

## Duas proteínas (*TcTIA6* e *TcTIA7*) envolvidas na biossíntese da tiamina codificadas pelo genoma do *Theobroma cacao* L

*Two proteins (TcTIA6 and TcTIA7) involved in the biosynthesis of thiamine coded by Theobroma cacao L. genome*

Recebido em: 30/08/2016

Aceito em: 20/05/2017

**Juliano Oliveira SANTANA; Gonçalo Santos SILVA;  
Luciano Angelo de Souza BERNARDES**

*Universidade Estadual de Santa Cruz (UESC), Ilhéus – BA, Brasil. km 16, Rodovia  
Ilhéus-Itabuna, Salobrinho, CEP 45662-900, Ilhéus, BA, Brasil.  
E-mail: julianosantana20@hotmail.com*

### ABSTRACT

Thiamine functions as a cofactor for activity of various enzymes that act on the metabolism of carbohydrates and amino acids. This study aimed to analyze through bioinformatic techniques the genes responsible for the biosynthesis of thiamine encoded by the genome of *Theobroma cacao* L. We found two genes, one in chromosome 6 and one in chromosome 7, being appointed as *TcTIA6* and *TcTIA7*. The multiple alignment revealed a strong identity with the *Arabidopsis thaliana* protein and its direction is destined to the secretory route of the chloroplast. Conserved motifs in the C-terminal region were found, indicating participation in the DNA repair mechanism. Analyzing the dendrogram it was evident that the genes studied had a common ancestor with *Gossypium raimondii*. The analysis showed that the *TcTIA6* and *TcTIA7* proteins are probably bifunctional, since they are involved in the biosynthesis of thiamine and stability of the DNA molecule.

**Key Words:** cacao, computational biology, genome, protein, thiamine

### RESUMO

A tiamina funciona como um cofator para atividades de várias enzimas que atuam no metabolismo de carboidratos e aminoácidos. Este trabalho objetivou analisar através de técnicas de bioinformática os genes responsáveis pela biossíntese da tiamina codificadas pelo genoma do *Theobroma cacao* L. Foram encontrados dois genes, um no cromossomo 6 e outro no 7, sendo nomeados *TcTIA6* e *TcTIA7*. O alinhamento múltiplo da *TcTIA6* e *TcTIA7* revelou alta identidade com a proteína da *Arabidopsis thaliana* e seu direcionamento é destinado para a rota secretora do cloroplasto. Foram observados motifs conservados na região C-terminal, indicando participação no mecanismo de reparo de DNA. Analisando o dendrograma ficou evidenciado que os genes estudados tiveram um ancestral comum com *Gossypium raimondii*. As análises mostraram que as proteínas *TcTIA6* e *TcTIA7* provavelmente são bifuncionais, uma vez que estão envolvidas na biossíntese da tiamina e na estabilidade da molécula de DNA.

**Palavras-chave:** cacau, bioinformática, genoma, proteína, tiamina

## INTRODUÇÃO

A vitamina B1, também conhecida como tiamina, é uma vitamina hidrossolúvel que atua como um cofator na descarboxilação do piruvato e do  $\alpha$ -cetoglutarato, essencial para o metabolismo aeróbico (1, 2). A tiamina é indispensável para o crescimento e desenvolvimento do organismo, atuando no sistema digestivo, cardiovascular, nervoso e ainda melhora o estado psicoemocional (3, 4).

Como um nutriente essencial, a tiamina não pode ser produzida pelo organismo humano e deve ser adquirida a partir da dieta (5). A ingestão dessa vitamina é essencial para manter o metabolismo saudável, pois atua nos carboidratos, lipídios e proteínas. Sua deficiência nutricional pode ocasionar sintomas neuropsiquiátricos, cansaço, perturbações no sono e alterações cardíacas (6).

