IDENTIFICAÇÃO DA FAMÍLIA GÊNICA bZIP EM CAJUEIRO (*Anacardium Occidentale L.*) UTILIZANDO FERRAMENTAS DE BIOINFORMÁTICA

Autores: Gilza Barcelos de Souza[1], Paulo Cezar Da Cunha Junior[1], Natanael Farias Almeida[1], Vinicius Sartori Fioresi[2], Angélica Maria Nogueira Portilho[2].

[1] Instituto Federal de Sergipe/Campus Glória, Povoado Piabas, Zona Rural- 49680000 - Nossa Senhora da Glória-SE, Brasil, gilza.barcelos@ifs.edu.br, paulo.junior@ifs.edu.br, natanaelfarias066@gmail.com.

[2] Universidade Federal do Espírito Santo, Centro de Ciências Agrárias e Engenharias, Alto Universitário, s/n, Guararema – 29500-000 – Alegre - ES, Brasil, vinicius_fioresi@hotmail.com, axnogueira@hotmail.com.

Resumo

O cajueiro (*Anacardium occidentale* L.) é uma espécie de importância socioeconômica para o Brasil, principalmente para o Nordeste. Apesar de sua relevância, os recursos genômicos do cajueiro ainda são escassos, e a família de fatores de transcrição bZIP (*Basic Leucine Zipper*) ainda não foi caracterizada na espécie. Este estudo teve como objetivo identificar e caracterizar, por meio de análises *in silico*, os genes da família bZIP presentes no genoma do cajueiro (*A. occidentale*). As proteínas bZIP de *Arabidopsis thaliana* foram utilizadas como referência em buscas BLASTP na plataforma Phytozome. Os domínios conservados foram confirmados no InterPro, e as propriedades físico-químicas foram determinadas pelo ExPASy ProtParam. Foram identificados 106 genes AobZIP, cujas proteínas apresentaram comprimento entre 75 e 781 aminoácidos, massa molecular de 8,44 a 87,23 kDa e ponto isoelétrico entre 4,29 e 10,56. Este é o primeiro estudo de caracterização da família bZIP em *A. occidentale*, fornecendo informações que poderão contribuir para a identificação de genes associados à tolerância a estresses.

Palavras-chave: Fatores de transcrição. Análise in sílico. Proteínas.

Área do Conhecimento: Ciências Biológicas.

Introdução

O cajueiro *(Anacardium ocidentalize* L.) é uma espécie frutífera nativa da América do Sul, mas atualmente é cultivada em várias regiões tropicais do planeta (Chen et al., 2023). Esta espécie tem grande relevância socioeconômica para o Brasil. Em 2024, a produção brasileira de castanha-de-caju foi estimada em 159.212 toneladas, concentrando-se principalmente na região Nordeste, com o Ceará como o maior produtor nacional (IBGE, 2026).

A castanha-de-caju constitui o principal produto comercial do cajueiro (Yao *et al.*, 2026). Entretanto, apesar do seu potencial econômico e nutricional infelizmente o pseudofruto são descartados como subprodutos nos campos após a colheita da castanha-de-caju (Chen et al., 2023). A castanha-de-caju apresenta em sua composição ácidos graxos, proteínas, vitaminas do complexo B, ácidos fenólicos, flavonoides e outros constituintes bioativos. Por sua vez, o pseudofruto é abundante em fibras alimentares, açúcares redutores, ácido ascórbico, carotenoides, polissacarídeos e uma ampla variedade de compostos fenólicos (Chen et al., 2023).

Apesar da importância socioeconômica do cajueiro, os recursos genômicos disponíveis ainda são limitados. A primeira montagem e anotação do genoma foi realizada a partir de dados sequenciamento obtidos pelas tecnologias Oxford Nanopore e Illumina (Savadi, et al., 2022). Embora esse avanço tenha expandido as possibilidades de estudos genômicos na espécie, diversas famílias gênicas ainda permanecem pouco caracterizadas. Entre elas, destaca-se a família de fatores de transcrição bZIP (*Basic Leucine Zipper*), cuja anotação e caracterização ainda não foi descrita para o cajueiro.

