

Full Length Research Paper

# Whole genome homology-based identification of candidate genes for drought tolerance in sesame (*Sesamum indicum* L.)

Komivi Dossa<sup>1,2,3\*</sup>, Marème Niang<sup>1</sup>, Achille E. Assogbadjo<sup>2</sup>, Ndiaga Cissé<sup>1</sup> and Diaga Diouf<sup>3</sup>

<sup>1</sup>Centre d'Etudes Régional pour l'Amélioration de l'Adaptation à la Sécheresse (CERAAS), BP 3320 Route de Khombole, Thiès, Sénégal.

<sup>2</sup>Laboratoire d'Ecologie Appliquée, Faculté des Sciences Agronomiques, Université d'Abomey-Calavi, 01 BP 526, Calavi, Bénin.

<sup>3</sup>Laboratoire Campus de Biotechnologies Végétales, Département de Biologie Végétale, Faculté des Sciences et Techniques, Université Cheikh Anta Diop, BP 5005 Dakar-Fann, Dakar, Sénégal.

Received 19 April, 2016; Accepted 29 June, 2016

**Sesame (*Sesamum indicum* L.) is one of the most important oilseed crops. It is mainly grown in arid and semi-arid regions with occurrence of unpredictable drought which is one of the major constraints of its production. However, the lack of gene resources associated with drought tolerance hinders sesame genetic improvement towards this osmotic stress. The present research aimed at identifying candidate genes associated with drought tolerance in the whole genome of sesame through homology search of known drought associated genes from three relative species, viz., potato, tomato and the well-described model plant *Arabidopsis*. Based on 2,495 sequences including 1,150 from *Arabidopsis*, 1,075 from potato and 270 from tomato, comparative analysis against sesame genome led to the identification of a set of 75 candidate genes (42, 22 and 11 from *Arabidopsis*, potato and tomato, respectively). Mapping results showed that the candidate genes were distributed on the 16 sesame linkage groups. Wide range of genes with various functions identified in this study confirmed that drought tolerance in sesame is under the control of several genes. Based on their functional classification, 2 groups of candidate genes were identified: (a) genes which protect the plant against drought effect; (b) signal transduction genes and transcription factors. Many transcription factors were retrieved including 6 AP2/ERF genes among which 3 were more expressed in drought tolerant material compared to the sensitive one and might play some important roles in drought tolerance in sesame. Our results provided genomic resources for further functional analysis and genetic engineering towards drought tolerance improvement in sesame.**

**Key words:** *Sesamum indicum*, candidate genes, drought tolerance, orthologous gene, whole genome.

## INTRODUCTION

Sesame (*Sesamum indicum* L.,  $2n = 2x = 26$ ) is one of the most commonly grown oilseed crops a seed

production of more than 4.8 million tons worldwide in year 2013 (FAOSTAT, 2013) and has been suggested as the

most ancient oil crop (Nayar and Mehra, 1970). Its seeds are an important source of high-quality oil and contain natural antioxidants such as sesamin and sesamol (Zhang et al., 2013). Sesame is a good source of vitamins (pantothenic acid and vitamin E), minerals such as calcium (1.450 mg/100 g), phosphorous (570 mg/100 g) for human consumption and the seed cake is also an important nutritious livestock feed (Balasubramaniyan and Palaniappan, 2001).

It is mainly grown in tropical and subtropical regions of Asia, Africa and South America, in marginal lands or under very difficult conditions with drought, high temperatures, high solar radiation and high evaporation demand which make sesame a drought tolerant plant (Langham, 2007; Witcombe et al., 2007). Despite its tolerance, drought is one of the most important environmental factors that limit sesame production by affecting the number of capsules per plant, grain yield as well as oil yield and quality depending on the genotypes and drought intensity (Betram et al., 2003; Hassanzadeh et al., 2009; Bahrami et al., 2012). Recently, drought will be a serious threat in the coming decades as the Intergovernmental Panel on Climate Change (IPCC) has concluded that elevated greenhouse gas concentrations are likely to lead to a general drying of the subtropics by the end of this century, creating widespread drought stress in agriculture (IPCC, 2007). Therefore, improvement of drought tolerance in sesame genotypes is one of the major objectives of sesame breeding programs which can be achieved by integrating new approaches (Pathak et al., 2014). In the past years, many investigations have been carried out to enhance our understanding on the genetic basis of drought tolerance by using the genomics, transcriptomics and transgenesis approaches in the model plant *Arabidopsis thaliana* (Shinozaki et al., 2003; Jang et al., 2004, 2007; Ramirez et al., 2009; Lata et al., 2011; Harshavardhan et al., 2014). These studies showed that the main genes involved in drought tolerance were transcription factors (TFs). In *Arabidopsis*, about 1,500 TFs are considered to be involved in stresses response (Riechmann et al., 2000) including drought. So far, many drought associated genes have been identified including TFs belonging to basic leucine zipper (bZIP), AP2/EREBP, ABA-binding factor (ABF), MYC, MYB, NAM, ATAF1-2, NAC, CCAAT-binding and zinc-finger families and have been characterized in detail (Abe et al., 1997; Bartels and Sunkar, 2005; Sakuma et al., 2006; Nakashima et al., 2014; Mondini et al., 2012, 2015).

The availability of genome sequences in a number of plant species combined with comparative genomics

analysis can improve our understanding of the fundamental aspects of plant biology including the identification and analysis of genes involved in adaptive traits of crops (Foucher et al., 2003). In fact, plant genomes share extensive similarities known as synteny, even between distantly related species (Guyot et al., 2012). Through comparative analysis against the *Arabidopsis* genome, many functional genomics regions and candidate genes such as flowering time *FLC* genes (Schranz et al., 2002), clubroot resistance genes (Suwabe et al., 2006), aliphatic glucosinolate biosynthetic pathway (Bisht et al., 2009) and genes for male fertility (Ashutosh et al., 2012) have been identified in *Brassica*. Similar strategies have been used to predict stress-responsive TFs in soybean, maize, sorghum, barley and wheat based on *Arabidopsis* and rice genome analyses (Mochida et al., 2009; Tran and Mochida, 2010b). In addition, alike analyses have been performed in tomato (*Solanum lycopersicum*) and potato (*Solanum tuberosum*) two economically important and naturally drought sensitive crops (Li et al., 2013; Obidiegwu et al., 2015) belonging to the *Asteridae* subclass which includes sesame, leading to the identification of drought tolerant genes in these crops (Reiter and Vanzin, 2001; Vasquez-Robinet et al., 2008; Evers et al., 2010; Anithakumari et al., 2011, 2012; Solankey et al., 2014). Thus, it is well documented that using synteny approach in closely related species is suitable for the identifying orthologous genes (Rubin, 2001).