Somente alguns procariotas, fungos e plantas são capazes de realizar a síntese da tiamina (7, 8). Em *Arabidopsis thaliana*, o gene *thi1* está relacionado com a via da biossíntese de tiamina, sendo composto por uma ORF de 1047 pares de bases, e codifica uma proteína de 349 resíduos de aminoácidos (9). A dedução da sequência polipeptídica da proteína THI1 revelou um alto grau de identidade com as proteínas envolvidas na biossíntese de tiamina em *Saccharomyces cerevisiae* (51,7% de identidade) e *Schizosaccharomyces pombe* (55,1% de identidade) (10).

O gene *thi1* também pode estar envolvido na proteção do DNA nas organelas (cloroplastos e mitocôndrias) contra danos, por meio do mecanismo de reparo de DNA por excisão de bases (11). Em bactéria (*Escherichia coli*) este gene tem a capacidade de restabelecer o mecanismo de reparo em organismos que eram deficientes para esse tipo de reparo (10, 12).

*Theobroma cacao* L. é uma árvore da família Malvaceae conhecida como cacaueteiro, e seu fruto, o cacau, possui grande relevância industrial e alimentícia (13). Das amêndoas fermentadas e torradas são obtidos licor, manteiga e torta, que são consideradas matérias-primas. A sua importância econômica é inserida principalmente pelo consumo de chocolate sob as mais variadas formas (14).

O presente trabalho relata pela primeira vez a caracterização de duas proteínas codificadas pelo genoma do *T. cacao*. Esse projeto foi executado utilizando técnica de bioinformática por meio de análises *in silico* e propõe a existência de duas proteínas que ainda não estão descritas e que participam no mecanismo de biossíntese da tiamina, como também no reparo do DNA da planta.

## MÉTODO

As sequências primárias dos resíduos de aminoácidos preditas para as duas proteínas codificadas pelo genoma do *T. cacao* foram obtidas no banco de dados do CocoaGenDB (<http://cocoagenadb.cirad.fr/>) (15). A identificação foi realizada com a submissão da sequência da proteína “*thi1*” de *Arabidopsis thaliana* no *Basic Local Alignment and Search Tool* – BLAST (16), empregando o algoritmo de busca *Programs Search Protein Databases Using a Protein Query* (BLASTp).

As sequências das proteínas identificadas foram submetidas a diferentes análises de bioinformática: alinhamento de sequências (Clustal Omega); descrição da família (Pfam); direcionamento da proteína (TargetP 1.1) (17); e predição da massa molecular (Compute pI/mW) (18).

O dendrograma filogenético foi construído com o programa MEGA 7 (19), utilizando como método de agrupamento, o Neighbor-Joining (20) com bootstrap de 1000 repetições, tendo seus valores anexados ao longo de cada ramo.

O programa BioEdit (Versão 7.2.1) foi utilizado para alinhamento das sequências primárias de aminoácidos e ao mesmo tempo gerar os arquivos no formato “fasta” para serem processados no MEGA 7. O dendrograma foi construído com 15 sequências de proteínas homólogas, distribuídas entre sete espécies de plantas, sendo elas: *Camelina sativa*, *Capsella rubella*, *Gossypium raimondii*, *Populus trichocarpa*, *Glycine max*, *Arabidopsis thaliana* e *Theobroma cacao*. Cada sequência de proteína foi identificada com seu número de acesso no banco de dados do NCBI.

## RESULTADOS E DISCUSSÃO

Devido à importância do cacaueteiro como fonte de matéria-prima para as indústrias de alimentos, vários trabalhos têm sido realizados almejando a obtenção de ferramentas para compreender e caracterizar as proteínas codificadas pelo seu genoma e ao mesmo tempo aumentar a produção e melhorar o aproveitamento do fruto (21, 22).

