CAPÍTULO 5

ANOTAÇÃO DE SEQUÊNCIAS GÊNICAS DE ARROZ (*Oryza sativa* L.) IDENTIFICADOS NO GENOMA DA ESPÉCIE PELO BANCO DE DADOS GENÔMICOS Phytozome

Data de aceite: 02/10/2023

Newton Chaves de Oliveira
Universidade Estadual do Piauí
Uruçuí- Piauí (Brasil)
https://orcid.org/0009-0005-6266-8369

Leonardo França da Silva
Universidade de Federal Viçosa
Viçosa – Minas Gerais (Brasil)
https://orcid.org/0000-0002-9710-8100

Fabrício Custódio de Moura Gonçalves
Universidade Estadual do Piauí
Uruçuí- Piauí (Brasil)
https://orcid.org/0000-0003-4319-635X

Anarlete Ursulino Alves
Universidade Estadual do Piauí
Uruçuí- Piauí (Brasil)
https://orcid.org/0000-0003-1208-5841

Francisco de Assis Gomes Junior
Universidade Estadual do Piauí
Uruçuí- Piauí (Brasil)
https://orcid.org/0000-0002-0854-821X

Victor Crespo de Oliveira
Universidade de Federal Viçosa
Viçosa – Minas Gerais (Brasil)
https://orcid.org/0000-0003-2719-9972

Ariadna Faria Vieira
Universidade Estadual do Piauí
Uruçuí- Piauí (Brasil)
https://orcid.org/0000-0002-1185-4269

RESUMO: O arroz (*Oryza sativa* L.) é o segundo cereal mais cultivado no mundo, sendo base alimentar de aproximadamente 3 bilhões de pessoas, contudo, nos próximos anos, à redução da área arável e de fatores bióticos e abióticos afetaram a produção da cultura. Uma das alternativas para manter a estabilidade produtiva da cultura é a identificação de genes relacionados à caracteres de interesse de melhoramento para o arroz. Essa identificação é feita utilizando ferramentas biotecnológicas, como os marcadores moleculares SNPs (*Single Nucleotides Polymorphisms*), que são definidos como marcadores de DNA que possibilita a identificação de genes interesse para a agricultura. Diante disso, o objetivo do trabalho foi identificar genes de interesse em oito marcadores SNPs identificados como polimórficos para a cultura de arroz através de análises *in silico*, utilizando o banco de dados genômicos *Phytozome*. Esses genes foram descritos com base na sua função, além de serem identificados a sequência genômica, sequência CDS (*Coding* DNA *Sequence*), sequência do peptídeo e sequência do transcrito. Como resultados, foi possível identificar seis SNPs dentro de genes, localizados em seis cromossomos distintos. Diversas

funções foram identificadas nos genes, como proteínas hipotéticas, proteínas relacionadas a calmodulina, proteínas G, domínio de ligação ARK e expressão de proteínas. Em arroz, uma das funções mais importantes estão relacionadas a expressão das proteínas quinases localizada no gene (LOC_Os04g45920, SNP S04_27195948) que, por sua vez, reconhece o estresse hídrico em arroz, representando fontes potencias a serem incorporados em novas cultivares, através da identificação desses genes em materiais como variedades tradicionais. Sendo assim podem ser de interesse de melhoramento para a cultura do arroz ou para outras culturas.
PALAVRAS-CHAVE: genômica, banco genômico; marcadores moleculares

ABSTRACT: Rice (Oryza sativa L.) is the second most cultivated cereal in the world, being the food base of approximately 3 billion people, however, in the coming years, the reduction of the arable area and of biotic and abiotic factors affected the production of the crop. One of the alternatives to maintain the productive stability of the crop is the identification of genes related to traits of interest for rice improvement. This identification is carried out using biotechnological tools, such as molecular markers SNPs (Single Nucleotides Polymorphisms), which are defined as DNA markers that enable the identification of genes of interest to agriculture. Therefore, the objective of this work was to identify genes of interest in eight SNPs markers identified as polymorphic for the rice crop through in silico analysis, using the Phytozome genomic database. These genes were sensed based on their function, in addition to identifying the genomic sequence, CDS sequence (Coding DNA Sequence), peptide sequence and transcript sequence. As a result, it was possible to identify six SNPs within genes, located on six different chromosomes. Several functions were identified in the genes, such as hypothetical proteins, calmodulin related proteins, G proteins, ARK binding domain and protein expression. In rice, one of the most important functions is related to the expression of protein kinases located in the gene (LOC_Os04g45920, SNP S04_27195948) which, in turn, recognizes water stress in rice, representing sources of potency to be incorporated into new cultivars, through identification of these genes in materials such as traditional varieties. Therefore, it may be of interest to improve rice or other crops.
KEYWORDS: genomics, genomic bank; molecular markers

1 | INTRODUÇÃO

O arroz (*Oryza sativa* L.) é o segundo cereal mais cultivado no mundo e representa a base alimentar para cerca de três bilhões de pessoas. O Brasil é o maior produtor da cultura fora do continente asiático, com produção em 2020 de 11.074.245 toneladas, em uma área de 1.680.100 hectares com produtividade média de 6.591 kg/ha (SILVA, 2020). O Rio Grande do Sul é o principal estado produtor de arroz, no qual é responsável por cerca de 70% da produção brasileira conforme dados da safra de 2018/2019 (SOSBAI, 2018).

Mesmo sendo um cereal com grande produção e produtividade mundial, a demanda por alimentos irá aumentar de forma significativa nos próximos anos devido à redução das áreas agricultáveis e pelos fatores bióticos e abióticos. Neste contexto, é preciso adotar estratégias para minimizar os problemas na produtividade do arroz, principalmente

relacionado a genética e o melhoramento da cultura (CHAVES; OLIVEIRA, 2004).

Os programas de melhoramento priorizam, dentre vários objetivos, a identificação de genes em progênies de linhagens, cultivares ou em variedades tradicionais para obtenção de genótipos que contenham um espectro, de por exemplo, fatores bióticos ou abióticos, Dessa maneira, um dos primeiros passos é a identificação de genes para estudos genéticos e genômicos posteriores utilizando marcadores moleculares. (RAO, 2004).

Os marcadores moleculares podem ser definidos como fragmentos de DNA que permitem a identificação de locos ou regiões genômicas, os quais podem conter genes diferencialmente expressos ou que podem distinguir indivíduos ou genótipos geneticamente divergentes. São ilimitados e fáceis de serem identificados no genoma e isso tem contribuído para a grande utilização nos programas de melhoramento (FALEIRO, 2007). Uma classe que é bem difundida no melhoramento para a cultura de arroz são os marcadores SNPs (*Single Nucleotides Polymorphisms*) ou polimorfismos de um único nucleotídeo (LIU; ZHANG, 2006).

Os SNPs são marcadores bialélicos e são caracterizados pela mudança em uma única base de DNA. São abundantes no genoma, com ampla distribuição no genoma, e podem ser encontrados em regiões que codificam proteínas, íntrons e em regiões flanqueadoras de genes, São utilizados para vários estudos, entre eles a identificação de genes com algum potencial agronômico. São amplamente utilizados em estudos genômicos de arroz, sendo um importante marcador para a cultura (SINGH *et al*., 2014).

Dessa maneira, o trabalho teve como objetivo identificar genes por marcadores SNPs na cultura do arroz (*O. sativa*.) em um banco de dados genômicos. Posteriormente a identificação, os genes foram descritos e anotados conforme as suas funções biológicas. Esse estudo permitirá que outros trabalhos poderão ser desenvolvidos para outras culturas, levando o entendimento da genética e melhoramento de plantas nos processos atuais para o desenvolvimento de cultivares e variedades de plantas de importância agronômica.

2 I MATERIAL E MÉTODOS

2.1 Local de execução do projeto

Empregou-se análises *in silico*, utilizando dados de marcadores de Polimorfismos de nucleotídeo único (SNPs). Esses marcadores foram identificados em possíveis genes no genoma do arroz, utilizando o banco de dados genômicos do Phytozome disponível em <https://phytozome-next.jgi.doe.gov>.

2.2 Marcadores SNPs e análises *in silico*

Foram utilizados oito marcadores SNPs obtidos na literatura identificados como polimórficos para a cultura do arroz (Tabela 1) com o objetivo de identificá-los em genes de interesse para a cultura do arroz.

SNP	Cromossomo	Posição
S01_23885433	1	23885433
S03_3311741	3	3311741
S04_27195948	4	27195948
S05_28201815	5	28201815
S05_29213630	5	29213620
S08_22353023	8	22353023
S09_18281732	9	18281732
S10_15699535	10	15699535

Tabela 1- Marcadores SNPs encontrados em arroz (*O. sativa*)

Em um primeiro momento, esses SNPs foram identificados no genoma do arroz no banco de dados do *Phytozome.* Para tanto algumas etapas principais serão necessárias, tais como:

1) No banco de dados foi selecionado a ferramenta *JBrowser* para o genoma de *O. sativa* v7.0.

2) Com o genoma disponível, foram localizados os cromossomos onde os SNPs foram identificados e a posição de cada um deles nos cromossomos.

3) Os genes identificados para cada marcador SNP foi descrito com base na sua sequência genômica, sequência do transcrito, sequência CDS (*Coding* DNA *Sequence*), sequência do peptídeo e outras informações tais como a identificação do gene, nome do transcrito.

Os genes identificados no *JBrowser* foram anotados quanto a sua função putativa no site do RGAP- *Rice Genome Annotation Project* acessado em (http://rice.plantbiology.msu.edu).

3 | RESULTADOS E DISCUSSÃO

De todos os marcadores single nucleotide polymorphisms avaliados, seis foram identificados em genes na cultura do arroz, exceto os SNPS S05_28201815 e S08_22353023, cujo polimorfismos não foram identificados em nenhum segmento gênico. Tais resultados evidenciam a importância do uso de marcadores moleculares SNPs na identificação de genes de importância agronômica para a cultura do arroz. (Tabela 1). Diversas informações dos genes foram indicadas, como: SG: Sequência genômica; ST: Sequência do Transcrito; SCDS: Sequência CDS; SP: Sequência do Peptídeo; FG: Função do gene.

SNP	CROMOSSOMO	POSIÇÃO	GENE	TRANSCRITO
S01_23885433	1	23885433	LOC_Os01g42130	LOC_Os01g42130.1
S03_3311741	3	3311741	LOC_Os03g06570	LOC_Os03g06570.1
S04_27195948	4	27195948	LOC_Os04g45920	LOC_Os04g45920.1
S05_28201815	5	28201815	-	-
S05_29213630	5	29213620	LOC_Os05g50910	LOC_Os05g50910.1
S08_22353023	8	22353023	-	-
S09_18281732	9	18281732	LOC_Os09g30070	LOC_Os09g30070.1
S10_15699535	10	15699535	LOC_Os10g30210	LOC_Os10g30210.1

Tabela 2. Marcador SNPs identificados em genes na cultura do arroz (*Oryza sativa* L.)

O marcador S01_23885433, localizado no cromossomo um e na posição 23885433 (Tabela 3), foi identificado no gene LOC_Os01g49070 (Figura 1). Este gene está associado *a hypothetical protein* (proteína hipotética).

LOC_Os01g42130 (S01_23885433)	
SG	ATGGCCGGCGGCCTCCTCAGCAAAGCGTCCAGCGCCGTGGCCGCCTGCGCGCGC-CGCGTGTCGCGCGCCACCCGCCGCCTCCTCCGCGCGCGCCTGCTCCGCCGGGGC-GGCGGCGGCGGTGGTGGCGAGACCGGGAAGCCCGACGGCGGTGGCGGCGGTGG-TGGCGAAGGGCTGTGGCGGCGGGCGATACTGATGGGGGAGCGGTGCGAGCCGCT-CAGCTTCCCGGGCGCCATCCCTACGACAGCCGCGGGCGGCGGCTCTCGCAGCCGC-GCCGCGCCAAGGCGAAGCCGGCGGCGGCCACCGCCGCGCTGCTCTGCCGCTCGT-CGGACGCCGTCGACGAGGCCGTCGCGGCGGCGAACAACAGCAAGGCCGCCAGATA-CGTG
ST	ATGGCCGGCGGCCTCCTCAGCAAAGCGTCCAGCGCCGTGGCCGCCTGCGCGCGCC-GCGTGTCGCGCGCCACCCGCCGCCTCCTCCGCGCGCGCCTGCTCCGCCGGGGCG-GCGGCGGCGGTGGTGGCGAGACCGGGAAGCCCGACGGCGGTGGCGGCGGTGGT-GGCGAAGGGCTGTGGCGGCGGGCGATACTGATGGGGGAGCGGTGCGAGCCGCTCA-GCTTCCCGGGCGCCATCCACTACGACAGCCGCGGGCGGCGGCTCTCGCAGCCGC-GCCGCGCCAAGGCGAAGCCGGCGGCGGCCACCGCCGCGCTGCTCTGCCGCTCGT-CGGACGCCGTCGACGAGGCCGTCGCGGCGGCGAACAACAGCAAGGCCGCCAGATA-CGTGGCGGTGTCGCTGCTCCGAGACTAGACATCGCGCGCGTGGCGCGC
SCDS	ATGGCCGGCGGCCTCCTCAGCAAAGCGTCCAGCGCCGTGGCCGCCTGCGCGCGCC-GCGTGTCGCGCGCCACCCGCCGCCTCCTCCGCGCGCGCCTGCTCCGCCGGGGCG-GCGGCGGCGGTGGTGGCGAGACCGGGAAGCCCGACGGCGGTGGCGGCGGTGGT-GGCGAAGGGCTGTGGCGGCGGGCGATACTGATGGGGGAGCGGTGCGAGCCGCTCA-GCTTCCCGGGCGCCATCCACTACGACAGCCGCGGGCGGCGGCTCTCGCAGCCGC-GCCGCGCCAAGGCGAAGCCGGCGGCGGCCACCGCCGCGCTGCTCTGCCGCTCGT-CGGACGCCGTCGACGAGGCCGTCGCGGCGGCGAACAACAGCAAGGCCGCCAGATA-CGTGGCGGTGTCGCTGCTCCGAGACTAG
SP	MAGGLLSKASSAVAACARRVSRATRRLLRARLLRRGGGGGGGETGKPDGGGGGGGE-GLWRRAILMGERCEPLSFPGAIHYDSRGRRLSQPRRAKAKPAAATAALLCRSSDAVDE-AVAAANNSKAARYVAVSLLRD*
FG	expressed protein

Tabela 3. Informações genômicas do gene LOC_Os01g42130, identificado pelo marcador SNP S01_23885433

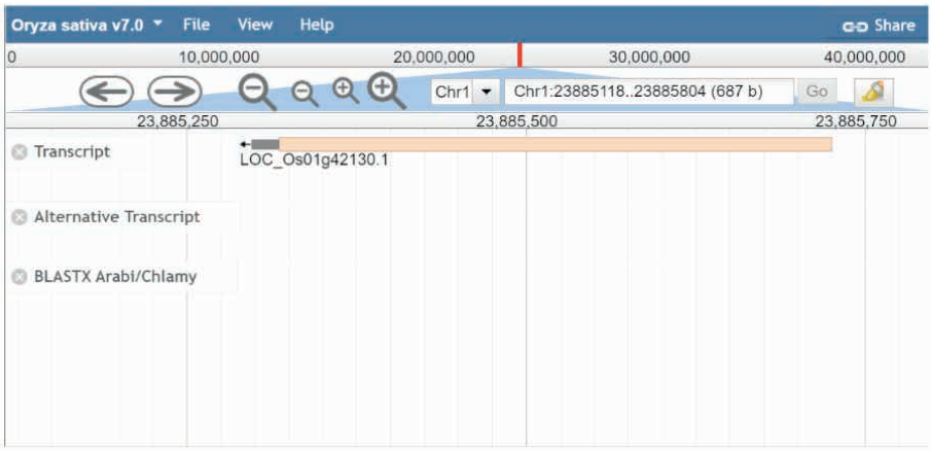

Figura 1. Identificação da posição do gene LOC_Os01g42130 no banco de dados genômicos

Mesmo com função generalizada (proteína expressa), esse gene foi identificado em dois QTLs (*Quantitative Trait Locus* ou Locos de características quantitativas) previamente relacionados a tolerância à seca em arroz (rfw1b, brt1d), sendo que o QTL brt1d está associado a espessura da raiz basal de arroz (YAMBAO *et al.*,1992) e o rfw1b está associado ao peso seco da raiz (LI *et al.*, 2005).

O estresse hídrico afeta diversos processos bioquímicos, fisiológicos e morfológicos nas plantas, sendo que as respostas das plantas a esses processos dependem do estádio de desenvolvimento da planta por exemplo em arroz, os estádios mais afetados são os reprodutivos, como o período de floração e enchimento de grãos, e na dependência e duração da severidade do estresse (VIDAL *et al.*, 2005). Nas plantas, sob a condição de déficit hídrico, ocorre o fechamento dos estômatos que reduz os níveis de transpiração e o suprimento de CO_2, além de diminuir o crescimento das células e aumentar a fotorrespiração (SHINOZAKI; YAMAGUCHI-SHINOZAKI, 2007). No mais, esses efeitos provocam a perda de vigor, diminuição de altura e consequentemente perda de produtividade na cultura (JIN *et al.*, 2013).

O marcador S03_3311741, localizado no cromossomo três e na posição 3311741 (Tabela 4) foi identificado no gene LOC_Os03g06570 (Figura 2), sendo a função associada relacionada a Calmodulina (IQ *calmodulin-binding motif*).