Os fatores de transcrição bZIP estão presentes em todos os organismos eucarióticos (Jakoby et al., 2002), particularmente em plantas, membros desta família estão envolvidos na regulação de vários processos fisiológicos importantes, tais como resistência e tolerância a estresses bióticos e abióticos

(Jakoby et al., 2002; Zhang et al., 2018). Em virtude de sua relevância funcional, a família bZIP tem sido sistematicamente investigada em várias espécies vegetais, demonstrando ampla distribuição e diversidade entre diferentes grupos de plantas. Em espécies de importância econômica, os fatores de transcrição bZIPs foram identificados em melão (*Cucumis melo*), melancia (*Citrullus lanatus*) (Unel et al., 2019), morango (*Fragaria vesca*) (Wang et al., 2017), maçã (*Malus domestica*) (Li et al., 2016), uva (*Vitis vinifera*) (LIU et al., 2014), e soja (*Glycine max*) (Zhang et al., 2018, Zhou et al., 2026).

Entretanto, apesar da relevância socioeconômica do cajueiro e da disponibilidade recente de seu genoma, a família bZIP ainda não foi anotada nessa espécie. Este estudo teve como objetivo identificar e caracterizar, por meio de análises *in silico*, os genes da família bZIP presentes no genoma do cajueiro (*A. occidentale*). Neste sentido, a identificação e caracterização dos genes AobZIPs, poderá contribuir para a ampliação dos recursos genéticos disponíveis para a espécie, bem como fornecer informações que subsidiem futuras investigações sobre a função dessas proteínas em diferentes processos fisiológicos.

Metodologia

As sequências das proteínas bZIP de *A. thaliana* (DRÖGE-LASER et al., 2018) foram coletadas do Arabidopsis Information Resource (TAIR) (The Arabidopsis Information Resource, 2026) e usadas para o BLASTP no Phytozome (Joint Genome Institute, 2026). A presença do domínio bZIP foi confirmado no programa ... (MISTRY et al., 2021). Os membros da família bZIP em *Anacardium occidentale* foram denominadas de AobZIP1-106. Além disso, foram determinamos os valores de massa molecular (MM), ponto isoelétrico (pI) das proteínas CcbZIPs utilizando o site proteômico ExPASy (Artimo et al., 2012). A localização subcelular de todas as proteínas CcbZIPs foram previstas através da ferramenta online "CELLO" (http://cello.life.nctu.edu.tw/) (YU et al., 2004; YU et al., 2006).

Resultados

Nesse estudo, as proteínas AtbZIP de *A.thaliana* foram utilizadas para identificar potenciais genes bZIP em *A. occidentale*. O locus ID, localização, comprimento da proteína, massa molecular (MM) e pI (ponto isoelétrico) dos genes AobZIPs e suas respectivas proteínas, estão apresentados na Tabela 1. Os genes foram nomeados de acordo com sua localização física nos *scaffold*, sendo 106 genes em *A. occidentale* (AobZIP1-AobZIP106). Destes, 87 genes AobZIP foram mapeados nos *scaffold*, que corresponde a 90% do genoma montado. A distribuição não é homogênea, o maior número de genes foi mapeado nos *scaffold* 3 (AobZIP11, AobZIP12, AobZIP13, AobZIP14, AobZIP15, AobZIP16, AobZIP17, AobZIP18 e AobZIP19) *scaffold* 16 (AobZIP72, AobZIP73, AobZIP74, AobZIP75, AobZIP76, AobZIP77, AobZIP78 e AobZIP79).

Com base nos dados disponíveis do genoma do cajueiro, foram identificados 106 genes *AobZIP*, que codificam, ao todo, 187 proteínas. O número de transcritos alternativos variou consideravelmente entre os genes da família AobZIP. Dos 106 genes identificados, 57 apresentaram apenas um transcrito, enquanto os demais apresentaram entre dois e 20 transcritos alternativos. O gene AobZIP2 contabilizou o maior número de transcritos alternativos (Tabela 1).