As duas proteínas relacionadas com a biossíntese de tiamina codificadas pelo genoma do *T. cacao* foram identificadas nos loci *Tc06\_g018680* e *Tc07\_g004790* no banco genômico do cacau – CocoaGenDB, respectivamente, sendo assim nomeadas: *TcTIA6* e *TcTIA7*. O gene no locus *Tc06\_g018680* situado no cromossomo

6 contém uma ORF de 1442 pb que codifica a *TcTIA6* com 357 resíduos de aminoácidos e com peso molecular de 38,1 kDa. O gene no locus *Tc07\_g004790* situado no

cromossomo 7 contém uma ORF de 1429 pb que codifica a *TcTIA7* com 359 resíduos de aminoácidos e com peso molecular de 37,9 kDa.

<b>Thi1</b>	-----MA-AIAST--LSLSSTKP-QRLFDSSFHG	25
<b>TcTIA6</b>	MPHVDTQSLITPHRISFIYHTKAPQNTPLQSM-AITATLTSLSSSPKPSFLDHKSSFHG	59
<b>TcTIA7</b>	-----MASSVATTLTSSSTKLYRN-TSLFESSFHG	28
	** :::* . : ..****	
<b>Thi1</b>	SAISAAPISIGLK--PR----SFSVRATTAGYDLNAFTFDPIKESIVSREMTRRYMTDMI	80
<b>TcTIA6</b>	TPIASRFTPIRSS--SQDS-AISMSLNTPPYDPLQSFNFQPIKESYVAREMTRRYMMDMI	116
<b>TcTIA7</b>	VPITPLSFHLKAKSSPRNASVMSAASSPPYDLNNFRFDPIKESIVSREMTRRYMMDMI	89
	* : . : : : * . ***: * *:***** *:***** **	
<b>Thi1</b>	TYAETD <u>DVVVVGAGSAGLS</u> AAYEISKPNVQVAII EQSVSPGGGAWLGGQLFSAMIVRKPA	140
<b>TcTIA6</b>	<u>TYADTDVIVGAGSAGLS</u> CAYEISKPNIRVAII EQSVSPGGGAWLGGQLFSAMVVRKPA	176
<b>TcTIA7</b>	<u>THADTDVVVVGAGSAGLS</u> CAYELSKNPSVQVAIVEQSVSPGGGAWLGGQLFSAMVVRKPA	149
	* *:***: :*****.***:***. : :***:*****:*****:*****	
<b>Thi1</b>	HLFLDEIGVAYDEQDQTYVVVKHAALFTSTIMSKLLARPVNKLFNAVAEEDLIVKGNRVGG	200
<b>TcTIA6</b>	<u>HI FLDELGIQYDEQENYVVIKHAALFTSTIMSKLLARPVNKLFNAVAEEDLIVKDNVRTG</u>	236
<b>TcTIA7</b>	<u>HRFLDELGIEYDEQDDYVVIKHAALFTSTIMSKLLARPVNKLFNAVAEEDLIVKGGRVGG</u>	209
	* ***: * : ***: ***:*****:*****:*****:*****:*****:*****	
<b>Thi1</b>	VVTNWALVAQNHHTQSCMDPNVMEAKIVVSSCGHDGPFGATGVKRLKSIGMIDHVPGMKA	260
<b>TcTIA6</b>	<u>VVTNWALVSMNHDTQSCMDPNVMEAKIVVSSCGHDGPFGATGVKRLKSIGMIDSVPGMKA</u>	296
<b>TcTIA7</b>	<u>VVTNWALVSMNHDTQSCMDPNVMEAKIVVSSCGHDGPFGATGVKRLKSIGMIDSVPGMKA</u>	269
	*****: * :*****: * :*****:*****:*****:*****:*****	
<b>Thi1</b>	LDMNTAE <u>DAIVRLTREVVP</u> GMIVTGMEVAEIDGAPRMGPTFGAMMISGQKAGQLALKALG	320
<b>TcTIA6</b>	<u>LDMNTAE DAIVRLTREIVP</u> GMIVTGMEVAEIDGAPRMGPTFGAMMISGQKAAHLALKALG	356
<b>TcTIA7</b>	<u>LDMNTAE DAIVRLTREIVP</u> GMIVTGMEVAEIDGSPRMGPTFGAMMISGQKAAHLALQSLG	329
	*****:*****:*****:*****:*****:*****:*****:*****:*****	
<b>Thi1</b>	LPNAIDGTLV--GNLSPELVLAAADSAETVDA	349
<b>TcTIA6</b>	QPNAIDGTFIEGGRVQPEFVLAAAETEGTVDA	387
<b>TcTIA7</b>	LPNALDGTYY--GNIHPELILAAADSAETADA	358
	***:*** : *.: ***:*****: * .**	