LOC_Os03g06570 (S03_3311741)	
SG	ATGTTCTTAGCAGTTTGTCCAAAAGCTTCACGTGGCTCTCTTCTTAATTCTTAT- TCAGCTCGCCACACTGTTGACTGAAAAATTGGAGCAGGCACTGCAGG GTG- GTTGTCTGAATTTGATCTTTGTATCTGTTTAGGTAGTAGCCATGGATGACAA- ATTGTTCGTTGTGATATGGTTTGAGATTTTGTGAAATTTTTGGTGAATTTTCAG TGATCGGTTAGGAGGGAGGAGGA ATGGGGAAGGCGGGGAGATGGCTGAGGA- GCTTCTTGGCCGGCGGCAAGAAGGGTGGCAAGAAGGGCGAGGCAATGGCGGC- GGCTTTGCCCGGAGAGGCGGCGAAGGAGAAGCGGTGGAGCTTCAGGCGACCG- GTGCACGGCGAGAAGGCGGCGGCGGAGGCGGCGGCGGCGGCGGACGGCGTG- GTGGTGGGCGAGGCCGAGGCGGGGTTTGATCTGTCCGCGTCGGAGTCGGAGT- TCGACCAGAAGAGACACGCCATGGCGGTCGCGGTGGCGACCGCGGCCGCCGC- CGACGCCGCCGTGGCCGCGGCGCACGCCGCGGCCGCCGCCGTCCGCCTCTCC- TCCCGCAAGGCGCACCAGCTGCCGGCGAGCGCCGTCGAGGAGGCGGCGGCTG- TCAGGATTCAGGCCACCTTCAGAGGCTACCTGGTAATAACGCACAAAATTTCCCTC- TCCTACGTGTTCGTCGGATCGATCACGACGTGTTCGTCCGCGTGACAATGCAG GCAAGAACAGCGCTGTGCGCGCTGAGGGGCATCGTGAAGTTGCAAGCTCTGGTGCGAGGC- CAGCTCGTGAGGAAGCAGGCGACCGCCACGCTCCGCTGCATGCAGGCTCTCCTCGCGGC- GCAGTCGCAGCTGCGCGCGCAGGCGCAGCGGGTGCGCGCGCTGCACGAACACCACCGGA- CGCCGCCCAGGCCACGGCCGCCGTCGCCGCCGCAGCACCCGAGGCACCGCCGATCCTAC GTAAGTTTTTCACCGGAAAATTTAATCTGATGATATGAACAGACGCCCTGACGGGGAT- GGGGGTGGAGCAGGAGATGGACAGGTCGTGCGAGGAGAACGCCAAGATCGTGGA- GGTGGACAGCGGCGCCGGCGAGCCGGCGCGGCGCGGCGGCGAGTACGGCCAC- CACGGGCGGTGGTCGCCGGCGCCGTCGGCGATGACGGAGGTGATGAGCCCGAG- GGCGTACAGCGGCCACTTCGAGGACATGGCGTTCGCCGCGACGGCGCACAGCA- GCCCGCATCACGCGTCGGCGTCGTCGGAGCTGCTGTGCTGCCCGAGCTACATGGC- CAACACGGAGTCGTCCCGCGCCAAGGCGCGGTCCCAGAGCGCGCCGAGGCAGC- GCACCGACGCGCTGGAGCGGCAGCCGAGCCGCCGGAAGAGCGGCGGCGGCGG- CGGCGGCGCCAAGATGCAGCGGTCGTCGTCGTCGCACGCCGCGGCGGCCCAGC- GCGGCGCGCAGTTCCCATGGCCGGTGATCAAGCTGGACACGTCGAGCGCGTCGC- TCAAGGACAGCGAGTGCGGGTCGACGAGCTCCGTTCTCACCGCCGCCACCGTCTA- CAGCCGGACGCGGTCGCTAGTCGGATTCGAGGTGAGTTACGACTTCCGAGACATC- GTCCATGTTTTTTTCTAACGAAGATGAAAATAAAGTTAATGTATGTAAAATAAAAAGT- TATTAGGATATGATTTATTGAGCTTTAACTATTACAAACTTAAAAAATAGATTTATTTAA- TATATTTAAGCAACTTATATATATAATTTTTTTAAAAAAAACGTAGAGTATGAAAATAGAG- GTAAATTCTACTACTCTGTCTAAAAAAATAAACTCTAGCTATGAATATAGACATGTATG- TGTTTAGATTCGTAGCTAGGATTGATTTTTTTTTTCTTTGGACCGAGGGGGTATATCT- TAGTGTGAAAAACATAAGGAAAAAATTATTCACTCGCACTTGTTCTTTGCATGATCCAG GTGCGCAGGGGTTTGTACTGA GCTACAAGAATAAAAAATGCCATTGCCGGACGAACC- GGGAGCTTCGCCAATCGCCATGGATTGAATGTTCGTCAGTTTGCTCGACGTGTGATG- GATTCAGGATCATGAACTGTACATTCTGTGTAGGTTATTCTGTGTTGTGGTGACTGTA- ACATACCAGGAATACATACTGCTAAGTGAAAGCGAAATGTAATAAGCAATCCACTGCA- ATTTCTTCCAAATTCCAGTGTTTGTTCC
ST	ATGTTCTTAGCAGTTTGTCCAAAAGCTTCACGTGGCTCTCTTCTTAATTCT- TATTCAGCTCGCCACACTGTTGACTGAAAAATTGGAGCAGGCACTGCAGG TGATCGGTTAGGAGGGAGGAGGA ATGGGGAAGGCGGGGAGATGGCTGAG- GAGCTTCTTGGCCGGCGGCAAGAAGGGTGGCAAGAAGGGCGAGGCAATGGCGG- CGGCTTTGCCCGGAGAGGCGGCGAAGGAGAAGCGGTGGAGCTTCAGGCGACC- GGTGCACGGCGAGAAGGCGGCGGCGGAGGCGGCGGCGGCGGCGGACGGCGT- GGTGGTGGGCGAGGCCGAGGCGGGGTTTGATCTGTCCGCGTCGGAGTCGGAGT- TCGACCAGAAGAGACACGCCATGGCGGTCGCGGTGGCGACCGCGGCCGCCGC- CGACGCCGCCGTGGCCGCGGCGCACGCCGCGGCCGCCGCCGTCCGCCTCTCC- TCCCGCAAGGCGCACCAGCTGCCGGCGAGCGCCGTCGAGGAGGCGGCGGCT- GTCAGGATTCAGGCCACCTTCAGAGGCTACCTGGCAAGAACAGCGCTGTGCGC- GCTGAGGGGCATCGTGAAGTTGCAAGCTCTGGTGCGAGCCAGCTCGTGAGGAAG- CAGGCGACCGCCACGCTCCGCTGCATGCAGGCTCTCCTCGGCGCAGTCGCAGC- TGCGCGCGCAGGCGCAGCGGGTGCGCGCGCTGCACGAACACCACCGGACGC- CGCCCAGGCCACGGCCGCTCGCCGCCAGCACCCGAGGCACCGCCGATCCTAC GTGCGCAGGGGTTTGTACTGAGCTACAAGAATAAAAAATGCCATTGCCGGACGAACC- GGGAGCTTCGCCAATCGCCATGGATTGAATGTTCGTCAGTTTGCTCGACGTGTGATG- GATTCAGGATCATGAACTGTACATTCTGTGTAGGTTATTCTGTGTTGTGGTGACTGTA- ACATACCAGGAATACATACTGCTAAGTGAAAGCGAAATGTAATAAGCAATCCACTGCA- ATTTCTTCCAAATTCCAGTGTTTGTTCC

SCDS	ATGGGGAAGGCGGGGAGATGGCTGAGGAGCTTCTTGGCCGGCGGCAAGAAGGG- TGGCAAGAAGGGCGAGGCAATGGCGGCGGCTTTGCCCGGAGAGGCGGCGAAG- GAGAAGCGGTGGAGCTTCAGGCGACCGGTGCACGGCGAGAAGGCGGCGGCGGA- GGCGGCGGCGGCGGCGGACGGCGTGGTGGTGGGCGAGGCCGAGGCGGGGTTT- GATCTGTCCGCGTCGGAGTCGGAGTTCGACCAGAAGAGACACGCCATGGCGGTC- GCGGTGGCGACCGCGGCCGCCGCCGACGCCGCCGTGGCCGCGGCGCACGCCG- CGGCCGCCGCCGTCCGCCTCTCCTCCCGCAAGGCGCACCAGCTGCCGGCGAGC- GCCGTCGAGGAGGCGGCGGCTGTCAGGATTCAGGCCACCTTCAGAGGCTACCTG GCAAGAACAGCGCTGTGCGCGCTGAGGGGCATCGTGAAGTTGCAAGCTCTGGTGC- GAGGCCAGCTCGTGAGGAAGCAGGCGACCGCCACGCTCCGCTGCATGCAGGCTC- TCCTCGCGGCGCAGTCGCAGCTGCGCGCGCAGGCGCAGCGGGTGCGCGCGCTG- CACGAACACCACCGGACGCCGCCCAGGCCACGGCCGCCGTCGCCGCCGCAGCACC- CGAGGCACCGCCGATCCTACGAGATGGACAGGTCGTGCGAGGAGAACGCCAAGATC- GTGGAGGTGGACAGCGGCGCCGGCGAGCCGGCGCGGCGCGGCGGCGAGTACGGC- CACCACGGGCGGTGGTCGCCGGCGCCGTCGGCGATGACGGAGGTGATGAGCCCGA- GGGCGTACAGCGGCCACTTCGAGGACATGGCGTTCGCCGCGACGGCGCACAGCA- GCCCGCATCACGCGTCGGCGTCGTCGGAGCTGCTGTGCTGCCCGAGCTACATGGC- CAACACGGAGTCGTCCCGCGCCAAGGCGCGGTCCCAGAGCGCGCCGAGGCAGCG- CACCGACGCGCTGGAGCGGCAGCCGAGCCGCCGGAAGAGCGGCGGCGGCGGCG- GCGGCGCCAAGATGCAGCGGTCGTCGTCGTCGCACGCCGCGGCGGCCCAGCGCG- GCGCGCAGTTCCCATGGCCGGTGATCAAGCTGGACACGTCGAGCGCGTCGCTCAAG- GACAGCGAGTGCGGGTCGACGAGCTCCGTTCTCACCGCCGCCACCGTCTACAGCCG- GACGCGGTCGCTAGTCGGATTCGAG GTGCGCAGGGGTTTGTACTGA
SP	MGKAGRWLRSFLAGGKKGGKKGEAMAAALPGEAAKEKRWSFRRPVHGEKAAAE- AAAAADGVVVGEAEAGFDLSASESEFDQKRHAMAVAVATAAAADAAVAAAHAAAAA- VRLSSRKAHQLPASAVEEAAAVRIQATFRGYL ARTALCALRGIVKLQALVRGQLVRK- QATATLRCMQALLAAQSQLRAQAQRVRALHEHHRTPPRPRPPSPPQHPRHRRSY EMDRSCEENAKIVEVDSGAGEPARRGGEYGHHGRWSPAPSAMTEVMSPRAYSGHFED- MAFAATAHSSPHHASASSELLCCPSYMANTESSRAKARSQSAPRQRTDALERQPSR- RKSGGGGGGAKMQRSSSSHAAAAQRGAQFPWPVIKLDTSSASLKDSECGSTSSVLTAA- TVYSRTRSLVGFE VRRGLY*
FG	IQ calmodulin-binding motif (IQ)

Tabela 4. Informações genômicas do gene LOC_Os03g06570, identificado pelo marcador SNP S03_3311741

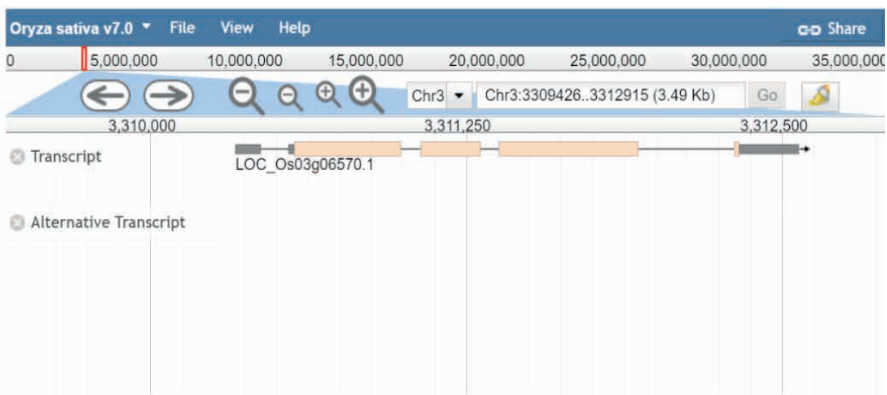

Figura 2. Identificação da posição do gene LOC_Os03g06570 no banco de dados genômicos

A calmodulina (CAM) é um sinalizador de Ca^{2+}, detectado em vários genes e encontrados em todos os eucariotos, apresentando diversas isoformas espécies-específica em plantas (BOUCHÉ *et al.,* 2005). Apesar CAM da não apresentar atividades enzimáticas, o complexo Ca^{2+}/CAM é capaz de funcionar como reguladores de diversos processos

celulares, modulando a atividade de muitas proteínas alvo (YANG; POOVAIAH, 2003).

Conforme estudos de Chinpongpanich *et al*. (2012), as CAM podem estar relacionadas as funções de regulações de estresses abióticos, que, para a cultura de arroz, é importante na relação para a tolerância à seca nas plantas.

O marcador SNP S04_27195948, localizado no cromossomo quatro e na posição 27195948 (Figura 3), foi identificado no gene LOC_Os04g45920 (Tabela 5) e está associado ao domínio de proteína quinase, que por sua vez é importante para regulação de sinais e desde alterações químicas (DROILLARD *et al*., 2000).

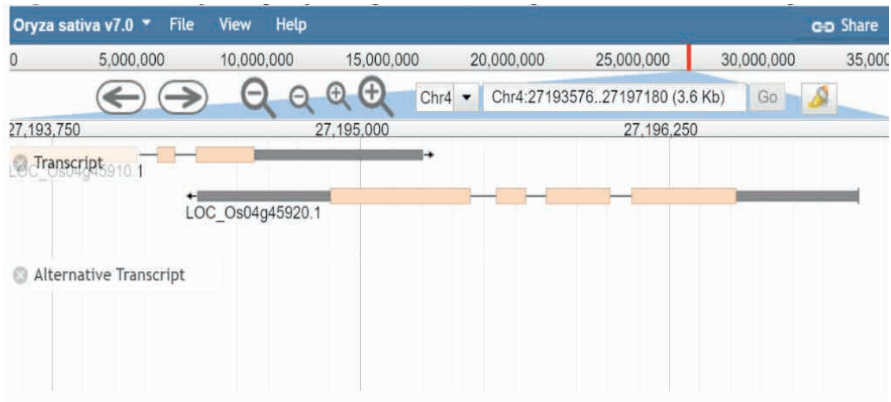

Figura 3. Identificação da posição do gene LOC_Os04g45920 no banco de dados genômicos

LOC_Os04g45920 (S04_27195948)	
SG	GCTTCGAGCAAAAGCAAGAAGGATAAAGAGGTTGGCAGGGGAGGTGACTGAGA- ACTGACTGGTTGAGCACTTTGATTCTTCGGAGCCGGTCAGCCGGATCACGGTAG- GCAGTACACAGTTCATGGTCCATGACTGTTGTATCAAACGATTCGGCCATTTGGAG- GCATGTCTCCTCTTGAGCATCTGCTGCCATTGAGTTTCTTGTGATCGGCAGCTGTC- TACTGTTGGATTTCTCTTCCATCTGGCAGAAATCTGGCTTTGCTGGATTTTTCTTC- TCTGATCGATTGGTTCTTCCTTCCACTGTGTTCTCTTTGCTCTTGGCCTTCTCCTG- GTTGGATTTTAATCACGTAGAAGTACAGACAAGTCATTGTCGAGCAACAGAAGAAA- TAGAAAATGTCTCTGTTTTAATCACCAATGCGTGATCTTTTTCTCTGCAACCTCTGA- GTTCTTGAAGTACCTGAAATTGTGGTAGCAGCCGGGCTTCTTCCTCCAAGAAATG- GGCTGCATGTGCCTCTTCCAGGACAAGCGCAGGAGCAAGAGACGGCCGGAGGT- GGCGGCGCCGGCGCCGGCACCGGCTCCGCCCTCCCCGAGTGCCGCTGCCGC- CGTTTCCACCGATGCGTGCTCGAACGTGTCCGTGCCAGCCGCCGCGCCGAGCA- CGTGCACGTCGTCGTGGGCGAGCACGAGGCCGTCGGCGAGCTCCTCGGCGGT- GAGCACCCCGGAGCCGTACGAGGCGAGGCAGGGCGCCCCCCGCGAGCTCGCTC- TCCGCGAGCTCCGCGGCGCCACGGGCGACTTCAGCCCGCTGCTCATGGTCGGC- CGGGGCGGCTTCGGGTGCGTCTACCGCGGCGTCCTCCGCCTCCCCGGCGAGC- CGCCCCACGGCACCCCCGTCGCCGTCAAGAGGCTCAACCCCGACAGCCGCCAG GCACGCACACACATGACGACACGCTCCCGGTTGATCTCGACCGCTTCCGCCATT- TGTGACGACTCCATTGTATCGTGCTCGTGCAG GGGCACAAGGAGTGGCTCGCC- GAGGTGCAGCTGCTGGGCGTCGTGGAGCACCCCAACCTCGTCAACCTCCTCG- GCTACTGCGCGGCGCAGACGGAGCGAGGCCCGCAGCGGCTGCTGGTGTACGA- GTTCGTGCCGAATAAGACGCTGGACGACCACCTGTTCGACCGATCCCACCCCG- TCCTCCCGTGGGGCGTCAGGCTGCAGATCGCGCTGGGCGCTGCCGAGGGGT- TACTGTACCTCCACGAGGGCCTGGAGTTTCAGGTAATCCACATCTTCTCCGAAAA- ACTGCGCGTCTTCATCAGAAAACACACCATTTTCACCTGATTTTCCTTGTTTGTCAG ATCATATACCGTGATTTCAAGGCTGCAAATGTGCTGCTGGATGATGAGTTCAGACCCA- AACTGTCAGATTTCGGATTAGCAAGGGAAGGGCCATCAGAAGGTCAGACACATGTCTC- CACAGCG GTATGTGCAAACATCACAGTTCAGAAAATGTGTCAACTCAACACGGATCA- TATCACTTCTGAATTCTTCAGAAATGAATCCGTTTTGGCTTTGGTGGATCAGGTGATGG- GGACGTACGGCTACGCGGCACCGGACTACGTCCGGACGGGGCACCTCACGACCAA- GAGCGACGTGTGGAGCTTCGGCGTTGTGCTGTACGAGATCCTCGCCGGCCGGCGGT- CGATCGACAAGAGTCGGCCCAAGGACGAGCAGAAGCTCCTCGAGTGGGTGCGGCG- GCACCCCGCCGGGAGCCCGCGGTTCGGCCGGATCATGGACGGGAGGCTGCAGGG- GAGGTACTCCGTGAGAGCGGCCAGGGAGGTCGCCGAGCTCGCCGCCGGCTGCCTC- GCGAAGCACGGCAAGGACCGCCCCGCGATGGCGGAGGTGGTCGAGCGGCTGAGGC- GGGCGACGCGGCACGCCGAGCTGGACGGCGAAGTGTATGATGATGCCGGGGAGGA- GAGCTCGAGCTCGCCGGCGGCGGCGGCCGTGGAGGACGACGTGGCCGTGGCCGC- GGCGGCGGCGAGGCGGCGGATGCTCCACCTTGCTGCGCTCGGGGAGAACGCGAGC- GCGAGCGCTCACGCGAGGAGAAGGCTCATGCTCATGAGGGCGGCTGCTGCTGCCAC- TGCCGCGACGTGACCTGTTCATTCTCTTGTCGGAGTTGACAATCACAATGCAGGATT- GACTTCTGAAGGATTTTTTTTTTGTGCTGAGATGATATTTTTTTGTTACGGGTTTTGAGAA- AAACAAATTACAGGACTTTGTGACTACAACTTGTACAGCATGGTTAATCCAAATAATACA- ATTTTCAGTAGAGACATTGTGACAACAAATTGAACAATGTGGTTCTTTTGTGTAAACGAC- CAAATCGAACAGCACACTAATAACTTCGTGAAACTGAAAGCTGAAGGAATCACATACAC- GCAGGAGCAGGCATGGAACGAATTCAGTTATCTGCAATTGGAAATGCTGACCCAATTTT- GCTGTTTCCCATTCGGTAGCACTTGATGTTTACTCCGTACGGGCCAATCTCAACGCGGT- TATCTTCCCCTCCCATAAAGAAGCAGAGCGTGTAGAGTGCAACCTCAAACTCAGGGCT- GACACCAATCAGAGTGCTGGATACCGATTTCAGAACACCGTGCCACTCAAACTGAATA- GTCAATAGCTGTGTCTCTGA