As 106 proteínas AobZIP referente aos respectivos transcritos principais de cada genes possuem o domínio bZIP confirmado no PFAM. As proteínas apresentaram comprimento variando de 75 (AobZIP39) a 781 (AobZIP84) aminoácidos, massa molecular entre 8436 Da (AobZIP39) e 87231 Da (AobZIP14) e ponto isoelétrico teórico entre 4,29 (AobZIP39) e 10,56 (AobZIP81). Todas as proteínas apresentaram predição de localização subcelular no núcleo, consistente com a função dos fatores de transcrição bZIP (tabela 1).

Tabela 1: As características detalhadas dos genes bZIP identificados em *Anacardium Occidentale L.*

Nome	Gene	scaffolds	SA	tamanho	pI	Massa molecular
AobZIP1	Anaoc.0001s0771	1	0	435	8.09	48129
AobZIP2	Anaoc.0001s0932	1	20	268	5.43	28580
AobZIP3	Anaoc.0001s1036	1	1	203	10.01	22973

AobZIP4	Anaoc.0001s1509	1	0	210	10.39	24220
AobZIP5	Anaoc.0001s2160	1	3	328	6.46	36823
AobZIP6	Anaoc.0002s0605	2	0	461	5.75	50851
AobZIP7	Anaoc.0002s0674	2	0	205	6.19	23342
AobZIP8	Anaoc.0002s1333	2	0	482	9.09	53891
AobZIP9	Anaoc.0002s1675	2	0	146	5.59	16693
AobZIP10	Anaoc.0002s1935	2	0	421	7.19	44610
AobZIP11	Anaoc.0003s0112	3	0	264	5.90	29411
AobZIP12	Anaoc.0003s0531	3	0	207	6.20	23652
AobZIP13	Anaoc.0003s0608	3	0	472	5.73	52279
AobZIP14	Anaoc.0003s0977	3	2	781	5.83	87231
AobZIP15	Anaoc.0003s1321	3	0	296	6.22	32803
AobZIP16	Anaoc.0003s1384	3	0	340	9.04	36166
AobZIP17	Anaoc.0003s1584	3	0	203	7.85	22870
AobZIP18	Anaoc.0003s1665	3	5	279	5.08	31196
AobZIP19	Anaoc.0003s1783	3	5	604	6.15	65663
AobZIP20	Anaoc.0004s0890	4	0	394	5.32	43135
AobZIP21	Anaoc.0005s0211	5	0	155	4.69	17467
AobZIP22	Anaoc.0005s0562	5	0	244	9.20	27570
AobZIP23	Anaoc.0005s1084	5	1	351	9.22	39470
AobZIP24	Anaoc.0005s1500	5	4	404	5.99	45834
AobZIP25	Anaoc.0005s1502	5	3	498	7.76	55337
AobZIP26	Anaoc.0005s1645	5	3	406	6.13	43933
AobZIP27	Anaoc.0006s0019	6	5	448	8.80	50109
AobZIP28	Anaoc.0006s0192	6	0	383	8.89	40756
AobZIP29	Anaoc.0006s0654	6	1	295	6.46	33172
AobZIP30	Anaoc.0006s1230	6	8	363	8.64	40725
AobZIP31	Anaoc.0006s1535	6	0	141	9.09	15926
AobZIP32	Anaoc.0006s1757	6	3	414	6.24	44098
AobZIP33	Anaoc.0007s1058	7	1	268	4.77	29386
AobZIP34	Anaoc.0008s0005	8	0	146	9.98	17095
AobZIP35	Anaoc.0008s0137	8	0	129	9.51	14762
AobZIP36	Anaoc.0008s0892	8	0	387	5.73	41937
AobZIP37	Anaoc.0008s1565	8	0	400	6.24	44094
AobZIP38	Anaoc.0009s0266	9	0	521	8.54	57955
AobZIP39	Anaoc.0009s0321	9	1	75	4.29	8436
AobZIP40	Anaoc.0009s0504	9	0	427	9.39	46416
AobZIP41	Anaoc.0009s0618	9	0	143	8.20	16753
AobZIP42	Anaoc.0009s0865	9	2	256	6.99	29419
AobZIP43	Anaoc.0009s1089	9	1	343	7.27	37942
AobZIP44	Anaoc.0009s1110	9	6	309	5.47	34325
AobZIP45	Anaoc.0010s0506	10	8	413	5.98	45825
AobZIP46	Anaoc.0010s0517	10	0	243	8.81	27568