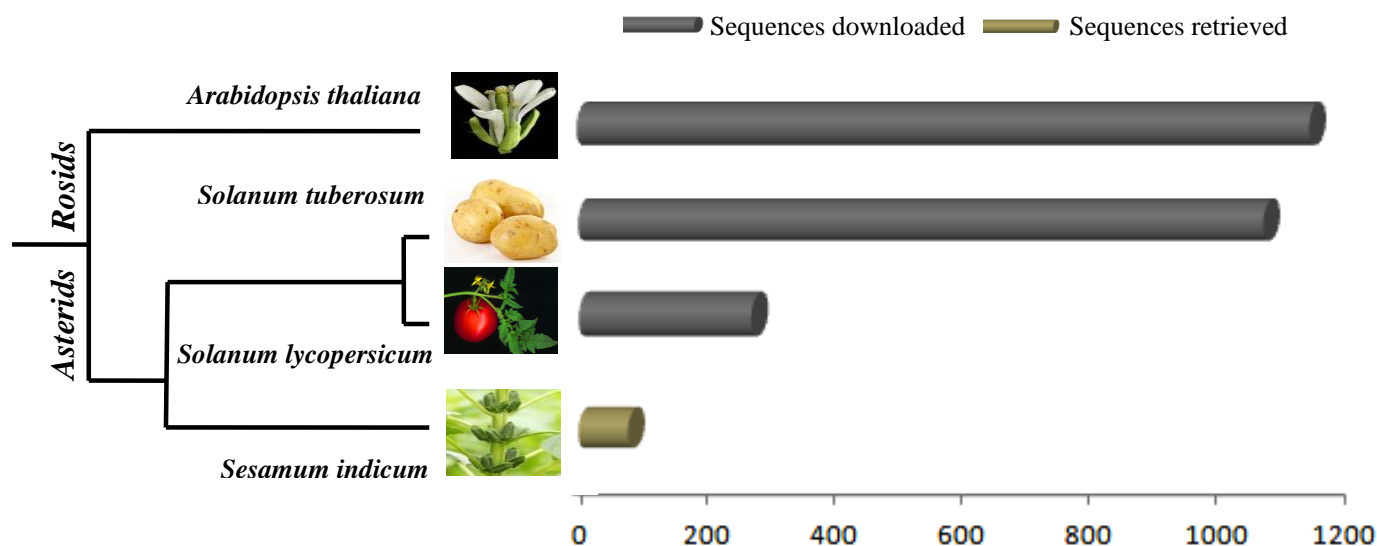
The identification of drought related candidate genes in sesame will provide useful information for its improvement. In the best of our knowledge no data have been reported regarding drought tolerance genes identified in sesame. Based on the sesame genome sequence recently released by Wang et al. (2014), a set of candidate genes in whole genome of sesame were identified in this study through homology search of known drought associated genes from three relatives species, viz. tomato, potato, and *Arabidopsis* and these genes were analyzed for further functional and validation experiments.

## MATERIALS AND METHODS

### Data resources

*Arabidopsis thaliana* genomic and annotation data were downloaded from the TAIR (<http://www.arabidopsis.org>) (Huala et al., 2001) while the potato and tomato genomes were downloaded from the Sol Genomics Network (<http://www.solgenomics.net>) (Fernandez-Pozo et al., 2015). In addition, the publicly known drought related

\*Corresponding author. E-mail: dossakomivi@gmail.com.



**Figure 1.** Phylogenetic relationships of the species studied and number of sequences downloaded.

genes and proteins sequences of these three species were downloaded from the Sol Genomics Networks, TAIR, NCBI database (<http://www.ncbi.nlm.nih.gov>) (Pruitt et al., 2007), and Drought Stress Gene Database (<http://pgsb.helmholtz-muenchen.de/droughtdb/>) (Alter et al., 2015). A total of 2,495 sequences including 1,150 from *Arabidopsis*, 1,075 from potato and 270 from tomato were downloaded. The phylogeny data of the four species were downloaded from NCBI Taxonomy common tree (<http://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi>) and a tree was constructed by MEGA 6.0 software (Tamura et al., 2013) as shown in Figure 1.

### Comparative genomics and genes expression assay

A local bank with the retrieved sequences was generated in order to make searches for similar sequences against the sesame genome (Wang et al., 2014) using the BLASTn and tBLASTn algorithms (Altschul et al., 1990) for DNA and protein sequences, respectively with a cut-off of  $1e^{-30}$ . A threshold value of 70% identity was considered as significant level (Roy et al., 2011). After removing redundant genes, analyses of the candidate drought related genes in the whole genome of sesame were carried out including their identification, classification in functional groups, sequences analysis and chromosomal location.

The sesame genes with unknown functions were submitted to the AutoFACT program (Koski et al., 2005), and annotated according to the data available in the largest functional annotation databanks (KEGG, PFAM, SMART). The homologous genes found in the sesame genome were mapped onto the 16 Linkage Groups (LGs) according to their physical positions using MapChart 2.3 (Voorrips, 2002). The comparative orthologous relationships of the candidate drought associated genes among sesame and *Arabidopsis* were illustrated using Circos program (Krzywinski et al., 2009). To find out the whole genome AP2/ERF genes, the Hidden Markov Model (HMM) profile of the AP2/ERF domain (PF00847) obtained from Pfam v28.0 database (<http://Pfam.sanger.ac.uk/>) (Finn et al., 2014) was searched against the sesame proteome using Unipro

UGENE (Okonechnikov et al., 2012).

Furthermore, a drought stress experiment was carried on to assess the expression of 6 AP2/ERF genes retrieved. For that, two contrasting sesame accessions (LC164-drought tolerant) and (hb168-drought sensitive), previously studied by Boureima et al. (2012), were sown in pots (25 cm diameter and 30 cm depth) filled with a mixture of soil, sand, and compost (5:2:2, v/v/v). The seedlings were grown and watered normally during 21 days before applying drought stress by withholding water for 5 days. At this stage, all plants were transferred under a plastic rain shelter. Total RNA of drought-stressed sesame seedlings were extracted from leaves using Trizol Reagent (Invitrogen, USA) according to the manufacturer's protocol and digested with DNase I (MBI, USA) to remove the genomic DNA contamination. One microgram RNA was reverse transcribed using the Reverse Transcription System (Promega). The semi quantitative reverse-transcription PCR (RT-PCR) amplification was carried out using gene specific primers (Table 1) and the cDNA libraries synthesized by using the following protocol: 4-min incubation step at 95°C for complete denaturation, followed by 30 cycles consisting of 95°C for 30 s, 55°C for 30 s, 72°C for 30 s and the final cycle at 72°C for 5 min. RT-PCR products were run on 2.0% (w/v) agarose gel, stained with ethidium bromide ( $5 \mu\text{g ml}^{-1}$ ) and the expression of each gene was qualitatively evaluated after imaging under UV light (Kodak EDAS 290).