ST	GCTTCGAGCAAAAGCAAGAAGGATAAAGAGGTTGGCAGGGGAGGTGACTGAGA- ACTGACTGGTTGAGCACTTTGATTCTTCGGAGCCGGTCAGCCGGATCACGGTAG- GCAGTACACAGTTCATGGTCCATGACTGTTGTATCAAACGATTCGGCCATTTGGAG- GCATGTCTCCTCTTGAGCATCTGCTGCCATTGAGTTTCTTGTGATCGGCAGCTGTC- TACTGTTGGATTTCTCTTCCATCTGGCAGAAATCTGGCTTTGCTGGATTTTTCTTC- TCTGATCGATTGGTTCTTCCTTCCACTGTGTTCTCTTTGCTCTTGGCCTTCTCCTG- GTTGGATTTTAATCACGTAGAAGTACAGACAAGTCATTGTCGAGCAACAGAAGAAA- TAGAAAATGTCTCTGTTTTAATCACCAATGCGTGATCTTTTTCTCTGCAACCTCTGA- GTTCTTGAAGTACCTGAAATTGTGGTAGCAGCCGGGCTTCTTCCTCCAAGAAATG- GGCTGCATGTGCCTCTTCCAGGACAAGCGCAGGAGCAAGAGACGGCCGGAGGT- GGCGGCGCCGGCGCCGGCACCGGCTCCGCCCTCCCCGAGTGCCGCTGCCGC- CGTTTCCACCGATGCGTGCTCGAACGTGTCCGTGCCAGCCGCCGCGCCGAGCA- CGTGCACGTCGTCGTGGGCGAGCACGAGGCCGTCGGCGAGCTCCTCGGCGGT- GAGCACCCCGGAGCCGTACGAGGCGAGGCAGGGCGCCCCCCGCGAGCTCGCTC- TCCGCGAGCTCCGCGGCGCCACGGGCGACTTCAGCCCGCTGCTCATGGTCGGC- CGGGGCGGCTTCGGGTGCGTCTACCGCGGCGTCCTCCGCCTCCCCGGCGAGC- CGCCCCACGGCACCCCCGTCGCCGTCAAGAGGCTCAACCCCGACAGCCGCCAG GGGCACAAGGAGTGGCTCGCCGAGGTGCAGCTGCTGGGCGTCGTGGAGCACCC- CAACCTCGTCAACCTCCTCGGCTACTGCGCGGCGCAGACGGAGCGAGGCCCG- CAGCGGCTGCTGGTGTACGAGTTCGTGCCGAATAAGACGCTGGACGACCACC- TGTTCGACCGATCCCACCCCGTCCTCCCGTGGGGCGTCAGGCTGCAGATCGC- GCTGGGCGCTGCCGAGGGGTTACTGTACCTCCACGAGGGCCTGGAGTTTCAG ATCATATACCGTGATTTCAAGGCTGCAAATGTGCTGCTGGATGATGAGTTCAGACCCA- AACTGTCAGATTTCGGATTAGCAAGGGAAGGGCCATCAGAAGGTCAGACACATGTCTC- CACAGCGGTGATGGGGACGTACGGCTACGCGGCACCGGACTACGTCCGGACGGGG- CACCTCACGACCAAGAGCGACGTGTGGAGCTTCGGCGTTGTGCTGTACGAGATCCT- CGCCGGCCGGCGGTCGATCGACAAGAGTCGGCCCAAGGACGAGCAGAAGCTCCTC- GAGTGGGTGCGGCGGCACCCCGCCGGGAGCCCGCGGTTCGGCCGGATCATGGACG- GGAGGCTGCAGGGGAGGTACTCCGTGAGAGCGGCCAGGGAGGTCGCCGAGCTCGC- CGCCGGCTGCCTCGCGAAGCACGGCAAGGACCGCCCCGCGATGGCGGAGGTGGTC- GAGCGGCTGAGGCGGGCGACGCGGCACGCCGAGCTGGACGGCGAAGTGTATGATGA- TGCCGGGGAGGAGAGCTCGAGCTCGCCGGCGGCGGCGGCCGTGGAGGACGACGT- GGCCGTGGCCGCGGCGGCGGCGAGGCGGCGGATGCTCCACCTTGCTGCGCTCGG- GGAGAACGCGAGCGCGAGCGCTCACGCGAGGAGAAGGCTCATGCTCATGAGGGCG- GCTGCTGCTGCCACTGCCGCGACGTGACCTGTTCATTCTCTTGTCGGAGTTGACAAT- CACAATGCAGGATTGACTTCTGAAGGATTTTTTTTTTTGTGCTGAGATGATATTTTTTTGT- TACGGGTTTTGAGAAAAACAAATTACAGGACTTTGTGACTACAACTTGTACAGCATGGT- TAATCCAAATAATACAATTTTCAGTAGAGACATTGTGACAACAAATTGAACAATGTGGT- TCTTTTGTGTAAACGACCAAATCGAACAGCACACTAATAACTTCGTGAAACTGAAAGC- TGAAGGAATCACATACACGCAGGAGCAGGCATGGAACGAATTCAGTTATCTGCAATTG- GAAATGCTGACCCAATTTTGCTGTTTCCCATTCGGTAGCACTTGATGTTTACTCCGTAC- GGGCCAATCTCAACGCGGTTATCTTCCCCTCCCATAAAGAAGCAGAGCGTGTAGAGTG- CAACCTCAAACTCAGGGCTGACACCAATCAGAGTGCTGGATACCGATTTCAGAACACC- GTGCCACTCAAACTGAATAGTCAATAGCTGTGTCTCTGAA

SCDS	ATGGGCTGCATGTGCCTCTTCCAGGACAAGCGCAGGAGCAAGAGACGGCCGGAG- GTGGCGGCGCCGGCGCCGGCACCGGCTCCGCCCTCCCCGAGTGCCGCTGCCGC- CGTTTCCACCGATGCGTGCTCGAACGTGTCCGTGCCAGCCGCCGCGCCGAGCA- CGTGCACGTCGTCGTGGGCGAGCACGAGGCCGTCGGCGAGCTCCTCGGCGGT- GAGCACCCCGGAGCCGTACGAGGCGAGGCAGGGCGCCCCCCGCGAGCTCGCTC- TCCGCGAGCTCCGCGGCGCCACGGGCGACTTCAGCCCGCTGCTCATGGTCGGC- CGGGGCGGCTTCGGGTGCGTCTACCGCGGCGTCCTCCGCCTCCCCGGCGAGC- CGCCCCACGGCACCCCCGTCGCCGTCAAGAGGCTCAACCCCGACAGCCGCCAG GGGCACAAGGAGTGGCTCGCCGAGGTGCAGCTGCTGGGCGTCGTGGAGCACCC- CAACCTCGTCAACCTCCTCGGCTACTGCGCGGCGCAGACGGAGCGAGGCCCGCA- GCGGCTGCTGGTGTACGAGTTCGTGCCGAATAAGACGCTGGACGACCACCTGTTC- GACCGATCCCACCCCGTCCTCCCGTGGGGCGTCAGGCTGCAGATCGCGCTGGGC- GCTGCCGAGGGGTTACTGTACCTCCACGAGGGCCTGGAGTTTCAGATCATATACC- GTGATTTCAAGGCTGCAAATGTGCTGCTGGATGATGAGTTCAGACCCAAACTGTCA- GATTTCGGATTAGCAAGGGAAGGGCCATCAGAAGGTCAGACACATGTCTCCACAGCG GTGATGGGGACGTACGGCTACGCGGCACCGGACTACGTCCGGACGGGGCACCTCAC- GACCAAGAGCGACGTGTGGAGCTTCGGCGTTGTGCTGTACGAGATCCTCGCCGGCC- GGCGGTCGATCGACAAGAGTCGGCCCAAGGACGAGCAGAAGCTCCTCGAGTGGGT- GCGGCGGCACCCCGCCGGGAGCCCGCGGTTCGGCCGGATCATGGACGGGAGGCTG- CAGGGGAGGTACTCCGTGAGAGCGGCCAGGGAGGTCGCCGAGCTCGCCGCCGGCT- GCCTCGCGAAGCACGGCAAGGACCGCCCCGCGATGGCGGAGGTGGTCGAGCGGCT- GAGGCGGGCGACGCGGCACGCCGAGCTGGACGGCGAAGTGTATGATGATGCCGGG- GAGGAGAGCTCGAGCTCGCCGGCGGCGGCGGCCGTGGAGGACGACGTGGCCGTG- GCCGCGGCGGCGGCGAGGCGGCGGATGCTCCACCTTGCTGCGCTCGGGGAGAACG- CGAGCGCGAGCGCTCACGCGAGGAGAAGGCTCATGCTCATGAGGGCGGCTGCTGCT- GCCACTGCCGCGACGTGA
SP	MGCMCLFQDKRRSKRRPEVAAPAPAPAPPSPSAAAAVSTDACSNVSVPAAAPSTCT- SSWASTRPSASSSAVSTPEPYEARQGAPRELALRELRGATGDFSPLLMVGRGGFGC- VYRGVLRLPGEPPHGTPVAVKRLNPDSRQ GHKEWLAEVQLLGVVEHPNLVNLLGYCA- AQTERGPQRLLVYEFVPNKTLDDHLFDRSHPVLPWGVRLQIALGAAEGLLYLHEGLEFQ IIYRDFKAANVLLDDEFRPKLSDFGLAREGPSEGQTHVSTA VMGTYGYAAPDYVRTGHLT- TKSDVWSFGVVLYEILAGRRSIDKSRPKDEQKLLEWVRRHPAGSPRFGRIMDGRLQGRYS- VRAAREVAELAAGCLAKHGKDRPAMAEVVERLRRATRHAELDGEVYDDAGEESSSSPAA- AAVEDDVAVAAAAARRRMLHLAALGENASASAHARRRLMLMRAAAAATAAT*
FG	protein kinase domain containing protein, expressed

Tabela 5. Informações genômicas do gene LOC_Os04g45920, identificado pelo marcador SNP S04_271995948

As proteínas quinases são encontradas em eucariotos e possuem as funções no controle intracelular, transdução e regulação de sinais. O mecanismo de regulação de sinais inclui desde alterações químicas até o controle transcricional. Na literatura, as famílias MAPKs (MAP quinases) são ativadas por estresses abióticos em plantas. Na planta modelo genômica, a *Arabidopsis*, as quinases (AtMAPK6) mostraram ser ativadas por hiperosmolaridade, sal, frio e seca (DROILLARD *et al*., 2000).

As plantas desenvolvem vias de sinalização complexas que traduzem estímulos bióticos ou abióticos em respostas celulares apropriados para se adaptar aos desafios ambientais em constante mudança. Essas vias de sinalização geralmente requerem a ativação de proteínas quinases para coordenar vários processos celulares por meio da fosforilação de vários substratos (XIE *et al*., 2014).

O marcador SNP S05_28201815, localizado no cromossomo cinco e na posição 28201815 não foi identificado em nenhum gene no banco de dados para a cultura de arroz, bem como o SNP S08_22353023, que está localizado no cromossomo três e na posição

22353023 e não foi identificado em região gênica no genoma de arroz.

O marcador SNP S05_29213630 foi localizado no cromossomo cinco (Figura 4) e na posição 22353023, identificado no gene LOC_Os05g50910 (Tabela 6). Esse gene está provavelmente associado em arroz a proteínas G. que, por sua vez, apresenta importante função que constituem papel crucial na sobrevivência celular podendo regular a expressão de genes envolvidos na sobrevivência, proliferação, diferenciação e outros processos celulares.

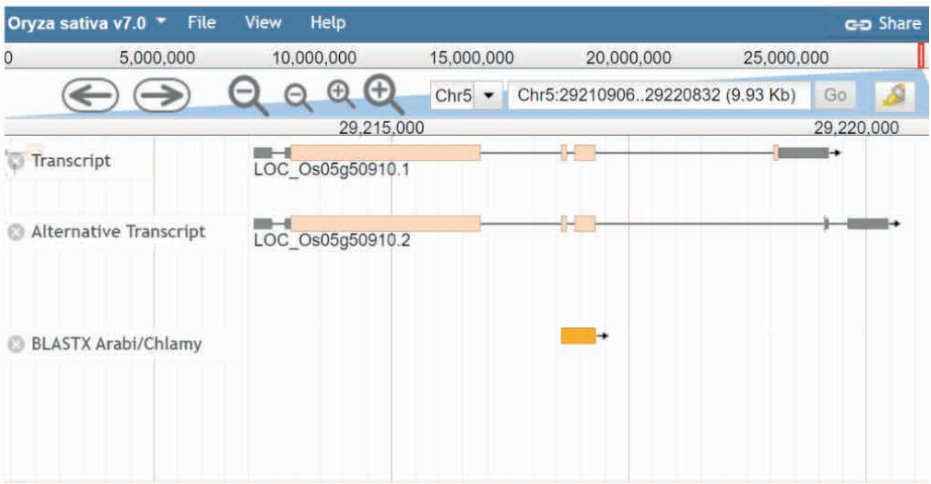

Figura 4. Identificação da posição do gene LOC_Os04g45920 no banco de dados genômicos

LOC_Os05g50910 (S05_29213630)