AobZIP47	Anaoc.0010s1197	10	2	510	7.35	56850
AobZIP48	Anaoc.0010s1199	10	0	348	6.18	39129
AobZIP49	Anaoc.0010s1308	10	0	368	8.74	41415
AobZIP50	Anaoc.0010s1357	10	1	682	7.02	74458
AobZIP51	Anaoc.0010s1556	10	0	181	5.74	21163
AobZIP52	Anaoc.0011s0016	11	0	602	6.49	65486
AobZIP53	Anaoc.0011s0376	11	2	270	6.55	29774
AobZIP54	Anaoc.0011s0524	11	2	437	9.57	47470
AobZIP55	Anaoc.0011s0545	11	0	167	5.93	18839
AobZIP56	Anaoc.0011s0943	11	0	350	6.33	39198
AobZIP57	Anaoc.0012s0215	12	2	305	5.56	33803
AobZIP58	Anaoc.0012s0232	12	1	351	7.26	38976
AobZIP59	Anaoc.0012s0414	12	0	328	9.16	37478
AobZIP60	Anaoc.0012s0618	12	0	140	6.19	16358
AobZIP61	Anaoc.0012s0729	12	0	429	9.31	46462
AobZIP62	Anaoc.0012s0885	12	0	496	6.94	54389
AobZIP63	Anaoc.0012s0916	12	0		5.92	18074
AobZIP64	Anaoc.0012s1012	12	1	262	9.33	28916
AobZIP65	Anaoc.0013s0007	13	2	165	9.51	18023
AobZIP66	Anaoc.0013s0008	13	0	165	9.51	18023
AobZIP67	Anaoc.0014s0145	14	0	380	6.55	41963
AobZIP68	Anaoc.0014s0256	14	0	218	6.60	25062
AobZIP69	Anaoc.0014s0646	14	2	301	6.32	33852
AobZIP70	Anaoc.0014s1089	14	0	415	8.35	44293
AobZIP71	Anaoc.0014s1183	14	0	185	6.76	20808
AobZIP72	Anaoc.0016s0762	16	8	283	8.43	31278
AobZIP73	Anaoc.0016s0796	16	0	320	9.08	35888
AobZIP74	Anaoc.0016s0852	16	8	469	6.74	51998
AobZIP75	Anaoc.0016s1015	16	0	325	5.56	36310
AobZIP76	Anaoc.0016s1418	16	8	466	6.58	51569
AobZIP77	Anaoc.0016s1462	16	4	325	7.03	36787
AobZIP78	Anaoc.0016s1478	16	2	307	8.54	34035
AobZIP79	Anaoc.0016s1558	16	3	273	6.08	30972
AobZIP80	Anaoc.0017s0283	17	8	394	5.76	41665
AobZIP81	Anaoc.0017s0526	17	0	180	10.56	20809
AobZIP82	Anaoc.0018s0571	18	8	455	6.18	50224
AobZIP83	Anaoc.0018s0591	18	0	195	9.67	21561
AobZIP84	Anaoc.0018s0621	18	1	751	7.07	81946
AobZIP85	Anaoc.0018s0833	18	0	182	6.30	21214
AobZIP86	Anaoc.0019s0226	19	1	340	6.08	37356
AobZIP87	Anaoc.0020s0496	20	0	226	5.83	26173
AobZIP88	Anaoc.0021s0018	21	3	371	7.77	41360
AobZIP89	Anaoc.0029s0004	29	5	371	7.27	41355