## RESULTS AND DISCUSSION

The identification of drought tolerance candidate genes which have a high potential to be used for breeding drought tolerant crops presents a challenge (Krannich et al., 2015). For species with largely unexplored genomes such as sesame, comparative genomics is a promising tool to gain information by utilizing the conservation between closely related plant species (Akpınar et al.,

**Table 1.** Primers used for the RT-PCR.

Locus	Primer code	Forward sequence (5'-3')	Reverse sequence (5'-3')
LOC105155867	SSAp1	ATTTCCATCATCTAACCCCTAAC	CCTCATCCGTATCCCTCTAT
LOC105165339	SSAp2	GAAAGCAAAGGCAACA	TATCTCGGCTACCCAC
LOC105160523	SSAp3	AGGGAGCCCAACAAGA	TTAGCATTTCGCAGACG
LOC105162917	SSAp4	GTCGCCTTGTCTTCTGT	TTCCTCGTTGTTGTCTGG
LOC105157874	SSAp5	CGATGACTGACGACGGAATG	GGTGAGGGTGCCCAACAA

2013). Hence, comparative genomics will broaden the ability to transfer information from model plants to other species that are fundamental to food production or as a source of alternative energy (Ma et al., 2012). From a total of 2,495 sequences downloaded from *Arabidopsis*, tomato and potato genomes, 75 candidate genes were found in the whole genome of sesame with high identity (Figure 1). Among these candidate genes, 42, 22 and 11 genes were homologous to *Arabidopsis*, potato and tomato genes, respectively. *Arabidopsis*, potato, tomato and sesame belong to the Asteridae subclass which includes nearly 60,000 species. Lee et al. (2005) reported that genes involved in adaptive processes tend to be highly conserved. Therefore, interspecies sequence comparison is a powerful tool to extract functional or evolutionary information from the genomes of organisms (Chiba et al., 2008). After functional annotation and classification, only 2 candidate genes with unknown functions remained (Table 2). The whole set of candidate genes could be classified into 2 main categories according to earlier reports on osmotic stress responsive genes in *Arabidopsis* (Seki et al., 2002, 2003): (a) genes which protect the plant against drought effect and (b) signal transduction genes and TFs. Drought tolerance is a quantitative trait that exhibits complex genetic control (Mc William, 1989). It greatly affects the plant both at the micro (that is, membrane structure), and at the macro level (that is, the physiology of the whole plant), with results that reflect the variety of responses involved in the acquisition of tolerance (Soares et al., 2012). The complexity of this trait explains why there is slow progress in crop improvement in drought-prone areas (Cattivelli et al., 2008).

The candidate genes were mapped onto the 16 Linkage Groups (LGs) in the sesame genome (Wang et al., 2014) (Figure 2). All LGs were represented and the distribution of these genes was globally uneven. However, some gene clusters existed on LGs 1, 3, 5, 6, 7, 8, 10 and 14. Recent works of Wei et al. (2015) on MADS gene family in sesame also find similar clustering patterns of some MADS genes along 14 LGs. In fact, there is evidence that functionally related genes tend to cluster more commonly than expected by chance (Boutanaev et al., 2002; Cohen et al., 2002). Our results suggested that these clustering regions of the genome

might be highly active in drought tolerance in sesame. The maximum number of genes (10; 13.16%) was localized on LG 3, whereas LGs 13 and 16 have the lowest number of genes (1; 1.3%).

To trace orthologous relationships of the candidate genes associated to drought, the physically mapped candidate genes of sesame were compared with those of *Arabidopsis* since most of the genes in *Arabidopsis* have been functionally characterized (Figure 3). According to Wang et al. (2014), sesame and *Arabidopsis* share more than 2,200 homologous genes. Forty three orthologous gene pairs were detected between *Arabidopsis* and sesame, including 43 sesame candidate genes and 32 *Arabidopsis* drought associated genes. Most of the candidate genes in sesame showed syntenic bias towards the chromosomes 1, 2 and 4. The comparative mapping information offers a useful preface for understanding the evolution of genes among different species.

Many drought stress associated genes encode TFs that in turn control other various genes involved in diverse physiological and molecular responses to drought stress. TFs are therefore good candidates for genetic engineering to improve crop tolerance to drought because of their role as master regulators of clusters of genes (Rabara et al. 2014). AP2/ERF transcription factors are reported to be involved in drought stress in many plants (Licausi et al., 2010; Lata et al., 2014; Rabara et al., 2014). Whole genome scanning of AP2/ERF genes resulted in 132 putative AP2/ERF genes. Regarding the importance of this gene superfamily in abiotic stress tolerance in plants, many AP2/ERF genes were expected to be found among the set of candidate genes identified in this study. Only five candidate AP2/ERF genes (3.8%) were retrieved as probably associated to drought tolerance in sesame (Figure 4), suggesting that these AP2/ERF genes should be targeted for drought research in sesame. According to the classification of Sakuma et al. (2002), the AP2/ERF genes found in this study, were classified into the ERF (2), AP2 (2) and DREB (1) subfamilies, respectively.