SG	ACTCGTCTTCTTCCTCTCGCCGCCGCAGCGCGCGGCCGCCACTCCATCCACCACAATGGCGAACCCCACGTCCCACCGCCTCCTCCTCGC- CGCGTAGCCCTATCGGCGAGCCGCTGCCGGTGGCTTCTCCGCATCCCCGCGCCGTAGAGTCTCCGCCGCCGCCGTGTGCTCCGCAACATT- GGGGTCGAGGTAACCGCCACCGCCGCCTCCTCCGCTTCTCGTAGCATTCAGAAGCTTCGCTGCTCTGATCTACATTGGCGTCTTCCCATTT- CGTGTTTGGTGGCTACGAATTTCGAGCGACTCATTGGAAGTTTTCCCCGCAGGGTTTTCGATTCGGCGAAGGGGGGGGGGGGGGGGGGGAA- AGCTCGTCCCGTGGGAGGCGCCATGGGGAGCGAGGCGAAGGAGATGAAGTACAGGAGAAGGGCTAGGGTTCCAGAGCCCTTCGATTACGG- GCAATGCGGCGGCGACCGGAGTGGCGTGCTGGATTGGGGCGCGCTCAAGGAGAACCCCGTGGAGCTGCTCCGGAAGCTCGACGAGCTG- CGGGACCACATCACCAGGTCCTGCGAGATCACCGACCAGCCGCGGGAGCGCCACCGCATGAGCCGCCGCACGGCCTCGCTGCGCCCCTC- GCACGCCGAGCCGCCGCCGCTAGGCCGTGGGCCGGAGCACTACCGCTCACGCTACACGGGAAGGTACGGGTCTGGCTTTCCGCACAGTC- CGAATGACCAGCTGCATCGGTCTATGCATAGAGATAGGTATGAGAGGCAGCCGAGTGGGCGGTTCCGGCAATGGCCGGAGAGGCAGTGGGA- GAATTCTGGGTATCTTGGGGGGAATCACCATCAGAGCACCTGCCAGTGTGCACAGTGCCTTCATGGCCAGAGAGCTGTGATGCAGGAGGAG- CACATTCCCATGACGAGGTACTTTGCAGGCCAGCAGGGGTCTCACCTGTTCGATAGGTCCCCATCAGTCTCATCGGAGCTTGACCGGAGGTC- TGTCGCTTCGTCATTGTACTCTCATTTCTCGGTGTCGAAGAGGAGGACAGAGTTTTTCAGGAAGAAGGCAGAGAGTTTCTGTCGTCCTGTGA- GAGGCGCCGCTCCTTTCGTTGTATGCAGTTCTTGTAACCAGCTGCTGCAGCTGCCTCCTGGGAAATGCACACAGCTCGGAAGCAGATTCAAGTT- CGGTGTGGTTCGTGCTCGGAGATCGTTAGTTTCAAGCTTAAGGAAGTGAAAATTCATCCTTTGGTTGCGCCGACATCTTTCCCTGCATCAAAA- ACAGTGGGGAGTTCCAGTCGTCAGGTTAACAAAAGTTTTGGGTGGTACCAACATCAGGATGAAGGAAATTCTAGTTTCCATAAGCTGCAAGCA- CAAGAGAGATGGCAGCAGAACAAGGATCTTGCAGACAATATTTCTGTGTCTTCTACTTCTAGTTATGACAGAATAGACAAAGAGTGTGGATCAA- ACAGGAGCAGTCAGTTACTATCAGTATCTGTTAGAAGGTCCAGACTTGCAAATATCCCAAAGGATATACTATGTCAAGGAGATGCATACAGTCAA- GTAGAAACTTCAGCATTTAATACAGGCAATCTACAGGCCCCAGTTATAGAGGACAAATGTGTCGATCCATTTTCCAGTCGTCTAAAAGATTGTAG- TGGTGGAGATCGCACAAGCAAGGAATGTAGCCTAAATATCATGGCAGATTCTGTTGATGCTAATGTTAGGGATGAAAGATCGGACGTGACATAT- GAGCAGAACAGCAAAGATCATAAAGAAGGATTTGGAGAAGAAACTGTCAGCAGTAGACATGAGCAAAAGCTCAAAGAAAGTACAAGTGGCTTT- TGTGATGATGGAAGCATGGGTAACATAGACAAGCTGACGGCTGATAATGATGATACCAGCAGCCTTGAAGATGGAGACGTGAGCAAAAAGTAT- GAAGAAAAAATCAAACAAGATGACAACAACTTTCAATCAGAATTCATTACTGAGCGATATAGCAAATGCAGCAAAGAAGATAATAACAGTGTCAT- TCAAGTTGAAACTATAGCTACGATATGCAAACAAGATGACTTAGATGATTGTTATAGTGAATTGCTCTCCCCAAATTCTGAACATGCCATAATGC- CATCGAAGCTTGAGTCATCAGTTAATGAGCGCACAAATTCTAGTTCTCGTGTTTCTTCTGAGGCTGAACTAGACGAAGTTCAGTCTGCAGCTAC- TAAGAACGGGGATTCAAAGTTTTTTGCTGGTTTTTTGAAGAAAGGTTTGAAGGACATTTCTTTGTTCAATCAATCAGTGGACAGTGCTAAGGTTT- CAATCAATGGCCATTCAATCTCCGAACGTGTCCTACGGAAGGCCGAGAAGAAGTCTGGTCCAGTTGGCCCTGGTTCATATTGGTAAGTCACCT- TCTCCCTTCATAAATTTGGAAATTGTTATGGACTGATGTTCTTAGCGTTTGATATCTACCGTGTTGGCACTTGTCATTGTTTTCTTTAGCATCTTTT- TATCATTTGCCTTTGAAATCAAGAGACGCACAACTACAGGCAAAAGAGATTGGTGTTATTCTTCAGAAGAATCTATCTTAACAGTAGGAATTAAGA- TGACTTATACTGCACAGCGGTGATTAAATGACTGTGTAATTCTCAATTGCATATTAACCAACCTATGTACCTGATTTCCTTGATAACTTTTCTCAGC- CAAATCAACCACTTCCCTGAGTTGATTACTAGAACATTTTGTACAGTGATGAGATACCTGGAAGATTCTTGTCCTGCAGAGTTTTTTTTTTTTTT- TGGCTGTTGATTGCTCCCATTAAATGCTAGTAGCTCAAAAAATTCTTGCATGGCCTAGTTCATGTATTAACATCTGTTAATTTATTATCACGGTCT- TACTTTGTCTCGGCATTCAGTAACTATGTGCAACATTAGTTTTGTTCTGTTATTGTTAATCGTAATGGTACAGTGTGAACTATAATTTATTTTCACA- CATCTGCATGCTTGCGACATGCAAGCTCCATAATTCAATGGTCTGCCTTTTACCAGCTACAGTTTTAAAATTAATCTGATTAGTATACAAAGACGA- GGAAAATGCAAAGTTAGTACAGTAAAAATTACTCTGGGAATATCATCTACCTAATGTACTTTCCAAAGCAACTTTGTTGGATCTTCACCAACTTA- AATTCACAAGGCATTGGGAATTTCCGTGTAAGGTAACATTCCGTATTCTCTAATTGTGATTCAGGTATGACTACCGCGCTGGATTTTGGGGTGT- CATGGGACATGAGTGTAGTGGCATCATCCCTGTAAGATGATTATTTTTCTCGTAGTTCTTTCTCTAGAACTTTGTGGTAGCACGTCATGACAGA- GCTTTGTTTTTGCAGCCATTTATAAAGGAATTTAATTATCCAATGCCGAAAAATTGTGCCGGTGGGAACACAGGAGTTATTGTCAATGGTAGAGA- ACTTCATCAGAAAGACTTTGAATTGCTTGCAGGAAGAGGTCTCCCACGGATATCTGGAAAATCATATTCTGTTGAGATCAATGGAAATGTGATT- GATGAAACAACTGGCAAAAAATTACGAAAGCTTGGAAAGCTTGCTCCTACGTAAGTTTCTTCTTCAGCTCAGTTCGAACAACACCATGTGAAA- TAGCATTTCCAGTATTTTGTGGTCACGTAGTGTCCATTCATTTCTTGCACAATATTTTCACCTTTTGGCCTCCTTACACTAGTCATTTTAAAGAACA- ACCTAATGAACAGAGTCTGACTAATTATTGAAGTTGGGAGTTGAGTGTGTGATCCAACAAGAATAGAAGAGCCAGGAGAAGGTTCAGATCATG- GCTTAAGATGTATTAAATTGGTAAGGCATTGAACACATGCCAAACTGGTGAAACAACTCTGTACAACATAACCAATAAGAGAAGAAAAGCCCGA- GCTTACTAACTGAAAAAAGCATGAGAAAAATGCAAAATTTTAAATAAATAAAGATGTTGTGCCAGGGGGACAGAAGATCTAGGCTGGACAGGATC- TAGCAGCAATATCTTTCAAGTTAAACAACCCTAGTCAAATAGCTCTAGTGTAGTCACAGGTTCACAGCATCTTCTGACAAAGCTAAAGACAACCA- GATCTTTATGTTGCCAGATTAACAAAAAAAAAAATCAGGCATCAACAAACATGAATACCAGGCAATTCAGTCAATTGCAGTGGTGTTTTCATGTT- GCACCTGTTGTTCTTGTGTCCCACGTTTTGAAATCATGGCCTCCAATGCACCGTATCATCTATGCCAAAGAATTAGTTAATTTGTTGTTGAATAG- GAGTATATCATTTCAAAGTAAGGCCTGGTTTAGTTCTCAACTTTTTCTTCAAACTTCCAACTTTTCCATCACATCAAAACTTCCCTACACACATAA- ACTTCCAACTTTCCGTCACATTGTTCCAATTTCAATCAAACTTCCAATTTTGGCGTGAACTAAACACACCCTAAATATAAAAACCCAATAACAAG- CAGCCAACGACTGTACATGATTTTTTTTCCCTTTTCACCTCAACTCTAATCACAGAATACGAAAAGTCCTCTGTAGTCTCTGCCTCTCTGGACA- TCTAGGCCAAGCTAATGCATCTCACATAACAGGCCAGAGGAAATTTAGGGGTTGTTTGGATTAGTCCCTGTCACATTGGATATTTGACACTAAT- TAGAAGTATTAAACGTAGACTATTGACAAAAACCATTCCATAACCCTGGACTAATTCGCGAGACGACTCTATTGAACCTAATTAATCCATGATTA- TAGCCTATGTGATGCTACAGTAAACATGTGCTAATTATGGATTAATTAGACTTAAAAAATTTATCACACGAATTAGTTCTCATTTATGTAATTAGTTT- TGTAAATAGTCTATGTTTAATACTCCAAATTAATGTCCAAACATCCGATTTGACAGGGACTAAAGTTTAGTCTTTGGATCCAAACACCACCTTAGC- TATAGAACACTAGAGCATGACATGATCAATACTTTCAGCTTGCGTGCATAGTAGTTACAACTCTTCCAACCTGTCTGGTCTTATTTCTTTAATTA- GATTTCCGGTTTTGTTTGTACTAACTAGATTGGTCGTTATGTCCTATCCAAGCTTCATTGTCCACAATTTACGGGCATCGGATCTTTTGCCCTC- CAAAGAATAGAGCGCTCTATGCATAACCTTGTCTTTTAGCTATTTTCATCTTGCTGTGGTTGTTTCAGTATTTTCATTGTTTGTAGTCTATTGTATA- CGTAGAAGTAAATCATCAGATGATATTGTGCAAGCTCATCATATCGTTCCTTTGGATGAAATTACGAGAGTTTTGTGTTACTAGTTTCTTCTATT- CATAATGATTACTCCCGTGTCAACGCTTAGACTCAGTAACATAAATTTTATCAGCAACTACGTTTTGATCAGATTTCTTAAAACCAGGCACTAACA- ATTTCAAATTTTAATGATCAGAGTTGAGAAGTTAAAACGTGGATTTGGCATGCACGTCCCTGAAGAGAAAAGTTAGCTCTCTGGCTCTGATGA- TGTTCCCTTGCTTTGCAAAAACTCATTTCCACGTAAAAGCTGTCAAGGGAGTTAAATTGGATTGTGGCATGGAAGAAGCCTTTAGATTGTACT- TCTTGAAGCTTGAGTGTGCAGCAACAGTAATACACGAGTACCTAAGACGACCTGTTTTTTATATATATTTACCACTTACTCATTCAAAGTTGACAA- ATTGTCGTACAAATTCCATTACTTCATTTGAAGCATTCTTTAGGTATATATGACCTAAGCATGCGAACAGCTCTAAGTTCCTCATGTGTCTGGGTA- TGTTGTTTCCTGTCCCCTGCATATGTTGTTGATACAGAGCAATAGAGTATAGACAATATTCTGGGTGTGTATCGTTGTGGGCTTGTAATATGTCT- GCTATACCAGTGCCTTTTTACATTATATTACACATGTATGTAAAGTTGTAAATATTTAATCTTAACCACCTTTTGCTAGGTGGCAGATGTGAATGTTT- GATGATTTGGGCTTGAGGTTTCATTTACAAAGGTATTACATTTGTTTTTTATACTGAAACTCAACTCTAGTTTGGCTGACCAAATGCGCCATTATAT- TACTCAATCATGGAAAACATGCATGAGCTTACCAGTGCATGGTTCGAAATATTTAATAACGTACTACTCTGTTGTGATTCCAGTTTTTATTAATAT- GCTTTTTTTAATCGATAAATGTATTTTCCTTGTCAACAGTGTTTTGGCACATGAACACCTTGGGGGGCTGAAATCAACATGTTATACATTGTAAAG- GATCTGTTAGTAGATTTTTTTTTTTAAGAACAAACTTCTACTATGTCATATGGAAAGCCAAAAGTACTTGGATATTTATATTTTGGGATGGGAGGAG- TAATATGGAAATTCGCATTATATATGCACATTGCACAGCTGCTTATCTGCCAGAACAATTCCATCTTCAGTTGGCTGGAGTCTGGAGGCATTTTTC- GTACAGTACCATGCATCTTGCATGTGTATCGTTGCTCGTTGGGCATTGCAGCGTTGACCTGTATATTTCCTACAGTTTTTCAGCACGCCCCAC- TGATCCCTCTAGTCTGTACGAACTAACGAACAGTTTATTATCGCTGCATTTGTGTACTGTACAAAACAAGCAAGTGGTGATCCCA