**Figura 1.** Alinhamento das proteínas de *Theobroma cacao* e *Arabidopsis thaliana*. Em preto: sequência de ligação a dinucleotídeos; em azul: motivo de ligação ao DNA; e resíduos sublinhados: representa o domínio funcional da proteína.

As análises das ORF's no Pfam indicaram que os genes codificam polipeptídeos característicos de proteínas da família *thi4* de plantas, ou seja, a família de enzimas responsáveis pela síntese do anel de tiazol. Essa hipótese ganha suporte ao analisar a sequência dos resíduos de aminoácidos das proteínas *TcTIA6* e *TcTIA7* com os resíduos de *Arabidopsis thaliana* já caracterizadas. Ainda no mesmo software, foram observados os domínios funcionais das duas proteínas. Na *TcTIA6* o domínio catalítico foi observado entre os resíduos de aminoácidos 103

e 336, na *TcTIA7* entre os resíduos 76 e 309 (Figura 1).

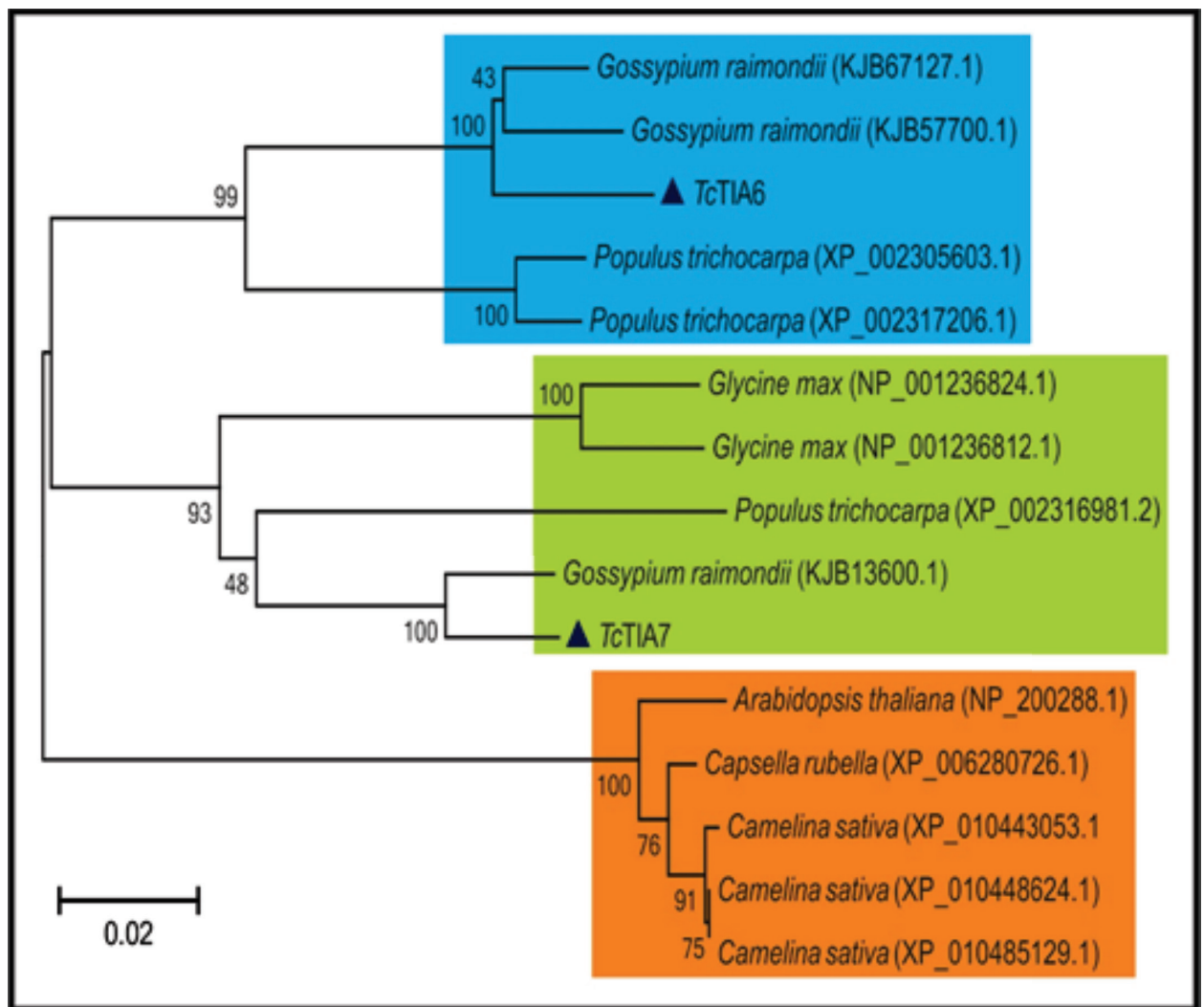
O alinhamento múltiplo entre as sequências primárias de aminoácidos (*TcTIA6* e *TcTIA7*) realizado com a proteína Thi1 da planta modelo *Arabidopsis thaliana* revelou alta identidade entre as duas espécies (*TcTIA6*: 77% e *TcTIA7*: 81%), com presença de vários resíduos de aminoácidos conservados (Figura 1). As sequências primárias das proteínas do cacau apresentam motivos conservados na região C-terminal (Figura 1), que segundo alguns autores, são resíduos que indicam a participa-