AobZIP90	Anaoc.0047s0009	47	8	410	6.23	43645
AobZIP91	Anaoc.0075s0002	75	0	301	6.16	33827
AobZIP92	Anaoc.0284s0001	284	0	332	7.78	37087
AobZIP93	Anaoc.0601s0015	601	0	171	6.98	19032
AobZIP94	Anaoc.0602s0012	602	1	356	5.90	38713
AobZIP95	Anaoc.0647s0002	647	0	320	9.08	35888
AobZIP96	Anaoc.0812s0001	812	3	432	6.16	47153
AobZIP97	Anaoc.0861s0001	861	0	454	6.78	49973
AobZIP98	Anaoc.0950s0006	950	0	260	4.99	28247
AobZIP99	Anaoc.1003s0004	1003	4	298	5.47	31437
AobZIP100	Anaoc.1043s0008	1043	3	270	6.55	29774
AobZIP101	Anaoc.1044s0003	1044	1	461	5.75	50882
AobZIP102	Anaoc.1129s0003	1129	0	556	6.38	60577
AobZIP103	Anaoc.1134s0004	1134	9	411	6.33	43865
AobZIP104	Anaoc.1212s0001	1212	1	648	6.88	71393
AobZIP105	Anaoc.1445s0004	1445	0	435	8.09	48129
AobZIP106	Anaoc.1512s0002	1512	0	343	7.27	37942

Fonte: Autores (2026).

Discussão

Observa-se uma diferença entre o número de genes nas principais linhagens evolutivas de plantas. A família bZIP passou por uma expansão significativa durante a evolução das plantas, sendo ainda mais acentuada nas angiospermas. As monocotiledôneas apresentaram os maiores números de cópias de bZIP, com média de 110,37 genes e um máximo de 563 genes, enquanto as eudicotiledôneas com uma média de 79,64 genes e um máximo de 304 genes (LIU et al., 2026). Além da variação no número de genes nas diferentes espécies vegetais, observa-se também um mecanismo de regulação pôs-transcricionais por meio de splicing alternativo, que resulta em aumento no número de isoformas de cada gene. Em caju,os 106 genes AobZIP correspondem a 187 proteínas bZIP, indicando uma variação considerável no número de transcritos por genes bZIP. Em soja, os 171 genes GmbZIP codificam 398 proteínas, dentre essas algumas perderam o domínio bZIP devido ao *splicing* alternativo (ZHOU et al., 2026). Em arroz, isoforma do gene OsbZIP58 pode representar um importante mecanismo regulatório da termotolerância. Esse mecanismo de modificação pós-transcricional que resulta em diversidade de transcritos pode ser importante para a adaptabilidade das plantas as variações ambientais (XU et al., 2020). Essas variações também são observadas quando se analisa as características físico-químicas (comprimento, massa molecular e ponto isoelétrico) das proteínas preditas em cajueiro. O ponto isoelétrico teórico entre 4,29 (AobZIP39) e 10,56 (AobZIP81) indica as proteínas vão do caráter ácido ao básico. Além disso, as proteínas AobZIP apresentam diferenças de massa molecular e comprimento da sequência de aminoácidos, entre 8436 a 87231 Da e 75 a 781, respectivamente.

Conclusão

Este estudo permitiu identificar e caracterizar 106 genes da família *AobZIP* no genoma de *A. occidentale*, dos quais 87 foram mapeados nos *scaffolds* que correspondem a 90% do genoma montado. Esses genes codificam 187 proteínas, evidenciando a ocorrência expressiva de transcritos alternativos, especialmente no gene *AobZIP2*. As proteínas correspondentes aos transcritos principais apresentaram o domínio bZIP conservado, ampla variação de comprimento, massa molecular e ponto isoelétrico, além de localização nuclear predita, compatível com sua atuação como fatores de transcrição. Esses resultados ampliam os recursos genômicos disponíveis para o cajueiro e fornecem subsídios para futuras análises filogenéticas, de expressão gênica e de caracterização funcional.