ST	ACTCGTCTTCTTCCTCTCGCCGCCGCAGCGCGCGGCCGCCACTCCATCCACCACAATGGCGAACCCCACGTCCCACCGCCTCCTCCTCGCC-GCGTAGCCCTATCGGCGAGCCGCTGCCGGTGGCTTCTCCGCATCCCCGCGCCGTAGAGTCTCCGCCGCCGCCGTGTGCTCCGCAACATTG-GGGTCGAGGGTTTTCGATTCGGCGAAGGGGGGGGGGGGGGGGAAAGCTCGTCCCGTGGGAGGCGCCATGGGGAGCGAGGCGAAGGA-GATGAAGTACAGGAGAAGGGCTAGGGTTCCAGAGCCCTTCGATTACGGGCAATGCGGCGGCGACCGGAGTGGCGTGCTGGATTGGGGCG-CGCTCAAGGAGAACCCCGTGGAGCTGCTCCGGAAGCTCGACGAGCTGCGGGACCACATCACCAGGTCCTGCGAGATCACCGACCAGCCG-CGGGAGCGCCACCGCATGAGCCGCCGCACGGCCTCGCTGCGCCCCTCGCACGCCGAGCCGCCGCCGCTAGGCCGTGGGCCGGAGCAC-TACCGCTCACGCTACACGGGAAGGTACGGGTCTGGCTTTCCGCACAGTCCGAATGACCAGCTGCATCGGTCTATGCATAGAGATAGGTATGA-GAGGCAGCCGAGTGGGCGGTTCCGGCAATGGCCGGAGAGGCAGTGGGAGAATTCTGGGTATCTTGGGGGGAATCACCATCAGAGCACCT-GCCAGTGTGCACAGTGCCTTCATGGCCAGAGAGCTGTGATGCAGGAGGAGCACATTCCCATGACGAGGTACTTTGCAGGCCAGCAGGGG-TCTCACCTGTTCGATAGGTCCCCATCAGTCTCATCGGAGCTTGACCGGAGGTCTGTCGCTTCGTCATTGTACTCTCATTTCTCGGTGTCGAA-GAGGAGGACAGAGTTTTTCAGGAAGAAGGCAGAGAGTTTCTGTCGTCCTGTGAGAGGCGCCGCTCCTTTCGTTGTATGCAGTTCTTGTAAC-CAGCTGCTGCAGCTGCCTCCTGGGAAATGCACAGCTCGGAAGCAGATTCAAGTTCGGTGTGGTTCGTGCTCGGAGATCGTTAGTTTCAAGCT-TAAGGAAGTGAAAATTCATCCTTTGGTTGCGCCGACATCTTTCCCTGCATCAAAAACAGTGGGGAGTTCCAGTCGTCAGGTTAACAAAAGTTT-TGGGTGGTACCAACATCAGGATGAAGGAAATTCTAGTTTCCATAAGCTGCAAGCACAAGAGAGATGGCAGCAGAACAAGGATCTTGCAGACA-ATATTTCTGTGTCTTCTACTTCTAGTTATGACAGAATAGACAAAGAGTGTGGATCAAACAGGAGCAGTCAGTTACTATCAGTATCTGTTAGAAG-GTCCAGACTTGCAAATATCCCAAAGGATATACTATGTCAAGGAGATGCATACAGTCAAGTAGAAACTTCAGCATTTAATACAGGCAATCTACAG-GCCCCAGTTATAGAGGACAAATGTGTCGATCCATTTTCCAGTCGTCTAAAAGATTGTAGTGGTGGAGATCGCACAAGCAAGGAATGTAGCCTA-AATATCATGGCAGATTCTGTTGATGCTAATGTTAGGGATGAAAGATCGGACGTGACATATGAGCAGAACAGCAAAGATCATAAAGAAGGATTTG-GAGAAGAAACTGTCAGCAGTAGACATGAGCAAAAGCTCAAAGAAAGTACAAGTGGCTTTTGTGATGATGGAAGCATGGGTAACATAGACAAGC-TGACGGCTGATAATGATGATACCAGCAGCCTTGAAGATGGAGACGTGAGCAAAAAGTATGAAGAAAAAATCAAACAAGATGACAACAACTTTCA-ATCAGAATTCATTACTGAGCGATATAGCAAATGCAGCAAAGAAGATAATAACAGTGTCATTCAAGTTGAAACTATAGCTACGATATGCAAACAAGA-TGACTTAGATGATTGTTATAGTGAATTGCTCTCCCCAAATTCTGAACATGCCATAATGCCATCGAAGCTTGAGTCATCAGTTAATGAGCGCACA-AATTCTAGTTCTCGTGTTTCTTCTGAGGCTGAACTAGACGAAGTTCAGTCTGCAGCTACTAAGAACGGGGATTCAAAGTTTTTTGCTGGTTTTT-TGAAGAAAGGTTTGAAGGACATTTCTTTGTTCAATCAATCAGTGGACAGTGCTAAGGTTTCAATCAATGGCCATTCAATCTCCGAACGTGTCC-TACGGAAGGCCGAGAAGAAGTCTGGTCCAGTTGGCCCTGGTTCATATTGGTATGACTACCGCGCTGGATTTTGGGGTGTCATGGGACATGAG-TGTAGTGGCATCATCCCTCCATTTATAAAGGAATTTAATTATCCAATGCCGAAAAATTGTGCCGGTGGGAACACAGGAGTTATTGTCAATGGTA-GAGAACTTCATCAGAAAGACTTTGAATTGCTTGCAGGAAGAGGTCTCCCACGGATATCTGGAAAATCATATTCTGTTGAGATCAATGGAAATGT-GATTGATGAAACAACTGGCAAAAAATTACGAAAGCTTGGAAAGCTTGCTCCTACAGTTGAGAAGTTAAAACGTGGATTGGCATGCACGTCCC-TGAAGAGAAAAGTTAGCTCTCTGGCTCTGATGATGTTCCCTTGCTTTGCAAAAACTCATTTCCACGTAAAAGCTGTCAAGGGAGTTAAATTG-GATTGTGGCATGGAAGAAGCCTTTAGATTGTACTTCTTGAAGCTTGAGTGTGCAGCAACAGTAATACACGAGTACCTAAGACGACCTGTTTTT-TATATATATTTACCACTTACTCATTCAAAGTTGACAAATTGTCGTACAAATTCCATTACTTCATTTGAAGCATTCTTTAGGTATATATGACCTAAGCAT-GCGAACAGCTCTAAGTTCCTCATGTGTCTGGGTATGTTGTTTCCTGTCCCCTGCATATGTTGTTGATACAGAGCAATAGAGTATAGACAATATTC-TGGGTGTGTATCGTTGTGGGCTTGTAATATGTCTGCTATACCAGTGCCTTTTTACATTATATTACACATGTATGTAAAGTTGTAAATATTTAATCTTA-ACCACCTTTTGCTAGGTGGCAGATGTGAATGTTTGATGATTTGGGCTTGAGGTTTCATTTACAAAG
SCDS	ATGGGGAGCGAGGCGAAGGAGATGAAGTACAGGAGAAGGGCTAGGGTTCCAGAGCCCTTCGATTACGGGCAATGCGGCGGCGACCGGAGT-GGCGTGCTGGATTGGGGCGCGCTCAAGGAGAACCCCGTGGAGCTGCTCCGGAAGCTCGACGAGCTGCGGGACCACATCACCAGGTCCTG-CGAGATCACCGACCAGCCGCGGGAGCGCCACCGCATGAGCCGCCGCACGGCCTCGCTGCGCCCCTCGCACGCCGAGCCGCCGCCGCTA-GGCCGTGGGCCGGAGCACTACCGCTCACGCTACACGGGAAGGTACGGGTCTGGCTTTCCGCACAGTCCGAATGACCAGCTGCATCGGTCTA-TGCATAGAGATAGGTATGAGAGGCAGCCGAGTGGGCGGTTCCGGCAATGGCCGGAGAGGCAGTGGGAGAATTCTGGGTATCTTGGGGGGA-ATCACCATCAGAGCACCTGCCAGTGTGCACAGTGCCTTCATGGCCAGAGAGCTGTGATGCAGGAGGAGCACATTCCCATGACGAGGTACTT-TGCAGGCCAGCAGGGGTCTCACCTGTTCGATAGGTCCCCATCAGTCTCATCGGAGCTTGACCGGAGGTCTGTCGCTTCGTCATTGTACTCT-CATTTCTCGGTGTCGAAGAGGAGGACAGAGTTTTTCAGGAAGAAGGCAGAGAGTTTCTGTCGTCCTGTGAGAGGCGCCGCTCCTTTCGTTG-TATGCAGTTCTTGTAACCAGCTGCTGCAGCTGCCTCCTGGGAAATGCACAGCTCGGAAGCAGATTCAAGTTCGGTGTGGTTCGTGCTCGGA-GATCGTTAGTTTCAAGCTTAAGGAAGTGAAAATTCATCCTTTGGTTGCGCCGACATCTTTCCCTGCATCAAAAACAGTGGGGAGTTCCAGTCG-TCAGGTTAACAAAAGTTTTGGGTGGTACCAACATCAGGATGAAGGAAATTCTAGTTTCCATAAGCTGCAAGCACAAGAGAGATGGCAGCAGAA-CAAGGATCTTGCAGACAATATTTCTGTGTCTTCTACTTCTAGTTATGACAGAATAGACAAAGAGTGTGGATCAAACAGGAGCAGTCAGTTACTAT-CAGTATCTGTTAGAAGGTCCAGACTTGCAAATATCCCAAAGGATATACTATGTCAAGGAGATGCATACAGTCAAGTAGAAACTTCAGCATTTAATA-CAGGCAATCTACAGGCCCCAGTTATAGAGGACAAATGTGTCGATCCATTTTCCAGTCGTCTAAAAGATTGTAGTGGTGGAGATCGCACAAGCA-AGGAATGTAGCCTAAATATCATGGCAGATTCTGTTGATGCTAATGTTAGGGATGAAAGATCGGACGTGACATATGAGCAGAACAGCAAAGATCA-TAAAGAAGGATTTGGAGAAGAAACTGTCAGCAGTAGACATGAGCAAAAGCTCAAAGAAAGTACAAGTGGCTTTTGTGATGATGGAAGCATGG-GTAACATAGACAAGCTGACGGCTGATAATGATGATACCAGCAGCCTTGAAGATGGAGACGTGAGCAAAAAGTATGAAGAAAAAATCAAACAA-GATGACAACAACTTTCAATCAGAATTCATTACTGAGCGATATAGCAAATGCAGCAAAGAAGATAATAACAGTGTCATTCAAGTTGAAACTATAGC-TACGATATGCAAACAAGATGACTTAGATGATTGTTATAGTGAATTGCTCTCCCCAAATTCTGAACATGCCATAATGCCATCGAAGCTTGAGTCAT-CAGTTAATGAGCGCACAAATTCTAGTTCTCGTGTTTCTTCTGAGGCTGAACTAGACGAAGTTCAGTCTGCAGCTACTAAGAACGGGGATTCAA-AGTTTTTTGCTGGTTTTTTGAAGAAAGGTTTGAAGGACATTTCTTTGTTCAATCAATCAGTGGACAGTGCTAAGGTTTCAATCAATGGCCATTCA-ATCTCCGAACGTGTCCTACGGAAGGCCGAGAAGAAGTCTGGTCCAGTTGGCCCTGGTTCATATTGGTATGACTACCGCGCTGGATTTTGGGG-TGTCATGGGACATGAGTGTAGTGGCATCATCCCTCCATTTATAAAGGAATTTAATTATCCAATGCCGAAAAATTGTGCCGGTGGGAACACAGGA-GTTATTGTCAATGGTAGAGAACTTCATCAGAAAGACTTTGAATTGCTTGCAGGAAGAGGTCTCCCACGGATATCTGGAAAATCATATTCTGTTGA-GATCAATGGAAATGTGATTGATGAAACAACTGGCAAAAAATTACGAAAGCTTGGAAAGCTTGCTCCTACAGTTGAGAAGTTAAAACGTGGATTT-GGCATGCACGTCCCTGAAGAGAAAAGTTAG
SP	MGSEAKEMKYRRRARVPEPFDYGQCGGDRSGVLDWGALKENPVELLRKLDELRDHITRSCEITDQPRERHRMSRRTASLRPSHAEPPPLGR-GPEHYRSRYTGRYGSGFPHSPNDQLHRSMHRDRYERQPSGRFRQWPERQWENSGYLGGNHHQSTCQCAQCLHGQRAVMQEEHIPMTRYFA-GQQGSHLFDRSPSVSSELDRRSVASSLYSHFSVSKRRTEFFRKKAESFCRPVRGAAPFVVCSSCNQLLQLPPGKCTARKQIQVRCGSCSEIVS-FKLKEVKIHPLVAPTSFPASKTVGSSSRQVNKSFGWYQHQDEGNSSFHKLQAQERWQQNKDLADNISVSSTSSYDRIDKECGSNRSSQLLSVS-VRRSRLANIPKDILCQGDAYSQVETSAFNTGNLQAPVIEDKCVDPFSSRLKDCSGGDRTSKECSLNIMADSVDANVRDERSDVTYEQNSKDHKE-GFGEETVSSRHEQKLKESTSGFCDDGSMGNIDKLTADNDDTSSLEDGDVSKKYEEKIKQDDNNFQSEFITERYSKCSKEDNNSVIQVETIATICK-QDDLDDCYSELLSPNSEHAIMPSKLESSVNERTNSSSRVSSEAELDEVQSAATKNGDSKFFAGFLKKGLKDISLFNQSVDSAKVSINGHSISERVL-RKAEKKSGPVGPGSYWYDYRAGFWGVMGHECSGIIPPFIKEFNYPMPKNCAGGNTGVIVNGRELHQKDFELLAGRGLPRISGKSYSVEINGNVI-DETTGKKLRKLGKLAPTVEKLKRGFGMHVPEEKS*
FG	*extra-large G-protein-related, putative, expressed*

Tabela 6. Informações genômicas do gene LOC_Os05g50910, identificado pelo marcador SNP S05_29213630

O complexo de proteína G é composta por subunidades Gα, Gβ e Gγ e é um dos principais transdutores de sinais em células eucarióticas, relacionados a diversos processos

biológicos (PANDEY, 2009). Essas proteínas são encontradas em diversas plantas, sendo relacionadas a regulação da divisão celular, formação de órgãos, adaptação das plantas a estresses abióticos e respostas à estímulos abióticos (URANO; JONES, 2014).

Em arroz, as funções proteínas G extragrandes (*rice extra-large* G ou XLG) foram caracterizadas para resistência a doenças (Funções de *OsXLG1, OsXLG2 e OsXLG3*). Os mutantes de arroz com *OsXLG2 e OsXLG3* mostraram resistência ao patógeno *Magnaporthe oryzae*, enquanto os mutantes de *OsXLG1* conferiram resistência ao patógeno *Xanthomonas oryzae* pv *oryzae* (ZHAO *et al.,* 2022).

O marcador SNP S09_18281732, localizado no cromossomo nove e na posição 18281732 (Figura 7), foi identificado no gene LOC_Os09g30070 (Tabela 5) cuja função está associada ao domínio de ligação ARK.

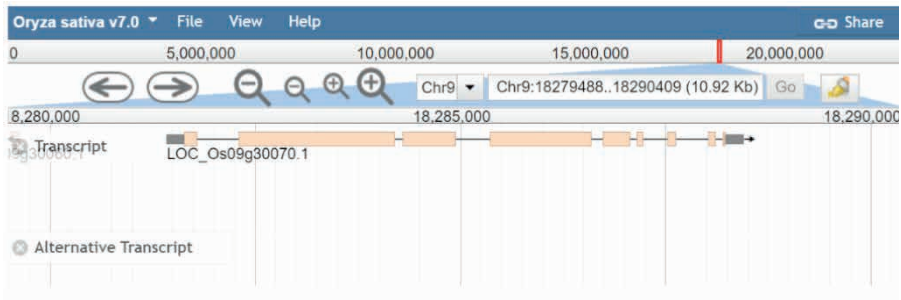

Figura 5. Identificação da posição do gene LOC_Os09g30070 no banco de dados genômicos

Concomitantemente, o marcador SNP S10_15699535 (Figura 6), localizado no cromossomo 10, posição 15699535, e no gene LOC_Os10g30210 (Tabela 8) foi associado a função geral de expressão de proteínas.

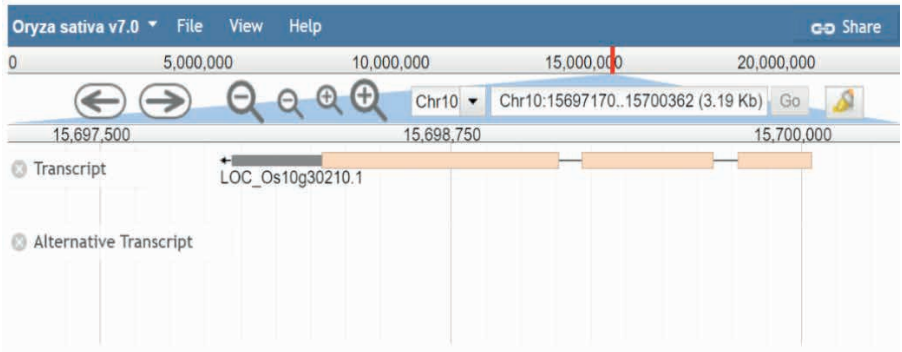

Figura 6. Identificação da posição do gene LOC_Os10g30210 no banco de dados genômicos