ção no mecanismo de reparo de DNA (10, 12, 23). Em outro estudo foi demonstrado que esses genes em *Citrus sinensis* (laranja) também possuem atividades durante a maturação e a senescência dos frutos (24).

Analisando os resultados do software TargetP, a predição do direcionamento celular revelou o encaminhamento para a rota secretora do cloroplasto. A probabilidade para a *TcTIA6* foi de 0.957 (cloroplasto) e 0.095 (retículo endoplasmático rugoso); para *TcTIA7*, a probabilidade foi de 0.916 (cloroplasto) e 0.012 (retículo endoplasmático rugoso), onde provavelmente nessas organelas realizam a biossíntese de tiamina. Porém, existem relatos em outros trabalhos que a síntese da tiamina possa ocorrer também no citoplasma da célula (25, 26).

Analisando o dendrograma, três grupos foram formados (Figura 2). As duas proteínas do *T. cacao* ficaram agrupadas em grupos distintos, porém permaneceram interligadas por uma subclasse comum com a espécie *Gossypium raimondii* (algodão). Não foi possível obter uma correlação das proteínas *TcTIA6* e *TcTIA7* com outras espécies, pois as duas sequências foram incluídas em grupos onde as proteínas homólogas foram apenas sequenciadas, no qual suas funções ainda não foram estabelecidas através de estudos em *in vitro*. Porém, foi descrito que a espécie *G. raimondii* e *T. cacao* pertencem à mesma subclasse, e que provavelmente divergiram de um ancestral comum, há cerca de 33.700 mil anos (27).