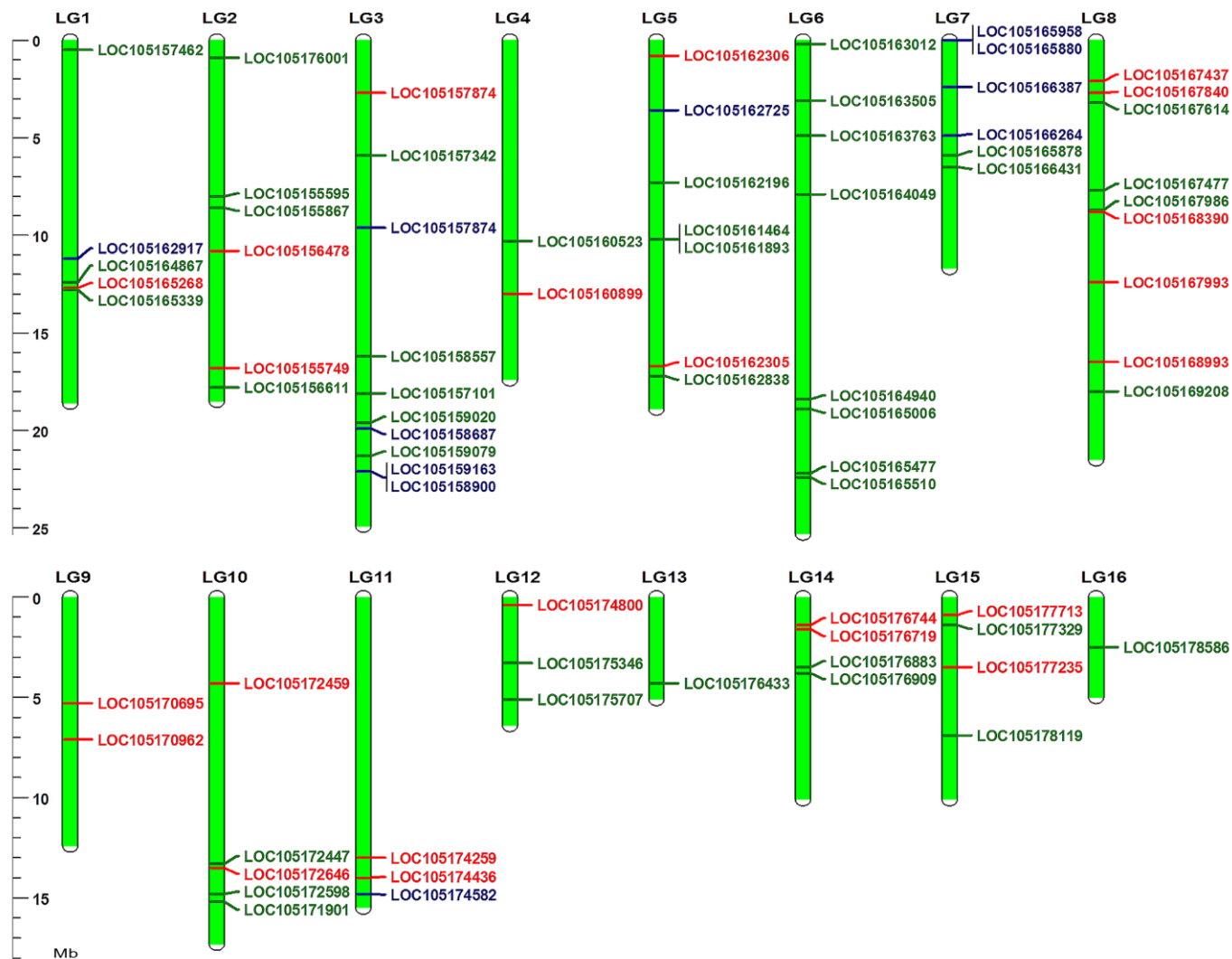
Although, the power of similarity-based gene discovery at a genome scale has been demonstrated in many works and partially reviewed by Windsor et al. (2006), the importance of functional characterization cannot be

**Table 2.** List and functions of orthologous genes retrieved from sesame genome.

Homologous species	Gene ID (Genbank)	Linkage group (LG)	Start (bp)	End (bp)	Predicted Function
<i>Arabi-dopsis</i>	LOC105166387	7	5906451	5910956	ATPase 8, plasma membrane-type
	LOC105162305	5	10204583	10209025	ATPase 8, plasma membrane-type
	LOC105167840	8	7731490	7737270	protein cellulose synthase A catalytic subunit 8
	LOC105172646	10	15149055	15149891	Squalene monooxygenase-like
	LOC105162306	5	10212601	10213008	WD repeat-containing protein 76
	LOC105158687	3	18059219	18059416	Importin-7 homolog
	LOC105163505	6	3064706	3064831	U6 snRNA-associated Sm-like LSm5
	LOC105166431	7	6449340	6449752	Pyrophosphate-energized vacuolar membrane proton pump-like
	LOC105172598	10	14826310	14826859	Chlorophyll a-b binding CP24 10A
	LOC105169208	8	18023545	18024092	Chlorophyll a-b binding CP24 10A
	LOC105165339	1	12753457	12753552	AP2/ERF
	LOC105155867	2	8573570	8573654	AP2/ERF
	LOC105155595	2	8022269	8022013	Zinc finger protein ZAT10-like
	LOC105159020	3	19550162	19550800	Annexin D1-like
	LOC105157342	3	5932731	5932993	Protein argonaute 1
	LOC105178119	15	6916417	6916015	RING-H2 finger protein ATL78-like
	LOC105156611	2	17750487	17750029	Galactinol synthase 1-like
	LOC105163012	6	238678	238355	Galactinol synthase 1-like
	LOC105172447	10	13338815	13338497	Galactinol synthase 1-like
	LOC105159079	3	21325879	21325607	Squalene epoxidase 1-like
	LOC105162196	5	7272028	7271749	Squalene monooxygenase-like
	LOC105164049	6	7911616	7911192	Respiratory burst oxidase protein A
	LOC105167614	8	3228940	3229962	LRR receptor-like serine/threonine-protein kinase RPK2
	LOC105177329	15	1389614	1390636	LRR receptor-like serine/threonine-protein kinase RPK2
	LOC105176001	2	867511	868042	ABC transporter G family member 25
	LOC105165477	6	22166960	22167350	Alpha,alpha-trehalose-phosphate synthase
	LOC105176909	14	3817870	3818260	Alpha,alpha-trehalose-phosphate synthase
	LOC105164867	1	12396709	12395749	Glycerol kinase
	LOC105162838	5	17176580	17177344	Plasma membrane ATPase 4-like
	LOC105164940	6	18440702	18441474	Plasma membrane ATPase 4-like
	LOC105176433	13	4333541	4333880	Homeobox-leucine zipper protein ATHB-6
	LOC105163763	6	4928626	4929458	Indole-3-acetaldehyde oxidase
	LOC105165510	6	22408587	22407052	9-cis-epoxycarotenoid dioxygenase NCED1, chloroplastic
LOC105176883	14	3512386	3513850	9-cis-epoxycarotenoid dioxygenase NCED1, chloroplastic	
LOC105178586	16	2456138	2458102	Protein far-red elongated hypocotyl 3 isoform X2	
LOC105167986	8	8723674	8724107	Calcium-dependent protein kinase-like	
LOC105175346	12	3311638	3311229	Calcium-dependent protein kinase-like	
LOC105167437	8	2135795	2137643	Far-red impaired response 1	

Table 2. Contd.