LOC_Os09g30070 (S09_18281732)	
SG	ACACATCCTCCGCAACGCAGCGGATTCGCCACAGATCTCTCGCCCCACGTCGCCGCCCTCCTCTCTCCAAATTCAAAAATTCCCGATCCCCTCC- TCTCCTCCGCCATCTCTCTCAAACTCAACGCATCGGCCTCCCCCTCGCCCGCCTCCTCGAATGCCCAAACCCTAGCCCGATCCAGCCGCCGAC- CGACCCACCTCGAGGGAGCGCC ATGGAGGAGCTGTTCATGCAGGTGTTCGAGCGGCGGGACTGGGTGGCGGCGCAGCTGCGGCAGCAGGT- GGAGTCCTACGACCAGTCCCTCGCCTGCGCGCTCCTCGCCGCCGGCCGCCCCCCCCCGCCGTGGCTCCTCCCGTCGCGCCCCGCTGCGCCG- CAGGGTGAGTTGCTTGCTGCTGCTGCTGCCCCCTTTTCCTCTCGCTCGGTTCCATTTTCCGTGGGGTTTGGGTTTGCGATTGAAGCTAGGCTTGG- GCAAGTTGATGGGATTTTTTGGTAGATTTAAGGTGGCGCAGGCTTAATTTTACCTATGCTAGCTACCATTTCGGTTTAGTTTAGCGTTGCGAACATTT- CGTACTGGATTAGTCTGGTTAGCGTTCTTATCCGTTGCATTTTGAAAGTTTTGGGCCTCGTAGTGCACAGATTTAGAGCTTGTTTGCTAAAATTTGG- CAGTTAAAATTGCCGAAACAGCAGAAGGTCTCCAATTTCTTTTAATTATAAACCGAGGGAAAGTAGTTTGAACCAAGTTGTGTTTTCATTGGAAGGA- ATATAAGAGTTTTTTATACAGTAGTATATTTCGGGCCATGCATGTTGTTCGTTTTTGTTCTTATGCACTTTTCTGTTATCTTAAATTTGTTTGAATGCTGA- AATGCCTCGTTGACATGCAGGGCTGAATGGCAAACCAGCGCCATCGGAGTTTGTTTTCACTGGAAGCCATATTACCACACCAGCAATCAATAGAAC- TGTCTATCAGCCTTCAGCTGTTCCTAGTACTTCCTTAAGAAATGTGGGGCTTCCAAGTGGATATAGCCATCTATGGACTGCTTGTAATTCTCTGGATAC- TGACCAACACCAAGAGGTCCAGCAAGAGCAGACTAAAGTTAATGAGGAGTTTGTCAACACTAGAGCTGAAGCCAACATGTTTTCAAGGATTCAAC- GTTCCAGGTCAAGGCAAAGGAATATCGAAGACCGCTTACGTGAAAGGGATGAAGCTGCAAATGGTGGAAGTAGTGATGGTTTGCAAGACAGGAT- GGAGAGGTCTAAGATTGCTGGCGTGCGGCTAAACAGGACCACCACATCATCATCCTCTGAACCATGTGGTGGTGATGCAAACAATTCTGGAGCAA- CACACCCTTTTCGAGGTCAGGAGAATGATATTTATACCAACAAGAGGAACTCAGTTGAATTCTTAAAGTGCAGTAAAGAGGGTGGTCTTGGAAGTG- GAGTTCACTTAGATTGTTCCCCATCTCTGGTTCTAGAAAACAAGATTGTTAGTTCTGACAGTCTCTTCAAGGTTCCTAATGACTGCTCCGCTAGAGAT- TCATCAAGAACACAAGTTGCAGATAGCGTGTGCCATCCCCTCCCTGAAACTCATCTATTTGTTGAGCCCAAAATTCTTCAATTTGAAGGTGTTGAA- TCAGTATGCATGAATTTTTCAAGTGAAAAAATGGGGCAGCCACTGGAAAGTGCCCATCTTGATCTTGCTGAGGCACATCCTTTAAATGAAGACCCA- TCTTCTACTGGCTGTTATCATGTACCTTGTTCCGTGGGAAGCTCATTGGTTGATGGTGTTGAATTGGGACTTTTAAGCACTGACAGTGCAACATTGA- AGCAGCACCTTCAATGTGGCAGTCCTGATCTTAGTCCCACACATTCTCGAAACAAGGACCCGTGTCCAACCATTTCCTCTGAAGTACCTAATTACA- CGAGTGAACCATTAGTTGAGCAGGATACATATTGTAATCCTGAAATCAATTCTCTGGAAGGACCATGTTCTAAGGTCAGCCAACTGCTTGAAAAGGA- AGAAACGAAGGCATGCCCCGATGCCAATCCACTGCTTAAAACAGATGCATTGCATACCATAGGAAGCACCGAAAGAATAAGAAATCTTGCCTCTCG- GAACTCCACTCCTCTTGAACAACGGAGCTCTGATCCTCATGTCTTGCCTTGTCAACGCAGTAGATCTGTGCAACCAGCTGACAGTAGTTCTCGGCC- TCCTCTATCAACAGGAATATTGCCGGACAGTTTGTTGGAAGCGGTTGGTTTAGACCACCTTCCGCATTCAAATGATACAAATAGCCAATGCTCCCCA- TCTAGGTCTGCTGCAAGCCCAGATCTGCTTCCACTTCGACTCGTTAATTCTGGTGATGTTTATCAACCGAGCTTCTCTTGTTGTAAATCACAAAATAA- CAAAGATTCCAATGGTTGTGCTGTTGAGGGCACCACTGTATCCATAGAGAAACCGCCGTCCCAGGAGCAATATTTATTGGATAGACCACCTATGGA- ACTCAATGGTTTTGCAGATGAGGATACTCCCTTGGGCCACACTCTGGGTACACATAATGAGATGCTGAAAGGAAGGAAGGCAGACGACTTGGTGA- ACTGCCATTCTGGAAAGCTCAACAGTTCTCAGAAAAAACCCAAGGGTTTGACTGAAGCATCTGGTTTTTCTTCCGGAAAGAATGAGTCTGCAGGG- CAAAAGGTTGAAAGCAATATTTCAACTGGTGTGATGCACACAACAGAAAGGAGCAGGGGGTTTTGTGCCATGAACTGCACGGAAGATCTCCAACA- AGATGGTACTGCGCTTATTACTTGTAGGAATTATCTTAATTTGTCTTTTGTAGCAACTGCCACTCTGGCACGTGCTTACATAGTTTCTAATCTTGCAG GAACAGAACAAGAAACCTCTCCTTTTGATAATGCTGTTCAAATAAATGCTAATCGATGCACCGCAGATAACAATAAGCAAATCAAATCTTTACGACCT- TCTGTTCGATACTCTTTACGTAGCTTGATGTCACATGAGAAAATTAATCTGCTCCAGTCAGAAGGAAGAAGTGCTGCTTGTGGCCAGAAGAGGTCA- GATGCAGATGGAGTTCAAGTAAACGGTGGCCCATCATCTAAGAGAAGAAGAATAAAGCGTCAATCAAATGCTGCTCTCTCCAGCTCTCCTAACACA- AATTCATTGTCTGTGGTCCACCAAGTTGATATTGACAATCATGTGTTACCCCTTGGAAATTTCTCAGGGAAGTCTCAGCCTTCAGGCCGTTACTTCT- TAAGGGATTTAGGGTCCTCTGGGAGCATGTCTCTCAAGTCAGAGGAGAGGAATGCTGTGAGTCACGGCAACATATCAGTATCTAGTATCCATAATA- AAACCAGCAGTTCTCCTGAAAGATATAATAAGGCTAGTTTAGATAATGAAAATGGGAATTCTCCAGGACAGTTACAAAATACATTAGATGTTGTGAAA- ACAACCGCTGCTTTGCCTAGCTGCTATGGTACCTTAATTGATAATGAGAAATCATGCGCAGAAGAGGTACGTGTTCATCAAGTGCTCTGTTTCCTT- TCTGTTCTGAAATCTTACAAATTTTAGTTCACCTGAAGAGCTCAATCATACTCTTCTATTGAGATAATATGCAGCTAAGTACTGAACTAGTGATGCTCA- GACTGGCCGTAGGATCTGTTATACTTTTACTTTCTAATAAATCTTGAATATACAGCCATGTAGCCTGTCCATTTTGTTGTTTTGGAGTAAGTGATATGG- GTACAATTAGATCATAACTAAATTTTGCATACAGCAGGGCAATGGTAATGTAGTTTTGCATGTAATATTGCACTCATTTTCATAAAAAATAATTTATTAG- CATTTTTAGCCAGTATGTTAGGCGATGCTCACTTTTTTGGTTGTGAATCAGAATTTAATGCACTGAAAAATCGTCTTTGTTGGCAGGAAAATCCTTGCT- TAGAAGGTAAACACGCAAATGATACTTGTTCGAGTGTGGTACACCAGCAAATGACTCTTCAGATTGATAACATTGCATCTCAGAGTGTAATATTAAAT- TCAGAAAATTATTCTAGAGAGAACTCAATTACTATATCTGCAAGTTATGTTTCGGATCAAAATGGTGATCAAGCACATGCACCAAGTGCATTGGTTCG- TGAGAATTTAAGTTATGGCTCCAGTGTAGAACTTGACAGGAGATGTAAAAGTAATGGTTCAAAGGGAAGTTTGTTGTCTGGTGCTGCTATTACTATG- CAAGATGGGGATGAATCTGTTGACTGTGATGCCACAATGCCGGAATTCGAGCGATTTGATGTTCCTATCCAATTTGACAGCCCATGTGCTGAGACA- AAGACATCTGAAGCCCTTTGTGAGTCCAGAAAATTAGTCACCCTCAGTTCAAAGTTCTCAAATTATGACACTAACACAGAAAGTGGTGTGAGCCACC- TATTGTCAGCCATGTCTGGGAAGCCTATAAACTTCCCTGATGATTTGCAACAATACAGAGCAAATAATGATAGAAGCATAACTGACATTTTTGGAGCA- TGTGGATTGGGACTTGATGATTCCTTTTCTATTTATGATGTCACAGCTTCATGTTCTTCAAATGGTAGCAGTGCAAAAGAGAACAATGACAATCCATT- GACCCCATCAGTTGAAAAATATGGCCTGGGGAAACTGTCAGCTAGATCTGGATCTAGTTCAGAACATATGGGCTCTATTCCTGAACTAGAATGCTTC- CGAATTGATGAACATAGCAGCATCGCAGAAGAAAATGAATACCAAGGGATGTTACATGGATCTGCTGGTTTGAGTTACTCGCATCAGTTGCCATCT- GGCAGAAAAGCACTTCAGGATATCACTGGATTATGTCAAAACACTGTAAATTCTGCCTCTCTTTCCTCAATATTCTTGGATACAGGTAATGAGCTCA- ATCATCAGACAGATCTCATAAATGGTCATGCCAACGATAAGCCAAAAAACAGTCTTGCTGCTTCAACCAAGAGAGAGAGAAAAATATCTGATTCCCT- TCATCCTAGATTAAGGAGGACAGAACTACATAACAGAAATGGAAGACACCAGAGTGAAGCCAATATTGACAAACAATCTAAGCCTAGCAATATTGT- TGCCAATGTGACATCATTTATACCTCTTGTAAAACCAAAGTTGCAACCTACAACAGCATGTGGTAAGTGCTAGTACTTTTAATATCATTGTTTATTTA- AAAAAATGTTAGTCCATTTGGCTAGTGATGCTTTGAGCTTAAAAAGAATTTGAATGGTGCTCATATATAGCTTTCATAAATTGGACTTCTGTCTTGTAG TGAAAAAAGATGTCAGAGTGAAGGCACTTGAGGCAGCTGAAGCTGCAAAACGCCTTGAAGAAAAGAAACAAAATGAACGTGAAATGCGCAAAGCAGCTGCAAA- ACTAGAGCGTGAAAGACTGAAGCAAGAGAAAGAATTAAAGCAAAAACAGGAGGAACAAAAGAAGAAGAGAGATGCCGATGTGGCTGCTAAAAAACGGCAAAGA- GGAGAGGAGGAAAGGAAGGAAAAGCAGAGAAAAAGAAAATGCACTGAGGAGGCTCGGAAACAACAGAAGCAGCCTACAGAAAAAAGTCTTGCTGTGAATGA- TGAGAAAGATGTTTGTCGAAAAACATCTGTGAGTAGAAGCAAAATATTTTATTTCTATCAGTGTTTCCTAGAACATGTTTTTGTTACTGAATACATTAATTATTGCAG GATAATATAGAGCTGACAAAACCTGATGGGAGAACTACCGAACCTGCTATGACCAATATTCCAAACAGTCTTGAAGAGGTACGCTTTGTT- TATTTAACTTCGTTGTGAAAGAGGGGCATAGATAGAAATCTGTATTTATTGGTACTGGTCAATTATTGTGTATATTACTGTGTAACTGATGTCG- TAATTCATAAGCCTTAGAATGTGGGCAAAATGTTTAGATACACACACCAATAGTTTATTTGCAGATGGAGCAAAACAACTAAATGTTGCCAA- GTTGACCCGTTTGCTCTTCATAGCTTTTGTTCTAGAGTCAACAATAGTTCTAACAGTGTATTACTTCCTGTGTTTCTGAAAACTTCTTTCAG TCATACCAAATGTCTCCCTATAAGGATTCTGATGAAGAGGATGATGATGATTTTGAACATGAGCAAGAATCCAGACGTAGGAGGAAATTTATTCCTTC- GTGGGCTCG GTATGTATCCATCTAATCAAATCCCCTGTTCAAAATTGCAGCCTTTATTTGAGAAAAAGTTGTGTGTTTGATGTGATGCACTGATGCC- TAATGTGAAAGCTCATCACTTGTGGGAAACAAATGTAGCTTCCGTATTATGTCAAAGATTACATGATAGTCTTTTGAGCACGATGTGAAGTGACTATG- GAAGCTTCTTTACATAAATGTTAAACATATTTTCTATTTTAACACTGATTAGCATATTGTTGCTATGGAAGCTTCTTTACATAAATGTTGAACATATTTTC- TATTTTAACAACTTGCATACTTCTTTTTGCTGCTTGTTGCTTCTGTGTTCATAATGAAGTACTGGTCCTCTAGCTTTTCTATTTACAAAGTTGTCTGATT- TGCAGGAAAGAAAACTTGGATAAACTCTTGTTGTCCAATCAGAGTTTAGACCCTAGAGAACTTTTGCACCAAAAGTGCTCCTTCAACTTATCTGATG GTAACCTCCAGCTCATCTCTGTCATAAATTTATTTCGGTTATTACTGCTGCTGAATTTGTCTTCTCTTGCCTGTTCTCTCCAG TTCTTTCAGTTCACA- TACCACAACGTGGCTTCAGATAA GTTTCCATTTGACGGACTATCGGCAATACGTCATTGCCAAATTACATCATCAACAGTTGTGCATATGCTTGGACG- GCATTTTCTTGTATGCTAGAAATTAATCAACGTAGCAGAAAATTTGTTTCTTGTGGAATATTAGAATTTAGCAGGACATGTGTTGCCACATTTGTATACA- TATTTGGGCATAACCAATTTGTACCAGCCGAGTTATGTTTATG

ST	ACACATCCTCCGCAACGCAGCGGATTCGCCACAGATCTCTCGCCCCACGTCGCCGCCCTCCTCTCTCCAAATTCAAAAATTCCCGATCCCCTCCTC- TCCTCCGCCATCTCTCTCAAACTCAACGCATCGGCCTCCCCCTCGCCCGCCTCCTCGAATGCCCAAACCCTAGCCCGATCCAGCCGCCGACCGACC- CACCTCGAGGGAGCGCC ATGGAGGAGCTGTTCATGCAGGTGTTCGAGCGGCGGGACTGGGTGGCGGCGCAGCTGCGGCAGCAGGTGGAGTCC- TACGACCAGTCCCTCGCCTGCGCGCTCCTCGCCGCCGGCCGCCCCCCCCCGCCGTGGCTCCTCCCGTCGCGCCCCGCTGCGCCGCAGGGGCT- GAATGGCAAACCAGCGCCATCGGAGTTTGTTTTCACTGGAAGCCATATTACCACACCAGCAATCAATAGAACTGTCTATCAGCCTTCAGCTGTTCCTA- GTACTTCCTTAAGAAATGTGGGGCTTCCAAGTGGATATAGCCATCTATGGACTGCTTGTAATTCTCTGGATACTGACCAACACCAAGAGGTCCAGCAA- GAGCAGACTAAAGTTAATGAGGAGTTTGTCAACACTAGAGCTGAAGCCAACATGTTTTCAAGGATTCAACGTTCCAGGTCAAGGCAAAGGAATATCGA- AGACCGCTTACGTGAAAGGGATGAAGCTGCAAATGGTGGAAGTAGTGATGGTTTGCAAGACAGGATGGAGAGGTCTAAGATTGCTGGCGTGCGGC- TAAACAGGACCACCACATCATCATCCTCTGAACCATGTGGTGGTGATGCAAACAATTCTGGAGCAACACACCCTTTTCGAGGTCAGGAGAATGATATT- TATACCAACAAGAGGAACTCAGTTGAATTCTTAAAGTGCAGTAAAGAGGGTGGTCTTGGAAGTGGAGTTCACTTAGATTGTTCCCCATCTCTGGTTCTA- GAAAACAAGATTGTTAGTTCTGACAGTCTCTTCAAGGTTCCTAATGACTGCTCCGCTAGAGATTCATCAAGAACACAAGTTGCAGATAGCGTGTGCCA- TCCCCTCCCTGAAACTCATCTATTTGTTGAGCCCAAAATTCTTCAATTTGAAGGTGTTGAATCAGTATGCATGAATTTTTCAAGTGAAAAAATGGGGCA- GCCACTGGAAAGTGCCCATCTTGATCTTGCTGAGGCACATCCTTTAAATGAAGACCCATCTTCTACTGGCTGTTATCATGTACCTTGTTCCGTGGGAA- GCTCATTGGTTGATGGTGTTGAATTGGGACTTTTAAGCACTGACAGTGCAACATTGAAGCAGCACCTTCAATGTGGCAGTCCTGATCTTAGTCCCACA- CATTCTCGAAACAAGGACCCGTGTCCAACCATTTCCTCTGAAGTACCTAATTACACGAGTGAACCATTAGTTGAGCAGGATACATATTGTAATCCTGA- AATCAATTCTCTGGAAGGACCATGTTCTAAGGTCAGCCAACTGCTTGAAAAGGAAGAAACGAAGGCATGCCCCGATGCCAATCCACTGCTTAAAACA- GATGCATTGCATACCATAGGAAGCACCGAAAGAATAAGAAATCTTGCCTCTCGGAACTCCACTCCTCTTGAACAACGGAGCTCTGATCCTCATGTCTT- GCCTTGTCAACGCAGTAGATCTGTGCAACCAGCTGACAGTAGTTCTCGGCCTCCTCTATCAACAGGAATATTGCCGGACAGTTTGTTGGAAGCGGTT- GGTTTAGACCACCTTCCGCATTCAAATGATACAAATAGCCAATGCTCCCCATCTAGGTCTGCTGCAAGCCCAGATCTGCTTCCACTTCGACTCGTTAAT- TCTGGTGATGTTTATCAACCGAGCTTCTCTTGTTGTAAATCACAAAATAACAAAGATTCCAATGGTTGTGCTGTTGAGGGCACCACTGTATCCATAGAGA- AACCGCCGTCCCAGGAGCAATATTTATTGGATAGACCACCTATGGAACTCAATGGTTTTGCAGATGAGGATACTCCCTTGGGCCACACTCTGGGTACA- CATAATGAGATGCTGAAAGGAAAGAAGGCAGACGACTTGGTGAACTGCCATTCTGGAAAGCTCAACAGTTCTCAGAAAAAACCCAAGGGTTTGACT- GAAGCATCTGGTTTTTCTTCCGGAAAGAATGAGTCTGCAGGGCAAAAGGTTGAAAGCAATATTTCAACTGGTGTGATGCACACAACAGAAAGGAGCA- GGGGGTTTTGTGCCATGAACTGCACGGAAGATCTCCAACAAGATGGAACAGAACAAGAAACCTCTCCTTTTGATAATGCTGTTCAAATAAATGCTAAT- CGATGCACCGCAGATAACAATAAGCAAATCAAATCTTTACGACCTTCTGTTCGATACTCTTTACGTAGCTTGATGTCACATGAGAAAATTAATCTGCTC- CAGTCAGAAGGAAGAAGTGCTGCTTGTGGCCAGAAGAGGTCAGATGCAGATGGAGTTCAAGTAAACGGTGGCCCATCATCTAAGAGAAGAAGAATA- AAGCGTCAATCAAATGCTGCTCTCTCCAGCTCTCCTAACACAAATTCATTGTCTGTGGTCCACCAAGTTGATATTGACAATCATGTGTTACCCCTTGGA- AATTTCTCAGGGAAGTCTCAGCCTTCAGGCCGTTACTTCTTAAGGGATTTAGGGTCCTCTGGGAGCATGTCTCTCAAGTCAGAGGAGAGGAATGCTG- TGAGTCACGGCAACATATCAGTATCTAGTATCCATAATAAAACCAGCAGTTCTCCTGAAAGATATAATAAGGCTAGTTTAGATAATGAAAATGGGAATTC- TCCAGGACAGTTACAAAATACATTAGATGTTGTGAAAACAACCGCTGCTTTGCCTAGCTGCTATGGTACCTTAATTGATAATGAGAAATCATGCGCAGA- AGAGGAAAATCCTTGCTTAGAAGGTAAACACGCAAATGATACTTGTTCGAGTGTGGTACACCAGCAAATGACTCTTCAGATTGATAACATTGCATCTCA- GAGTGTAATATTAAATTCAGAAAATTATTCTAGAGAGAACTCAATTACTATATCTGCAAGTTATGTTTCGGATCAAAATGGTGATCAAGCACATGCACCAA- GTGCATTGGTTCGTGAGAATTTAAGTTATGGCTCCAGTGTAGAACTTGACAGGAGATGTAAAAGTAATGGTTCAAAGGGAAGTTTGTTGTCTGGTGCT- GCTATTACTATGCAAGATGGGGATGAATCTGTTGACTGTGATGCCACAATGCCGGAATTCGAGCGATTTGATGTTCCTATCCAATTTGACAGCCCATGT- GCTGAGACAAAGACATCTGAAGCCCTTTGTGAGTCCAGAAAATTAGTCACCCTCAGTTCAAAGTTCTCAAATTATGACACTAACACAGAAAGTGGTGT- GAGCCACCTATTGTCAGCCATGTCTGGGAAGCCTATAAACTTCCCTGATGATTTGCAACAATACAGAGCAAATAATGATAGAAGCATAACTGACATTTTT- GGAGCATGTGGATTGGGACTTGATGATTCCTTTTCTATTTATGATGTCACAGCTTCATGTTCTTCAAATGGTAGCAGTGCAAAAGAGAACAATGACAATC- CATTGACCCCATCAGTTGAAAAATATGGCCTGGGGAAACTGTCAGCTAGATCTGGATCTAGTTCAGAACATATGGGCTCTATTCCTGAACTAGAATGCT- TCCGAATTGATGAACATAGCAGCATCGCAGAAGAAAATGAATACCAAGGGATGTTACATGGATCTGCTGGTTTGAGTTACTCGCATCAGTTGCCATCTG- GCAGAAAAGCACTTCAGGATATCACTGGATTATGTCAAAACACTGTAAATTCTGCCTCTCTTTCCTCAATATTCTTGGATACAGGTAATGAGCTCAATCA- TCAGACAGATCTCATAAATGGTCATGCCAACGATAAGCCAAAAAACAGTCTTGCTGCTTCAACCAAGAGAGAGAGAAAAATATCTGATTCCCTTCATCC- TAGATTAAGGAGGACAGAACTACATAACAGAAATGGAAGACACCAGAGTGAAGCCAATATTGACAAACAATCTAAGCCTAGCAATATTGTTGCCAATGT- GACATCATTTATACCTCTTGTAAAACCAAAGTTGCAACCTACAACAGCATGTGTGAAAAAAGATGTCAGAGTGAAGGCACTTGAGGCAGCTGAAGCTG- CAAAACGCCTTGAAGAAAAGAAACAAAATGAACGTGAAATGCGCAAAGCAGCTGCAAAACTAGAGCGTGAAAGACTGAAGCAAGAGAAAGAATTAAA- GCAAAAACAGGAGGAACAAAAGAAGAAGAGAGATGCCGATGTGGCTGCTAAAAAACGGCAAAGAGGAGAGGAGGAAAGGAAGGAAAAGCAGAGAA- AAAGAAAATGCACTGAGGAGGCTCGGAAACAACAGAAGCAGCCTACAGAAAAAAGTCTTGCTGTGAATGATGAGAAAGATGTTTGTCGAAAAACATCT GATAATATAGAGCTGACAAAACCTGATGGGAGAACTACCGAACCTGCTATGACCAATATTCCAAACAGTCTTGAAGAG TCATACCAAATGTC- TCCCTATAAGGATTCTGATGAAGAGGATGATGATGATTTTGAACATGAGCAAGAATCCAGACGTAGGAGGAAATTTATTCCTTCGTGGGCTCG GAAAGAAAACTTGGATAAACTCTTGTTGTCCAATCAGAGTTTAGACCCTAGAGAACTTTTTGCACAAAAGTGCTCCTTCAACTTATCTGATG TTCTTTCAGTTCACATACCACAACGTGGCTTCAGATAAGTTTCCATTTGACGGACTATCGGCAATACGTCATTGCCAAATTACATCATCAACAGTTGTG- CATATGCTTGGACGGCATTTTCTTGTATGCTAGAAATTAATCAACGTAGCAGAAAATTTGTTTCTTGTGGAATATTAGAATTTAGCAGGACATGTGTTGC- CACATTTGTATACATATTTGGGCATAACCAATTTGTACCAGCCGAGTTATGTTTATG