**Figura 2.** Dendrograma de 15 proteínas homólogas de espécies de plantas. Três grupos foram formados e as duas proteínas de *Theobroma cacao* (triângulos: *TcTIA6* e *TcTIA7*) permaneceram inseridas em grupos distintos.

## CONCLUSÃO

Os resultados obtidos sugerem que o genoma do *T. cacao* possui dois genes que codificam duas proteínas que participam na biossíntese de tiamina, sugerindo que o cacau e seus derivados são fontes de alimentos de vitamina B1.

Alguns microrganismos e plantas possuem a habilidade de realizar a síntese de tiamina, porém esse me-

canismo ainda não é compreendido, pois algumas etapas não estão totalmente elucidadas. Mas sabe-se que os genes exercem outras funções nas plantas como: reparo do DNA e maturação de frutos.

A bioinformática possibilitou, neste estudo, compreender a nível estrutural os genes e proteínas por meio de análises *in silico*. No entanto, ainda é necessária a caracterização *in vitro* desses genes para uma melhor compreensão de suas prováveis funções na fisiologia do *T. cacao*.

## REFERÊNCIAS

- Manzetti S, Zhang J, van der Spoel D. Thiamin function, metabolism, uptake, and transport. *Biochemistry*. 2014; 53(5):821-835. DOI: 10.1021/bi401618y
- Schabelman E, Kuo D. Glucose before thiamine for Wernicke encephalopathy: a literature review. *J Emerg Med* 2012; 42(4):488-494. DOI: 10.1016/j.jemermed.2011.05.076
- Isenberg-Grzeda E, Shen MJ, Alici Y, Wills J, Nelson C, Breitbart W. High rate of thiamine deficiency among inpatients with cancer referred for psychiatric consultation: results of a single site prevalence study. *Psychoneurology*. 2016; 26:1-6. DOI: 10.1002/pon.4155.
- Bubko I, Gruber BM, Anuszewska EL. The role of thiamine in neurodegenerative diseases. *Postepy Hig Med Dosw*. 2015; 69:1096-1106.
- Isenberg-Grzeda E, Kutner HE, Nicholson SE. Wernicke-Korsakoff-syndrome: underrecognized and undertreated. *Psychosomatics*. 2012; 53(6):507-516. DOI: 10.1016/j.psych.2012.04.008.
- Martin PR, Singleton CK, Hiller-Sturmhöfel S. The role of thiamine deficiency in alcoholic brain disease. *Alcohol Res Health*. 2003; 27(2):134-142.
- Roje S. Vitamin B biosynthesis in plants. *Phytochemistry*. 2007; 68(14):1904-1921. DOI: 10.1016/j.phytochem.2007.03.038.
- Jurgenson CT, Begley TP, Ealick SE. The structural and biochemical foundations of thiamin biosynthesis. *Annu Rev Biochem*. 2009; 78:569-603. DOI: 10.1146/annurev.biochem.78.072407.102340.
- Godoi PH, Galhardo RS, Luche DD, Van Sluys MA, Menck CF, Oliva G. Structure of the thiazole biosynthetic enzyme THI1 from *Arabidopsis thaliana*. *JBC*. 2006; 281(41): 30957-30966. DOI: 10.1074/jbc.M604469200.
- Machado CR, De Oliveira RL, Boiteux S, Praekelt UM, Meacock PA, Menck C.F. Thi1, a thiamine biosynthetic gene in *Arabidopsis thaliana*, complements bacterial defects in DNA repair. *Plant Mol Biol*. 1996; 31(3):585-593. DOI: 10.1007/BF00042231.
- Chabregas SM, Luche DD, Farias LP, Ribeiro AF, Van Sluys MA, Menck C. F, Silva-Filho M. C. Dual targeting properties of the N-terminal signal sequence of *Arabidopsis thaliana* THI1 protein to mitochondria and chloroplasts. *Plant Molecular Biology*. 2001; 46(6): 639-650. DOI: 10.1023/A:1011628510711.
- Garcia AF, Dyszy F, Munte CE, Demarco R, Beltramini LM, Oliva G, Costa-Filho AJ, Araujo AP. THI1, a protein involved in the biosynthesis of thiamin in *Arabidopsis thaliana*: structural analysis of THI1(A140V) mutant. *Biochim Biophys Acta*. 2014; 1844(6): 1094-1103. DOI: 10.1016/j.bbapap.2014.03.005.
- Bastos, E. Cacau: a riqueza agrícola da América. São Paulo: **Ícone**. 1987.
- Purdy L.H, Schmidt RA. Status of cacao witches' broom: Biology, Epidemiology, and Management. *Annu. Rev. Phytopathol*. 1996; 34:573-594. DOI: 10.1146/annurev.phyto.34.1.573.
- Argout X, Ruiz M, Turnbull C, Lanaud C, Rosenquist E. CocoaGen DB: a Web portal for crossing cocoa phenotypic, genetic and genomic data from ICGD and TropGeneDB data bases. In: 15<sup>th</sup> International Cocoa Research Conference. 2007; 1:515-518.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990; 215(3):403-410. DOI: 10.1016/S0022-2836(05)80360-2.
- Emanuelsson O, Brunak S, Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature protocols*. 2007; 2(4):953-971. DOI: 10.1038/nprot.2007.131.
- Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. Protein Identification and Analysis Tools on the ExPASy Server. In: John M. Walker (ed): *The Proteomics Protocols Handbook*, p.571-607, 2005.
- Kumar S, Tamura K, Nei M. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform*. 2004; 5(2):150-163. DOI: 10.1093/bib/5.2.150.

20. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol.* 1987; 4(4):406-25.
21. Leite PB, Lannes SCS, Rodrigues AM, Soares FASM, Soares SE, Bispo ES. Estudo reológico de chocolates elaborados com diferentes cultivares de cacau (*Theobroma cacao* L.). *Braz. J. Food Technol.* 2013; 16(3):192-197. DOI: 10.1590/S1981-67232013005000024.
22. Lopes UV, Monteiro WR, Pires JL, Clement D, Yamada MM, Gramacho KP. Cacao breeding in Bahia, Brazil – strategies and results. *Crop Breed Appl Biotechnol.* 2011; 11:73-81. DOI: 10.1590/S1984-70332011000500011.
23. Chabregas SM, Luche DD, Farias LP, Ribeiro A F, Van Sluys MA, Menck CF, Silva-Filho MC. Dual targeting properties of the N-terminal signal sequence of *Arabidopsis thaliana* TH11 protein to mitochondria and chloroplasts. *Plant Molecular Biology.* 2001; 46(6):639-650. DOI: 10.1023/A:1011628510711.
24. Jacob-Wilk D, Goldschmidt EE, Riov J, Sadka A, Holland D. Induction of a Citrus gene highly homologous to plant and yeast thi genes involved in thiamine biosynthesis during natural and ethylene-induced fruit maturation. *Plant Mol Biol.* 1997; 35(5):661-666. DOI: 10.1023/A:1005833724582.
25. Belanger FC, Leustek T, Chu B, Kriz AL. Evidence for the thiamine biosynthetic pathway in higher-plant plastids and its developmental regulation. *Plant Mol Biol.* 1995; 29(4):809-821. DOI: 10.1007/BF00041170.
26. Ribeiro DT, Farias LP, De Almeida JD, Kashiwabara PM, Ribeiro AF, Silva-Filho MC, Menck CF, Van Sluys MA. Functional characterization of the th11 promoter region from *Arabidopsis thaliana*. *J Exp Bot.* 2005; 56(417):1797-1804. DOI: 10.1093/jxb/eri168.
27. Wang K, Wang Z, Li F, *et al.* The draft genome of a diploid cotton *Gossypium raimondii*. *Nature.* 2012; 44(10):1098-1104. DOI: 10.1038/ng.2371.