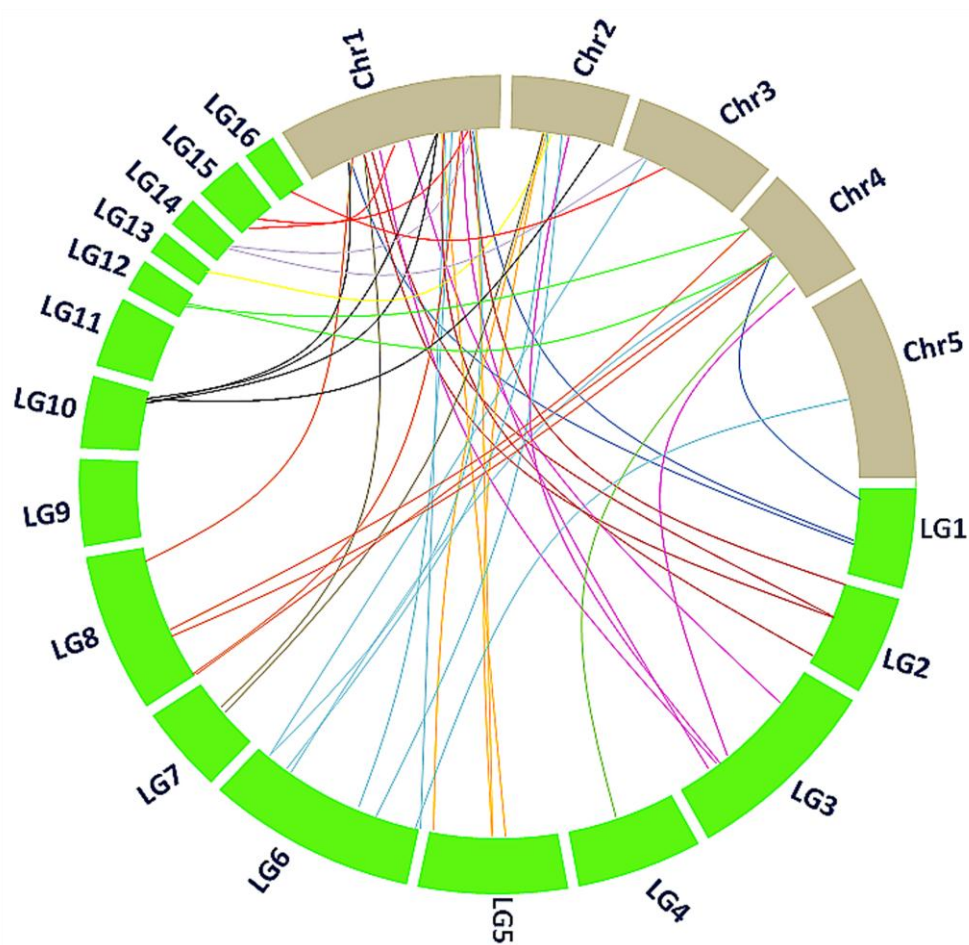
	LOC105175707	12	5079580	5079898	Potassium channel KAT3-like
	LOC105165006	6	18910517	18910130	Abscisic acid 8'-hydroxylase 1-like
	LOC105157462	1	530090	530409	Abscisic acid 8'-hydroxylase 1-like
	LOC105160523	4	10290113	10291144	DREB
	LOC105158557	3	16161829	16162134	Phosphoenolpyruvate carboxykinase [ATP]-like
	LOC105156478	2	16843822	16844116	Alpha-1,4-glucan-protein synthase
	LOC105177713	15	3453387	3454069	Oxygen-evolving enhancer protein 1
	LOC105167477	8	2684789	2685468	Oxygen-evolving enhancer protein 1
	LOC105165268	1	12691700	12692388	Heat shock 70 kDa protein-like
	LOC105155749	2	10818136	10818702	40S ribosomal protein S2-4-like
	LOC105162725	5	16732816	16733391	40S ribosomal protein S2-4-like
	LOC105171901	10	4259842	4260638	Heat shock cognate 70 kDa
	LOC105177235	15	867494	868130	Tubulin alpha chain
	LOC105170695	9	5296418	5296884	Elongation factor 1-alpha
	LOC105160899	4	13022275	13022673	60S acidic ribosomal protein P0
<b>Potato</b>	LOC105174800	12	387677	387970	60S ribosomal protein L8-1
	LOC105176744	14	1569814	1570107	60S ribosomal protein L8-1
	LOC105170962	9	7060127	7060576	Ethylene insensitive 3
	LOC105157101	3	2740513	2741378	Geranylgeranyl diphosphate reductase
	LOC105168390	8	12427153	12427919	NADH dehydrogenase Flavoprotein 1
	LOC105167993	8	8776046	8776600	Aminomethyltransferase
	LOC105168993	8	16465020	16465871	S-adenosylmethionine synthase 1
	LOC105172459	10	13464015	13464776	S-adenosylmethionine synthase 1-like
	LOC105174259	11	13028743	13029504	S-adenosylmethionine synthase 3
	LOC105174436	11	13956861	13957724	heat shock cognate 70 kDa protein 2
	LOC105161464	5	777968	778829	heat shock cognate 70 kDa protein 2
	LOC105176719	14	1410275	1412327	Polyubiquitin
	LOC105165880	7	15613	16461	Unknow function
	LOC105165878	7	5147	6771	Unknow function
	LOC105158900	3	19877583	19881182	Protein hypoxanthine-guanine phosphoribosyl transferase
	LOC105161893	5	3608867	3609445	Senescence-specific cysteine protease SAG39-like
<b>Tomato</b>	LOC105166264	7	4914769	4915824	Abscisic acid-insensitive 5
	LOC105159163	3	22078807	22079856	Abscisic acid-insensitive 5
	LOC105174582	11	14789561	14790058	Monothiol glutaredoxin-S16
	LOC105165958	7	2367785	2368681	Thioredoxin-like protein CDSP32
	LOC105157874	3	9547915	9547019	ERF RAP2-12-like
	LOC105162917	1	11232516	11233406	ERF RAP2-12-like



