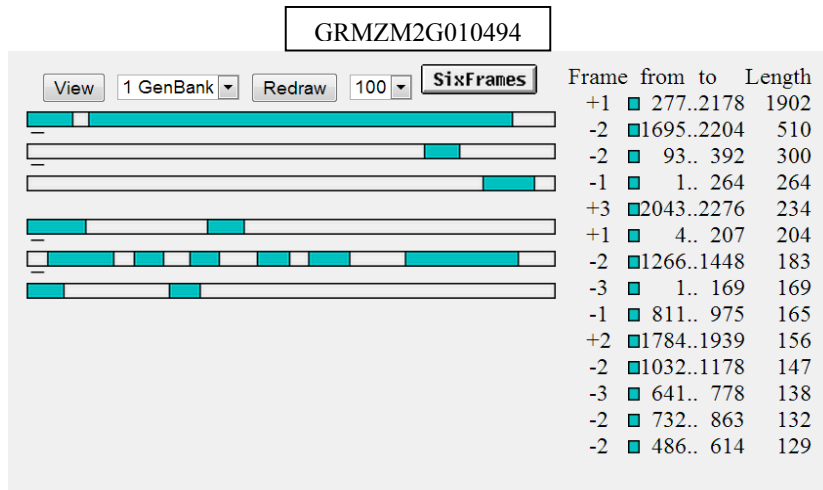
GRMZM2G083102



GRMZM2G178960



## Transquetolase



## Transaldolase