Referências

ARTIMO, Panu et al. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research*, Oxford, v. 40, n. W1, p. W597–W603, 2012. DOI: https://doi.org/10.1093/nar/gks400.

CHEN, Chengjie et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant*, Cambridge, v. 13, n. 8, p. 1194–1202, 2020. DOI: https://doi.org/10.1016/j.molp.2020.06.009.

DRÖGE-LASER, Wolfgang; SNOEK, Basten L.; SNEL, Berend; WEISTE, Christoph. The Arabidopsis bZIP transcription factor family: an update. *Current Opinion in Plant Biology*, London, v. 45, p. 36–49, 2018. DOI: https://doi.org/10.1016/j.pbi.2018.05.001.

JAKOBY, Marc et al. bZIP transcription factors in Arabidopsis. *Trends in Plant Science*, Oxford, v. 7, n. 3, p. 106–111, 2002. DOI: https://doi.org/10.1016/S1360-1385(01)02223-3.

JOINT GENOME INSTITUTE (JGI). *Phytozome: plant comparative genomics portal – BLAST search*. Walnut Creek, CA: Lawrence Berkeley National Laboratory, [2026]. Disponível em: https://phytozome-next.jgi.doe.gov/blast-search. Acesso em: 9 set. 2026.

LI, Yuan-Yuan; MENG, Dong; LI, Mingjun; CHENG, Lailiang. Genome-wide identification and expression analysis of the bZIP gene family in apple (*Malus domestica*). *Tree Genetics & Genomes*, v. 12, n. 4, art. 82, 2016. DOI: https://doi.org/10.1007/s11295-016-1043-6.

LIU, Jinyi et al. Genome-wide analysis and expression profile of the bZIP transcription factor gene family in grapevine (*Vitis vinifera*). *BMC Genomics*, v. 15, art. 281, 2014. DOI: https://doi.org/10.1186/1471-2164-15-281.

MISTRY, Jaina et al. Pfam: the protein families database in 2021. *Nucleic Acids Research*, Oxford, v. 49, n. D1, p. D412–D419, 2021. DOI: https://doi.org/10.1093/nar/gkaa913.

UNEL, Necdet Mehmet; CETIN, Fadime; KARACA, Yasin; CELIK ALTUNOGLU, Yasemin; BALOGLU, Mehmet Cengiz. Comparative identification, characterization, and expression analysis of bZIP gene family members in watermelon and melon genomes. *Plant Growth Regulation*, v. 87, n. 2, p. 227–243, 2019. DOI: https://doi.org/10.1007/s10725-018-0465-6.

WANG, Xiao-Long; CHEN, Xinlu; YANG, Tian-Bao; CHENG, Qunkang; CHENG, Zong-Ming. Genome-wide identification of bZIP family genes involved in drought and heat stresses in strawberry (*Fragaria vesca*). *International Journal of Genomics*, v. 2017, art. 3981031, 2017. DOI: https://doi.org/10.1155/2017/3981031.

YU, Chih-Hao; LIN, Chien-Ju; HWANG, Jenn-Kang. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on *n*-peptide compositions. *Protein Science*, Cambridge, v. 13, n. 5, p. 1402–1406, 2004. DOI: https://doi.org/10.1110/ps.03479604.

YU, Chih-Hao et al. Prediction of protein subcellular localization. *Proteins: Structure, Function, and Bioinformatics*, Hoboken, v. 64, n. 3, p. 643–651, 2006. DOI: https://doi.org/10.1002/prot.21018.

ZHANG, Man et al. Evolutionary and expression analyses of soybean basic leucine zipper transcription factor family. *BMC Genomics*, v. 19, art. 159, 2018. DOI: https://doi.org/10.1186/s12864-018-4511-6.

ZHOU, Enqiang et al. Global analysis of soybean bZIP gene family under stress and identification of salt-stress-responsive candidate genes. *Frontiers in Plant Science*, v. 17, art. 1806221, 2026. DOI: https://doi.org/10.3389/fpls.2026.1806221.