SCDS	ATGGAGGAGCTGTTCATGCAGGTGTTCGGCGGCGGGACTGGGTGGCGGCGCAGCTGCGGCAGCAGGTGGAGTCCTACGACCAGTCCCTCGCCT- GCGCGCTCCTCGCCGCCGGCCGCCCCCCCCCGCCGTGGCTCCTCCCGTCGCGCCCCGCTGCGCCGCAGGGGCTGAATGGCAAACCAGCGCCA- TCGGAGTTTGTTTTCACTGGAAGCCATATTACCACACCAGCAATCAATAGAACTGTCTATCAGCCTTCAGCTGTTCCTAGTACTTCCTTAAGAAATGTG- GGGCTTCCAAGTGGATATAGCCATCTATGGACTGCTTGTAATTCTCTGGATACTGACCAACACCAAGAGGTCCAGCAAGAGCAGACTAAAGTTAATGA- GGAGTTTGTCAACACTAGAGCTGAAGCCAACATGTTTTCAAGGATTCAACGTTCCAGGTCAAGGCAAAGGAATATCGAAGACCGCTTACGTGAAAGG- GATGAAGCTGCAAATGGTGGAAGTAGTGATGGTTTGCAAGACAGGATGGAGAGGTCTAAGATTGCTGGCGTGCGGCTAAACAGGACCACCACATCA- TCATCCTCTGAACCATGTGGTGGTGATGCAAACAATTCTGGAGCAACACACCCTTTTCGAGGTCAGGAGAATGATATTTATACCAACAAGAGGAACT- CAGTTGAATTCTTAAAGTGCAGTAAAGAGGGTGGTCTTGGAAGTGGAGTTCACTTAGATTGTTCCCCATCTCTGGTTCTAGAAAACAAGATTGTTAGT- TCTGACAGTCTCTTCAAGGTTCCTAATGACTGCTCCGCTAGAGATTCATCAAGAACACAAGTTGCAGATAGCGTGTGCCATCCCCTCCCTGAAACTCA- TCTATTTGTTGAGCCCAAAATTCTTCAATTTGAAGGTGTTGAATCAGTATGCATGAATTTTTCAAGTGTGAAAAAATGGGGCAGCCACTGGAAAGTGCCCA- TCTTGATCTTGCTGAGGCACATCCTTTAAATGAAGACCCATCTTCTACTGGCTGTTATCATGTACCTTGTTCCGTGGGAAGCTCATTGGTTGATGGTGT- TGAATTGGGACTTTTAAGCACTGACAGTGCAACATTGAAGCAGCACCTTCAATGTGGCAGTCCTGATCTTAGTCCCACACATTCTCGAAACAAGGACC- CGTGTCCAACCATTTCCTCTGAAGTACCTAATTACACGAGTGAACCATTAGTTGAGCAGGATACATATTGTAATCCTGAAATCAATTCTCTGGAAGGAC- CATGTTCTAAGGTCAGCCAACTGCTTGAAAAGGAAGAAACGAAGGCATGCCCCGATGCCAATCCACTGCTTAAAACAGATGCATTGCATACCATAGGA- AGCACCGAAAGAATAAGAAATCTTGCCTCTCGGAACTCCACTCCTCTTGAACAACGGAGCTCTGATCCTCATGTCTTGCCTTGTCAACGCAGTAGATC- TGTGCAACCAGCTGACAGTAGTTCTCGGCCTCCTCTATCAACAGGAATATTGCCGGACAGTTTGTTGGAAGCGGTTGGTTTAGACCACCTTCCGCAT- TCAAATGATACAAATAGCCAATGCTCCCCATCTAGGTCTGCTGCAAGCCCAGATCTGCTTCCACTTCGACTCGTTAATTCTGGTGATGTTTATCAACC- GAGCTTCTCTTGTTGTAAATCACAAAATAACAAAGATTCCAATGGTTGTGCTGTTGAGGGCACCACTGTATCCATAGAGAAACCGCCGTCCCAGGAG- CAATATTTATTGGATAGACCACCTATGGAACTCAATGGTTTTGCAGATGAGGATACTCCCTTGGGCCACACTCTGGGTACACATAATGAGATGCTGAAA- GGAAAGAAGGCAGACGACTTGGTGAACTGCCATTCTGGAAAGCTCAACAGTTCTCAGAAAAAACCCAAGGGTTTGACTGAAGCATCTGGTTTTTCT- TCCGGAAAGAATGAGTCTGCAGGGCAAAAGGTTGAAAGCAATATTTCAACTGGTGTGATGCACACAACAGAAAGGAGCAGGGGGTTTTGTGCCAT- GAACTGCACGGAAGATCTCCAACAAGATGGAACAGAACAAGAAACCTCTCCTTTTGATAATGCTGTTCAAATAAATGCTAATCGATGCACCGCAGATA- ACAATAAGCAAATCAAATCTTTACGACCTTCTGTTCGATACTCTTTACGTAGCTTGATGTCACATGAGAAAATTAATCTGCTCCAGTCAGAAGGAAGA- AGTGCTGCTTGTGGCCAGAAGAGGTCAGATGCAGATGGAGTTCAAGTAAACGGTGGCCCATCATCTAAGAGAAGAAGAATAAAGCGTCAATCAAAT- GCTGCTCTCTCCAGCTCTCCTAACACAAATTCATTGTCTGTGGTCCACCAAGTTGATATTGACAATCATGTGTTACCCCTTGGAAATTTCTCAGGGAA- GTCTCAGCCTTCAGGCCGTTACTTCTTAAGGGATTTAGGGTCCTCTGGGAGCATGTCTCTCAAGTCAGAGGAGAGGAATGCTGTGAGTCACGGCAA- CATATCAGTATCTAGTATCCATAATAAAACCAGCAGTTCTCCTGAAAGATATAATAAGGCTAGTTTAGATAATGAAAATGGGAATTCTCCAGGACAGTTA- CAAAATACATTAGATGTTGTGAAAACAACCGCTGCTTTGCCTAGCTGCTATGGTACCTTAATTGATAATGAGAAATCATGCGCAGAAGAGGAAAATCCT- TGCTTAGAAGGTAAACACGCAAATGATACTTGTTCGAGTGTGGTACACCAGCAAATGACTCTTCAGATTGATAACATTGCATCTCAGAGTGTAATATTA- AATTCAGAAAATTATTCTAGAGAGAACTCAATTACTATATCTGCAAGTTATGTTTCGGATCAAAATGGTGATCAAGCACATGCACCAAGTGCATTGGTTC- GTGAGAATTTAAGTTATGGCTCCAGTGTAGAACTTGACAGGAGATGTAAAAGTAATGGTTCAAAGGGAAGTTTGTTGTCTGGTGCTGCTATTACTATG- CAAGATGGGGATGAATCTGTTGACTGTGATGCCACAATGCCGGAATTCGAGCGATTTGATGTTCCTATCCAATTTGACAGCCCATGTGCTGAGACAAA- GACATCTGAAGCCCTTTGTGAGTCCAGAAAATTAGTCACCCTCAGTTCAAAGTTCTCAAATTATGACACTAACACAGAAAGTGGTGTGAGCCACCTAT- TGTCAGCCATGTCTGGGAAGCCTATAAACTTCCCTGATGATTTGCAACAATACAGAGCAAATAATGATAGAAGCATAACTGACATTTTTGGAGCATGTG- GATTGGGACTTGATGATTCCTTTTCTATTTATGATGTCACAGCTTCATGTTCTTCAAATGGTAGCAGTGCAAAAGAGAACAATGACAATCCATTGACCC- CATCAGTTGAAAAATATGGCCTGGGGAAACTGTCAGCTAGATCTGGATCTAGTTCAGAACATATGGGCTCTATTCCTGAACTAGAATGCTTCCGAATT- GATGAACATAGCAGCATCGCAGAAGAAAATGAATACCAAGGGATGTTACATGGATCTGCTGGTTTGAGTTACTCGCATCAGTTGCCATCTGGCAGAAA- AGCACTTCAGGATATCACTGGATTATGTCAAAACACTGTAAATTCTGCCTCTCTTTCCTCAATATTCTTGGATACAGGTAATGAGCTCAATCATCAGACA- GATCTCATAAATGGTCATGCCAACGATAAGCCAAAAAACAGTCTTGCTGCTTCAACCAAGAGAGAGAGAAAAATATCTGATTCCCTTCATCCTAGATTA- AGGAGGACAGAACTACATAACAGAAATGGAAGACACCAGAGTGAAGCCAATATTGACAAACAATCTAAGCCTAGCAATATTGTTGCCAATGTGACAT- CATTTATACCTCTTGTAAAACCAAAGTTGCAACCTACAACAGCATGTGTGAAAAAAGATGTCAGAGTGAAGGCACTTGAGGCAGCTGAAGCTGCAAA- ACGCCTTGAAGAAAAGAAACAAAATGAACGTGAAATGCGCAAAGCAGCTGCAAAACTAGAGCGTGAAAGACTGAAGCAAGAGAAAGAATTAAAGCA- AAAACAGGAGGAACAAAAGAAGAAGAGAGATGCCGATGTGGCTGCTAAAAAACGGCAAAGAGGAGAGGAGGAAAGGAAGGAAAAGCAGAGAAAA- AGAAAATGCACTGAGGAGGCTCGGAAACAACAGAAGCAGCCTACAGAAAAAAGTCTTGCTGTGAATGATGAGAAAGATGTTTGTCGAAAAACATCT GATAATATAGAGCTGACAAAACCTGATGGGAGAACTACCGAACCTGCTATGACCAATATTCCAAACAGTCTTGAAGAG TCATACCAAATGTCTCCCTA- TAAGGATTCTGATGAAGAGGATGATGATGATTTTGAACATGAGCAAGAATCCAGACGTAGGAGGAAATTTATTCCTTCGTGGGCTCG GAAAGAAAACT- TGGATAAACTCTTGTTGTCCAATCAGAGTTTAGACCCTAGAGAACTTTTTGCACAAAAGTGCTCCTTCAACTTATCTGATG TTCTTTCAGTTCACATAC- CACAACGTGGCTTCAGATAA
SP	MEELFMQVFERRDWVAAQLRQQVESYDQSLACALLAAGRPPPPWLLPSRPAAPQ GLNGKPAPSEFVFTGSHITTPAINRTVYQPSAVPST- SLRNVGLPSGYSHLWTACNSLDTDQHQEVQQEQTKVNEEFVNTRAEANMFSRIQRSRSRQRNIEDRLRERDEAANGGSSDGLQDRMER- SKIAGVRLNRTTTSSSSEPCGGDANNSGATHPFRGQENDIYTNKRNSVEFLKCSKEGGLGSGVHLDCSPSLVLENKIVSSDSLFKVPNDC- SARDSSRTQVADSVCHPLPETHLFVEPKILQFEGVESVCMNFSSEKMGQPLESAHLDLAEAHPLNEDPSSTGCYHVPCSVGSSLVDGVEL- GLLSTDSATLKQHLQCGSPDLSPTHSRNKDPCPTISSEVPNYTSEPLVEQDTYCNPEINSLEGPCSKVSQLLEKEETKACPDANPLLKTDALH- TIGSTERIRNLASRNSTPLEQRSSDPHVLPCQRSRSVQPADSSSRPPLSTGILPDSLLEAVGLDHLPHSNDTNSQCSPSRSAASPDLLPLRLV- NSGDVYQPSFSCCKSQNNKDSNGCAVEGTTVSIEKPPSQEQYLLDRPPMELNGFADEDTPLGHTLGTHNEMLKGKKADDLVNCHSGKLNS- SQKKPKGLTEASGFSSGKNESAGQKVESNISTGVMHTTERSRGFCAMNCTEDLQQDGTEQETSPFDNAVQINANRCTADNNKQIKSLRPS- VRYSLRSLMSHEKINLLQSEGRSAACGQKRSDADGVQVNGGPSSKRRRIKRQSNAALSSSPNTNSLSVVHQVDIDNHVLPLGNFSGKSQPS- GRYFLRDLGSSGSMSLKSEERNAVSHGNISVSSIHNKTSSSPERYNKASLDNENGNSPGQLQNTLDVVKTTAALPSCYGTLIDNEKSCAEE ENPCLEGKHANDTCSSVVHQQMTLQIDNIASQSVTKNSENYSRENSITISASYVSDQNGDQAHAPSALVRENLSYGSSVELDRRCKSNGSKGSLLSGAAITM- QDGDESVDCDATMPEFERFDVPIQFDSPCAETKTSEALCESRKLVTLSSKFSNYDTNTESGVSHLLSAMSGKPINFPDDLQQYRANNDRSITDIFGACGLGL- DDSFSIYDVTASCSSNGSSAKENNDNPLTPSVEKYGLGKLSARSGSSSEHMGSIPELECFRIDEHSSIAEENEYQQMLHGSAGLSYSHQLPSGRKALQDITGL- CQNTVNSASLSSIFLDTGNELNHQTDLINGHANDKPKNSLAASTKRERKISDSLHPRLRRTELHNRNGRHQSEANIDKQSKPSNIVANVTSFIPLVKPKLQPTTAC VKKDVRVKALEAAEAAKRLEEKKQNEREMRKAAAKLERERLKQEKELKQKQEEQKKKRDADVAAKKRQRGEEERKEKQRKRKCTEEARKQQKQPTEKS- LAVNDEKDVCRKTS DNIELTKPDGRTTEPAMTNIPNSLEESYQMSPYKDSDEEDDDDDFEHEQESRRRRKFIPSWAR KENLDKLLLSNQSLDPRELFAQK- CSFNLSDV LSVHIPQRGFR*
FG	Inner centromere protein, ARK-binding domain

Tabela 7. Informações genômicas do gene LOC_Os09g30070, identificado pelo marcador SNP S09_18281732.