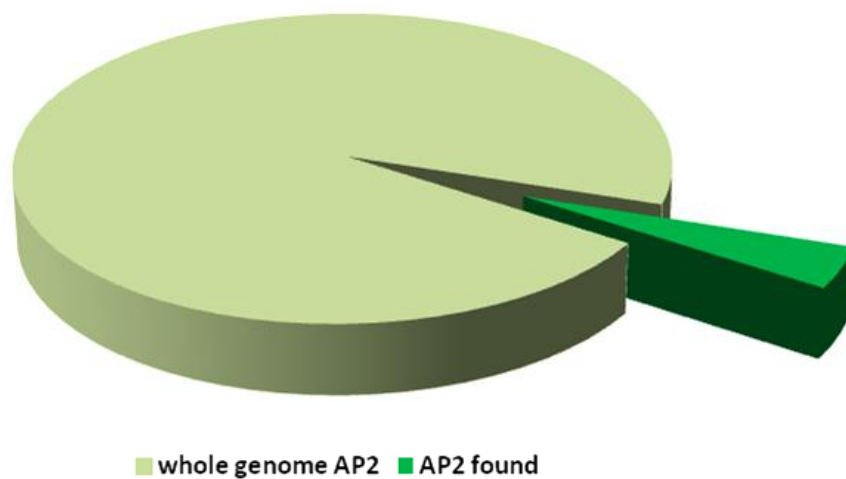
**Figure 2.** Distribution of the candidate genes on the sesame linkage groups. LG1~LG16 represent linkage groups of the sesame genome. Locus names in green, blue and red indicate orthologous genes of *Arabidopsis*, tomato and potato, respectively.

ignored. An expression profiling of these AP2/ERF genes under drought stress in two contrasted sesame lines through RT-PCR were further performed. Gene expression patterns are usually closely correlated with their functions (Peng et al., 2015). One primer pair (SSAp5) designed for the gene LOC105157874 did not amplify any of the two accessions suggesting probably an inadequate primer. Three out of the four remaining genes expressed highly in the drought tolerant material compared to the sensitive one under drought stress (Figure 5). The expression level of the gene LOC105160523 was more striking in drought tolerant material compared to the sensitive one whereas the gene LOC105162917 showed similar expression pattern under both water regimes. The BLASTp search against

*Arabidopsis* genome showed that the highest expressed gene LOC105160523 is the orthologs of CBF4 (*AT5G51990*) described as regulator of drought adaptation in *Arabidopsis* (Haake et al., 2002) suggesting that this gene plays a pivotal role in drought tolerance in sesame. Since, few sesame accessions have been used in this study, it was proposed that these genes should be more deeply studied on a large sample of contrasted materials to uncover their biological roles in drought tolerance in sesame. Our results corroborate well with that of Kamvysselis (2003), who reported that comparative genomics analysis can reveal biological findings that could not have been discovered by traditional genetic methods, regardless of the time or effort spent.

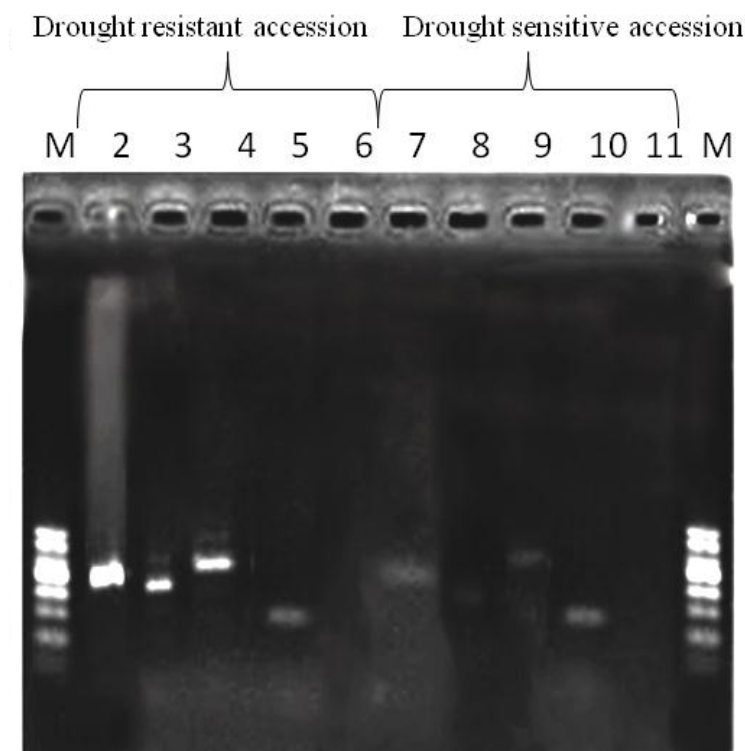


**Figure 3.** Syntenic relationships of drought associated genes between *Arabidopsis* and sesame genomes. Chr1–Chr5 represent pseudo-chromosomes of *Arabidopsis* genome and represented by gray bars. LG01–LG16 represent linkage groups of sesame genome and drawn in green bars. Colorful lines stand for the relationships of orthologous gene pairs between the two species.



**Figure 4.** AP2/ERF transcription factor genes found within the candidate genes compared to whole genome AP2/ERF genes.





**Figure 5.** Expression profile of five AP2/ERF genes in two contrasted sesame accessions under drought stress. M represents the migration profile of the ladder; 2 and 7- SSAp3; 3 and 8-SSAp3, 4 and 9- SSAp1, 5 and 10- SSAp4, 6 and 11- SSAp5.

Sesame is an oil crop that contributes to the daily oil and protein requirements of almost half of the world's population (Wei et al., 2015). One of the major constraints for its production is drought as it is mainly grown in semi-arid areas. Functions of most sesame genes are still uncharacterized. Hence, the identification and functional analysis of valuable genes in sesame genome is necessary for its improvement. Since reports on drought associated genes in sesame are lacking, this study provided, a set of candidate genes spanning the whole genome and including different functional genes for drought research in sesame using comparative genomic approach. Further and thorough functional experiments including transgenic studies could rely on these gene resources to validate their functions and decipher the mechanisms of drought tolerance in sesame.

#### Conflict of interests

The authors have not declared any conflict of interests.

#### ACKNOWLEDGEMENTS

The authors thank Kyky Komla Ganyo for his helpful

suggestions on the manuscript. This study was partially financially supported by the West African Research Association travel grant accorded to the first author.