LOC_Os10g30210 (S10_15699535)	
SG	ATGGCCGCCTCCTCCTCCAACCCCACGCCTCTGCTGCTGCCACCGCCGCCGC- CGCAAGAGCCCGCGCCTCTGTCGCCGCCGCCGCCGCTGCCTACGCCGAAGCC- GATCCCCACCGTCGCCGACAACTTCCGCAGCCTGCTCAGGTCCGGCGAGGCGC- TCCTCCGCTTCGCCTTCCGCGGCAACTCCGGCCAATTGACTCACCGGCATCCGCC- TCCGCCTCGGCCGCCGCCGCAGCAGCAGCATCCCCACCACCACAACCGCCCG GTGCGTACGCCTCGATCGTGAGCTCGAACACACTCGATCAGCAACCATGGCGAT- TACCTTCTTCACTGAGACACTCGCGTTTGCAGGCTGAGATCATGAAGCGGCTGCAGCG- GGAGAAGTTCGCCGACATGATCAAGCACATGGACGGGCACGAGCAGATCGACCGCCT- CGTCGCGCTGTACACGAGCAGCGCCAAGGGGTTCCACCTCCCGGAGCTCCCCGTCC- GGGTGAAGGTCGCACTCGACGCGGCCGGCGCGCTGCTCCTCGTCGACGGCGACGA- GCTCGAGCAGGCGAGGGACAGGCTCGCCAAGGCCAGGAACACGACGGGGCTCGGC- TCCAGGTTCGTGTTCGAGTCGAGCACCCGTGGCGGCAAGGACACCGTCGCCGCGGA- GCTCGCCACCGGGCTCGGGGCGGCGGCGGCGGCGGCGGCCGGCGGGAGGCCACT- GGAGCTCACCCGGCTGCAGTACTGCGCCCACGTCAGCGACCTGCTGTCGATGACCC- TCGTCCCGTTTGGCGCCCAGTGCAACAACTTCTTGCACGGTTCGAGCCTGATCCAG GTAACTTGCCACTGAGCTCGCTGGAGTATTGGCGGTTGCGGCGGCGAGCTTTTGTCT- GACGATGAAACTGTTCGCCGCAGAGTATCCAGAGCAGAGCGTTGTCCGGCGGGCCGC- CGTCGTATTCGGAGCGTCATGACTGCGGCGCCGGCGTGAGCATCAAGGGATCAAGAT- TCAGGGCCTCCATAGCTGAGCTGATCTTTGGTTCTCCAGGCGAACATGGCGGCGGCG- GCGGCGGCGGCAATGGCGATCACGAGGTGCCCAACCGGTTGACGACGTTCGGCAA- GGTGAGCTACGAGACGGCCGACGACATCAAGCTCAGCTTGTCGGGGCTGTGGCAGG- TCCGCTCGCCGGCGTCCCGGTTCAGCGACCTCGGCGCGCTCGCCGTGCCGCTGGG- CAGCCTGAAGACGCGGAGAGCCATCGCTCCTCCTCCGTCGCCGCCGGCGACGACG- CATCCTCCTCCGTCGCCGGACCTGATGGTGCAGGTGCCGGCGCCTCCTGCTCCTCC- CACCCCCATGATGGGTATAGGGTCGACGGTGGCGGTGCAGGGCTCCGTTGCCGCC- GCCGCCGCCGCCGCCACGGCGCCGTCGTCGCACACGGTGGCCGTCATGGTGGACTG- CGACATGTACGACACCTTGAGAGCAGAAGGCTGGGTCGAGATGGAGACGGCGGCGG- CGGCGACTCCGGCGAGGCGGCGCGGGCCGGTGGCGCGGTGGGGAGTCTGCGTGTC- CGACTGCCCGGAGCACGAGCTCGGGTGGGGGGTGCGGATCGGCGGCACGGCGGA- GAGGAACGCGCACCGGCCGCACGTGGAAGGGTTCCTCAGCTTCGACCTCGGCAAGG- GCGGCAGGGTGCAGCCGGGCCTCGTCATCGCCATGGACGGCGACAAGCGGACGCC- GGCGCTGGTGCTGAGGTCGTCGTGGCTCATGTGACGCCGGTGATCAAACTTGCCCGT- TCTTCCCCTTCATGCCGGTGTTTTGGTTTTTTGCTCTGAAATCTGTTGCAGATATAATAA- GCGTACGTGACAAATGCCCTCCTTCAATTTACACCTGAGTGAGATTTCATTTTTCTTTT- GCTTTTTGAGGGATTTTTTTTTTCTACCACCAACATATCGATCGTCGCTTGCTTTAGATAG- CAAAAGGGAACTCGAATTTGAGTAAATTTGATAACACCTTTAACTTTTTTGTCCAAAAGT- GCCCAAGTTAATTGTAATCTGTAGCTCTCAGAAATTTGAGCTAAAGTTGCCATTCCTT

ST	ATGGCCGCCTCCTCCTCCAACCCCACGCCTCTGCTGCTGCCACCGCCGCCGC- CGCAAGAGCCCGCGCCTCTGTCGCCGCCGCCGCCGCTGCCTACGCCGAAGCC- GATCCCCACCGTCGCCGACAACTTCCGCAGCCTGCTCAGGTCCGGCGAGGCGC- TCCTCCGCTTCGCCTTCCGCGGCAACTCCGGCCAATTGACTCACCGGCATCCGCC- TCCGCCTCGGCCGCCGCCGCAGCAGCAGCATCCCCACCACCACAACCGCCCG GCTGAGATCATGAAGCGGCTGCAGCGGGAGAAGTTCGCCGACATGATCAAGCACATG- GACGGGCACGAGCAGATCGACCGCCTCGTCGCGCTGTACACGAGCAGCGCCAAGGG- GTTCCACCTCCCGGAGCTCCCCGTCCGGGTGAAGGTCGCACTCGACGCGGCCGGC- GCGCTGCTCCTCGTCGACGGCGACGAGCTCGAGCAGGCGAGGGACAGGCTCGCCA- AGGCCAGGAACACGACGGGGCTCGGCTCCAGGTTCGTGTTCGAGTCGAGCACCCGT- GGCGGCAAGGACACCGTCGCCGCGGAGCTCGCCACCGGGCTCGGGGCGGCGGCG- GCGGCGGCGGCCGGCGGGAGGCCACTGGAGCTCACCCGGCTGCAGTACTGCGCC- CACGTCAGCGACCTGCTGTCGATGACCCTCGTCCCGTTTGGCGCCCAGTGCAACA- ACTTCTTGCACGGTTCGAGCCTGATCCAGAGTATCCAGAGCAGAGCGTTGTCCGGCG- GGCCGCCGTCGTATTCGGAGCGTCATGACTGCGGCGCCGGCGTGAGCATCAAGGGAT- CAAGATTCAGGGCCTCCATAGCTGAGCTGATCTTTGGTTCTCCAGGCGAACATGGCGG- CGGCGGCGGCGGCGGCAATGGCGATCACGAGGTGCCCAACCGGTTGACGACGTTCG- GCAAGGTGAGCTACGAGACGGCCGACGACATCAAGCTCAGCTTGTCGGGGCTGTGG- CAGGTCCGCTCGCCGGCGTCCCGGTTCAGCGACCTCGGCGCGCTCGCCGTGCCGCT- GGGCAGCCTGAAGACGCGGAGAGCCATCGCTCCTCCTCCGTCGCCGCCGGCGACGA- CGCATCCTCCTCCGTCGCCGGACCTGATGGTGCAGGTGCCGGCGCCTCCTGCTCCTC- CCACCCCCATGATGGGTATAGGGTCGACGGTGGCGGTGCAGGGCTCCGTTGCCGCC- GCCGCCGCCGCCGCCACGGCGCCGTCGTCGCACACGGTGGCCGTCATGGTGGACTG- CGACATGTACGACACCTTGAGAGCAGAAGGCTGGGTCGAGATGGAGACGGCGGCGG- CGGCGACTCCGGCGAGGCGGCGCGGGCCGGTGGCGCGGTGGGGAGTCTGCGTGTC- CGACTGCCCGGAGCACGAGCTCGGGTGGGGGGTGCGGATCGGCGGCACGGCGGA- GAGGAACGCGCACCGGCCGCACGTGGAAGGGTTCCTCAGCTTCGACCTCGGCAAGG- GCGGCAGGGTGCAGCCGGGCCTCGTCATCGCCATGGACGGCGACAAGCGGACGCC- GGCGCTGGTGCTGAGGTCGTCGTGGCTCATGTGACGCCGGTGATCAAACTTGCCCGT- TCTTCCCCTTCATGCCGGTGTTTTGGTTTTTTGCTCTGAAATCTGTTGCAGATATAATAA- GCGTACGTGACAAATGCCCTCCTTCAATTTACACCTGAGTGAGATTTCATTTTTCTTTT- GCTTTTTGAGGGATTTTTTTTTCTACCACCAACATATCGATCGTCGCTTGCTTTAGATAG- CAAAAGGGAACTCGAATTTGAGTAAATTTGATAACACCTTTAACTTTTTTGTCCAAAAGT- GCCCAAGTTAATTGTAATCTGTAGCTCTCAGAAATTTGAGCTAAAGTTGCCATTCCTT
SCDS	ATGGCCGCCTCCTCCTCCAACCCCACGCCTCTGCTGCTGCCACCGCCGCCGC- CGCAAGAGCCCGCGCCTCTGTCGCCGCCGCCGCCGCTGCCTACGCCGAAGCC- GATCCCCACCGTCGCCGACAACTTCCGCAGCCTGCTCAGGTCCGGCGAGGCGC- TCCTCCGCTTCGCCTTCCGCGGCAACTCCGGCCAATTGACTCACCGGCATCCGCC- TCCGCCTCGGCCGCCGCCGCAGCAGCATCCCCACCACCACAACCGCCCG GCTGAGATCATGAAGCGGCTGCAGCGGGAGAAGTTCGCCGACATGATCAAGCACAT- GGACGGGCACGAGCAGATCGACCGCCTCGTCGCGCTGTACACGAGCAGCGCCAAG- GGGTTCCACCTCCCGGAGCTCCCCGTCCGGGTGAAGGTCGCACTCGACGCGGCCG- GCGCGCTGCTCCTCGTCGACGGCGACGAGCTCGAGCAGGCGAGGGACAGGCTCGC- CAAGGCCAGGAACACGACGGGGCTCGGCTCCAGGTTCGTGTTCGAGTCGAGCACC- CGTGGCGGCAAGGACACCGTCGCCGCGGAGCTCGCCACCGGGCTCGGGGCGGCG- GCGGCGGCGGCGGCCGGCGGGAGGCCACTGGAGCTCACCCGGCTGCAGTACTGC- GCCCACGTCAGCGACCTGCTGTCGATGACCCTCGTCCCGTTTGGCGCCCAGTGCAA- CAACTTCTTGCACGGTTCGAGCCTGATCCAGAGTATCCAGAGCAGAGCGTTGTCCGG- CGGGCCGCCGTCGTATTCGGAGCGTCATGACTGCGGCGCCGGCGTGAGCATCAAGG- GATCAAGATTCAGGGCCTCCATAGCTGAGCTGATCTTTGGTTCTCCAGGCGAACATG- GCGGCGGCGGCGGCGGCGGCAATGGCGATCACGAGGTGCCCAACCGGTTGACGAC- GTTCGGCAAGGTGAGCTACGAGACGGCCGACGACATCAAGCTCAGCTTGTCGGGGC- TGTGGCAGGTCCGCTCGCCGGCGTCCCGGTTCAGCGACCTCGGCGCGCTCGCCGT- GCCGCTGGGCAGCCTGAAGACGCGGAGAGCCATCGCTCCTCCTCCGTCGCCGCCGG- CGACGACGCATCCTCCTCCGTCGCCGGACCTGATGGTGCAGGTGCCGGCGCCTCCT- GCTCCTCCCACCCCCATGATGGGTATAGGGTCGACGGTGGCGGTGCAGGGCTCCGTT- GCCGCCGCCGCCGCCGCCGCCACGGCGCCGTCGTCGCACACGGTGGCCGTCATGG- TGGACTGCGACATGTACGACACCTTGAGAGCAGAAGGCTGGGTCGAGATGGAGACGG- CGGCGGCGGCGACTCCGGCGAGGCGGCGCGGGCCGGTGGCGCGGTGGGGAGTCT- GCGTGTCCGACTGCCCGGAGCACGAGCTCGGGTGGGGGGTGCGGATCGGCGGCAC- GGCGGAGAGGAACGCGCACCGGCCGCACGTGGAAGGGTTCCTCAGCTTCGACCTCG- GCAAGGGCGGCAGGGTGCAGCCGGGCCTCGTCATCGCCATGGACGGCGACAAGCG- GACGCCGGCGCTGGTGCTGAGGTCGTCGTGGCTCATGTGA

SP	MAASSSNPTPLLLPPPPPQEPAPLSPPPPLPTPKPIPTVADNFR- SLLRSGEALLRFAFRGNSGQLTHRHPPPPRPPPQQQHPHHHNRP AEIMKRLQREKFADMIKHMDGHEQIDRLVALYTSSAKGFHLPE- LPVRVKVALDAAGALLLVDGDELEQARDRLAKARNTTGLGSRFVFESSTRGGKDTVA- AELATGLGAAAAAAAGGRPLELTRLQYCAHVSDLLSMTLVPFGAQCNNFLHGSSLIQ SIQSRALSGGPPSYSERHDCGAGVSIKGSRFRASIAELIFGSPGEHGGGGGGGNGDHE- VPNRLTTFGKVSYETADDIKLSLSGLWQVRSPASRFSDLGALAVPLGSLKTRRAIAPPPS- PPATTHPPPSPDLMVQVPAPPAPPTPMMGIGSTVAVQGSVAAAAAAATAPSSHTVAVMVD- CDMYDTLRAEGWVEMETAAAATPARRRGPVARWGVCVSDCPEHELGWGVRIGGTAER- NAHRPHVEGFLSFDLGKGGRVQPGLVIAMDGDKRTPALVLRSSWLM*
FG	expressed protein

Tabela 8. Informações genômicas do gene LOC_Os05g50910, identificado pelo marcador SNP S05_29213630

4 I CONCLUSÕES

Dos oito marcadores SNPs, cinco foram encontrados em genes para a cultura do arroz. Esses genes possuíam as funções de: proteína hipotética (LOC_Os01g49070, S01_23885433); função relacionada a calmodulina (LOC_Os03g06570, S03_3311741); domínio de proteína quinase (LOC_Os04g45920, SNP S04_27195948); proteínas G (LOC_Os05g50910, SNP S05_29213630); domínio de ligação ARK (LOC_Os09g30070, S09_18281732) e expressão de proteínas (LOC_Os10g30210, SNP S10_15699535).

Somente os marcadores SNP S05_28201815 e SNP S08_22353023 não foram identificados em genes no genoma da cultura, porém, os mesmos podem ter potencial futuramente para uso na seleção assistida por marcadores (SAM) para a cultura do arroz.

Os genes identificados podem ser de interesse de melhoramento para a cultura do arroz ou para outras culturas, pois diretamente ou indiretamente estão relacionados a tolerância à fatores abióticos, como tolerância à seca. Se considerarmos o plantio de arroz de terras altas no estado do Piauí, o caráter tolerância à seca é um dos mais importantes a serem estudados, pois a irregularidades de chuva no estado faz com que ocorra baixas produtividades da cultura no estado.

REFERÊNCIAS

BOUCHÉ, N.; YELLIN, A.; SNEDDEN, W. A.; FROMM, H. Plant-specific calmodulin-binding proteins. **Annu. Rev. Plant Biol**, v.56, p.435–466, 2005.

CHAVES, M. M.; OLIVEIRA, M. M. Mechanism's underlying plant resilience to water deficits: prospects for water-saving agriculture. **Journal of Experimental Botany**, Oxford, v. 55, n. 407, p. 2365-2384, 2004.

CHINPONGPANICH, A. *et al.* Expression analysis of calmodulin and calmodulin-like genes from rice, *Oryza sativa* L. **BMC Research Notes**, v.5, p. 625, 2012.

DROILLARD, M. J. *et al.* Protein kinases induced by osmotic stresses and elicitor molecules in tobacco cell suspensions: Two crossroad MAP kinases and one osmoregulation-specific protein kinase. **FEBS Letters**, v.474, p. 217–222, 2000.

FALEIRO, F. G. **Marcadores genético-moleculares aplicados a programas de conservação e uso de recursos genéticos**. 1. ed. Planaltina, DF: Embrapa Cerrados, 2007. 102 p.

JIN, Y.*et al*. Rice male development under drought stress: phenotypic changes and stage-dependent transcriptomic reprogramming. **Molecular Plant**, v.6, p.1630-1645, 2013.

LI, Z. *et al*. QTL mapping of root traits in a doubled haploid population from a cross between upland and lowland japonica rice in three environments. **Theor. Appl. Genet**, v.110, p. 1244–1252, 2005.

PANDEY, S. Heterotrimeric G-protein signaling in plants: conserved and novel mechanisms. **Annual Review of Plant Biology**, v. **70**, p. 213-238, 2019.

RAO, N. K. Plantgenetic resources: advancing conservation and use through biotechnology. **African Journal of Biotechnology**, Nairobi, v. 3, n. 2, p. 136-145, 2004.

SHINOZAKI, K.; YAMAGUCHI-SHINOZAKI, K. Gene networks involved in drought stress response and tolerance. **Journal of Experimental Botany**, v. 58, p.221-227, 2007.

SILVA, O.F. da (2020). Dados conjunturais do arroz (área, produção e rendimento). Brasil- 1986 a 2020* Disponível em: <www.cnpaf.embrapa.br/socioeconomia/index.htm>. Acesso em 11 de julho de 2022.

SINGH, U. *et al*. Molecular markers and their applications in cattle genetic research: A review. **Biomarkers and Genomic Medicine**, v. 6, n. 2, p. 49-58, 2014.

SOSBAI - SOCIEDADE SUL-BRASILEIRA DE ARROZ IRRIGADO. **Arroz irrigado: Recomendações Técnicas da Pesquisa Para o Sul do Brasil/XXXII.** Reunião Técnica da Cultura do Arroz Irrigado. Farroupilha, RS: SOSBAI, 2018. 205p.

URANO, D.; JONES, A. M. Heterotrimeric G protein-coupled signaling in plants. **Annual Review of Plant Biology** v. 65, p. 365-384, 2014.

VIDAL, M.S. *et al*. **Déficit hídrico: aspectos morfofisiológicos**. Campina Grande: Embrapa Algodão, 2005. 19p. (Embrapa Algodão. Documentos, 142).

LIU, C.G.; ZHANG, G.Q. Single nucleotide polymorphism (SNP) and its application in rice. **Yi Chuan**, v. 28, p. 737-44 2006.

XIE, K. *et al*. Direct Phosphorylation and Activation of a Mitogen-Activated Protein Kinase by a Calcium-Dependent Protein Kinase in Rice. **Plant Cell**, v. 26, n. 7, p. 3077-3089, 2014.

YAMBAO, E. B. *et al*. Root xylem influence on the water relations and drought resistance of rice. **Journal of Experimental Botany**, Oxford, v. 43, n. 252, p. 925-932, 1992.

YANG, T.; POOVAIAH, B. W. Calcium/calmodulin mediated signal netword in plants. **Trends Plant Science**, v.8, p. 505-512, 2003.

ZHAO, Y. *et al.* Rice extra-large G proteins play pivotal roles in controlling disease resistance and yield-related traits. **New Phytologist**, v. 234, n. 2, p. 607-617, 2022.