#### REFERENCES

- Abe H, Yamaguchi-Shinozaki K, Urao T, Iwasaki T, Hosokawa D, Shinozaki K (1997). Role of *Arabidopsis* MYC and MYB homologs in drought and abscisic acid-regulated gene expression. *Plant Cell* 9:1859-1868.
- Alter S, Bader KC, Spannagl M, Wang Y, Bauer E, Schön CC, Mayer KFX (2015). DroughtDB: An expert-curated compilation of plant drought stress genes and their homologs in nine species. *Database* pp. 1-7.
- Anithakumari AM, Dolstra O, Vosman B, Visser RGF, van der Linden CG (2011). In vitro screening and QTL analysis for drought tolerance in diploid potato. *Euphytica* 181:357-369.
- Anithakumari AM, Nataraja KN, Visser RGF, Van Der Linden CG (2012). Genetic dissection of drought tolerance and recovery potential by quantitative trait locus mapping of a diploid potato population. *Mol. Breed.* 30:1413-1429.
- Ashutosh SB, Shinada T, Kifuji Y, Kitashiba H, Nishio T (2012). Molecular mapping of a male fertility restorer locus of *Brassica oleracea* using EST-based SNP markers and analysis of a syntenic region in *Arabidopsis thaliana* for identification of genes encoding PPR proteins. *Mol. Breed.* 30:1781-1792.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.
- Bahrami H, Razmjoo J, Jafari AO (2012). Effect of drought stress on

- germination and seedling growth of sesame cultivars (*Sesamum indicum* L.). *Int. J. Agric. Sci.* 2(5):423-428.
- Balasuubramanian P, Palaniappan SP (2001). Field Crops: An overview. Principles and Practices of Agronomy, Agrobios, India. P 47.
- Bartels D, Sunkar R (2005). Drought and salt tolerance in plants. *Cr. Rev. Plant Sci.* 24:23-58.
- Betram K, Janssens MJJ, Abdalwahab A (2003). Breeding for drought tolerance in sesame (*Sesamum indicum*). Conference on Technological and Institutional Innovations for Sustainable Rural Development, 8–10 October, Deutscher Tropentag, Göttingen, P 135.
- Bisht NC, Gupta V, Ramchiary N, Sodhi YS, Mukhopadhyay A, Arumugam N, Pental D, Pradhan AK (2009). Fine mapping of loci involved with glucosinolate biosynthesis in oilseed mustard (*Brassica juncea*) using genomic information from allied species. *Theor. Appl. Genet.* 118:413-421.
- Boureima S, Oukarroum A, Diouf M, Cissé N, Van Damme P (2012). Screening for drought tolerance in mutant germplasm of sesame (*sesamum indicum*) probing by chlorophyll a fluorescence. *Environ. Exp. Bot.* 81:37-43.
- Boutanaev AM, Kalmykova AI, Shevelyov YY, Nurminsky DI (2002). Large clusters of co-expressed genes in the *Drosophila* genome. *Nat.* 420:666-669.
- Cattivelli L, Rizza F, Badeck FW, Mazzucotelli E, Mastrangelo AM, Francia E, Marè C, Tondelli A, Stanca AM (2008). Drought tolerance improvement in crop plants: An integrated view from breeding to genomics. *Field Crops Res.* 105:1-14.
- Chiba H, Yamashita R, Kinoshita K, Nakai K (2008). Weak correlation between sequence conservation in promoter regions and in protein-coding regions of human-mouse orthologous gene pairs. *BMC Genomics* 9:152.
- Cohen BA, Mitra RD, Hughes JD, Church GM (2000). A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nat. Genet.* 26:183-186.
- Evers D, Lefevre I, Legay S, Lamoureux D, Hausman JF, Rosales RO, Marca LR, Hoffmann L, Bonierbale M, Schaffleitner R (2010). Identification of drought-responsive compounds in potato through a combined transcriptomic and targeted metabolite approach. *J. Exp. Bot.* 61:2327-2343.
- FAOSTAT Food and Agriculture Organization (2013). Available online: <http://faostat.fao.org/>. (Accessed 2015-12-5).
- Fernandez-Pozo N, Menda N, Edwards JD, Saha S, Teclé IY, Strickler SR, Bombarely A, Fisher-York T, Pujar A, Foerster H, Yan A, Mueller LA (2015). The Sol Genomics Network (SGN)--from genotype to phenotype to breeding. *Nucleic Acids Res.* 43:1036-1041.
- Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M (2014). The Pfam protein families database. *Nucleic Acids Res.* 42:D222-D230.
- Foucher F, Morin J, Courtiade J, Cadioux S, Ellis N, Banfield MJ, Rameau C (2003). Determinate and late flowering are two terminal flower1/centro-radialis homologs that control two distinct phases of flowering initiation and development in pea. *Plant Cell* 15:2742-2754.
- Guyot R, Lefebvre-Pautigny F, Tranchant-Dubreuil C, Rigoreau M, Hamon P, Leroy T, Hamon S, Poncet V, Crouzillat D, de Kochko A (2012). Ancestral synteny shared between distantly-related plant species from the asterid (*Coffea canephora* and *Solanum Sp.*) and rosid (*Vitis vinifera*) clades. *BMC Genomics* 13:103.
- Haake V, Cook D, Riechmann J, Pineda O, Thomashow MF, Zhang JZ (2002). Transcription factor CBF4 is a regulator of drought adaptation in *Arabidopsis*. *Plant Physiol.* 130:639-648.
- Harshavardhan VT, Son LV, Seiler C, Junker A, Weigelt-Fischer K, Klukas C, Altmann T, Sreenivasulu N, Bäumlein H, Kuhlmann M (2014). AtRD22 and AtUSPL1, members of the plant-specific BURP domain family involved in *Arabidopsis thaliana* drought tolerance. *PLoS ONE* 9(10):e110065.
- Hassanzadeh M, Asghari A, Jamaati-e-Somarin Sh, Saeidi M, Zabihi-e-Mahmoodabad R, Hokmalipour S (2009). Effects of water deficit on drought tolerance indices of sesame (*Sesamum indicum* L.) genotypes in Moghan Region. *Res. J. Environ. Sci.* 3:116-121.
- Huala E, Dickerman AW, Garcia-Hernandez M, Weems D, Reiser L, LaFond F, Hanley D, Kiphart D, Zhuang M, Huang W, Mueller LA, Bhattacharyya D, Bhaya D, Sobral BW, Beavis W, Meinke DW, Town CD, Somerville C, Rhee SY (2001). The *Arabidopsis* Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis and visualization system for a model plant. *Nucleic Acids Res.* 29:102-105.
- IPCC (2007). Climate Change 2007: Contribution of working group to the fourth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge.
- Jang JY, Kim DG, Kim YO, Kim JS, Kang H (2004). An expression analysis of a gene family encoding plasma membrane aquaporins in response to abiotic stresses in *Arabidopsis thaliana*. *Plant Mol. Biol.* 54:713-725.
- Jang JY, Lee SH, Rhee JY, Chung GC, Ahn SJ, Kang H (2007). Transgenic *Arabidopsis* and tobacco plants overexpressing an aquaporin respond differently to various abiotic stresses. *Plant Mol. Biol.* 64:621-632.
- Kamvysseis M (2003). Computational comparative genomics: genes, regulation, evolution. PhD Thesis, Massachusetts Institute of Technology. [http://compbio.mit.edu/publications/M02\\_Kellis\\_Thesis\\_03.pdf](http://compbio.mit.edu/publications/M02_Kellis_Thesis_03.pdf).
- Krannich CT, Maletzki L, Kurowsky C, Horn R (2015). Network candidate genes in breeding for drought tolerant crops. *Int. J. Mol. Sci.* 16:16378-16400.
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19(9):1639-1645.
- Langham DR (2007). "Phenology of sesame," In: J. Janick and A. Whipkey (ed.), *Issues in New Crops and New Uses*, ASHS Press, Alexandria, VA, Pp. 144-182.
- Lata C, Mishra AK, Muthamilarasan M, Bonthala VS, Khan Y, M Prasad (2014). Genome-wide investigation and expression profiling of ap2/erf transcription factor superfamily in foxtail millet (*Setaria italica* L.). *PLoS ONE* 9(11):e113092.
- Lata C, Yadav A, Prasad M (2011). Role of plant transcription factors in abiotic stress tolerance, abiotic stress response in plants - physiological, biochemical and genetic perspectives, Prof. Arun Shanker (Ed.), ISBN: 978-953-307-672-0, InTech, Available from: <http://www.intechopen.com/books/abiotic-stress-response-in-plants-physiological-biochemical-and-genetic-perspectives/role-of-plant-transcription-factors-in-abiotic-stress-tolerance>.
- Lee S, Kohane I, Kasif S (2005). Genes involved in complex adaptive processes tend to have highly conserved upstream regions in mammalian genomes. *BMC Genomics* 6:168.
- Li C, Yan J-M, Li Y-Z, Zhang Z-C, Wang Q-L, Liang Y (2013). Silencing the SpMPK1, SpMPK2 and SpMPK3 genes in tomato reduces abscisic acid-mediated drought tolerance. *Int. J. Mol. Sci.* 14:21983-21996.
- Licausi F, Giorgi FM, Zenoni S, Osti F, Pezzotti M, Perata P (2010). Genomic and transcriptomic analysis of the AP2/ERF superfamily in *Vitis vinifera*. *BMC Genomics* 11:719.
- Ma Y, Qin F, Tran LSP (2012). Contribution of genomics to gene discovery in plant abiotic stress responses. *Mol. Plant* 5(6):1176-1178.
- McWilliam J (1989). The dimensions of drought, in drought resistance in cereals (Baker F, ed.), CAB International, Wallingford, UK. Pp. 1-11.
- Mochida K, Yoshida T, Sakurai T, Ogihara Y, Shinozaki K (2009). TriFLDB: a database of clustered full-length coding sequences from Triticeae with applications to comparative grass genomics. *Plant Physiol.* 150:1135-1146.
- Mondini L, Nachit MM, Pagnotta MA (2015). Allelic variants in durum wheat (*Triticum turgidum* L. var. durum) DREB genes conferring tolerance to abiotic stresses. *Mol. Genet. Genomics* 290:531-544.
- Mondini L, Nachit M, Porceddu E, Pagnotta MA (2012). Identification of SNP mutations in DREB1, HKT1 and WRKY1 genes involved in drought and salt stress tolerance in durum wheat (*Triticum turgidum* L. var durum). *OMICS* 16(4):178-187.
- Nakashima K, Yamaguchi-Shinozaki K, Shinozaki K (2014). The

- transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses including drought, cold, and heat. *Front. Plant Sci.* 5:170.
- Nayar NM, Mehra KL (1970). Sesame. Its uses, botany, cytogenetics and origin. *Econ. Bot.* 24:20-31.
- Obidiegwu JE, Bryan GJ, Jones HG, Prashar A (2015). Coping with drought: stress and adaptive responses in potato and perspectives for improvement. *Front. Plant Sci.* 6:542.
- Okonechnikov K, Golosova O, Fursov M (2012). UGENE team Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* 28:1166-1167.
- Pathak N, Rai AK, Kumari R, Thapa A, Bhat KV (2014). Sesame crop: an underexploited oilseed holds tremendous potential for enhanced food value. *Agric. Sci.* 5:519-529.
- Peng X, Zhao Y, Li X, Wu M, Chai W, Sheng L, Wang Y, Dong Q, Jiang H, Cheng B (2015). Genome wide identification, classification and analysis of NAC type gene family in maize. *J. Genet.* 94:377-390.
- Pruitt KD, Tatusova T, Maglott DR (2007). NCBI reference sequences (RefSeq): a curated nonredundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 35:D61-D65.
- Rabara RC, Tripathi P, Rushton PJ (2014). The Potential of transcription factor-based genetic engineering in improving crop tolerance to drought. *OMICS* 18(10):601-614.
- Ramirez V, Coego A, Lopez A, Agorio A, Flors V, Vera P (2009). Drought tolerance in *Arabidopsis* is controlled by the OCP3 disease resistance regulator. *Plant J.* 58:578-591.
- Reiter WD, Vanzin GF (2001). Molecular genetics of nucleotide sugar interconversion pathways in plants. *Plant Mol. Biol.* 47:95-113.
- Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D, Sherman BK, Yu G (2000). *Arabidopsis* transcription factor: genome wide comparative analysis among eukaryotes. *Sci.* 290:2105-2110.
- Roy S, Chauhan R, Maheshwari N, Gupta S, Gupta DK, Sharma A (2011). In silico approaches in comparative genomics, structure prediction and functional characterization of secondary metabolite proteins of *Mentha* sp. *OMICS* 4(7):354-363.
- Rubin GM (2001). The draft sequences: Comparing species. *Nat.* 409:820-821.
- Sakuma Y, Liu Q, Dubouzet JG, Abe H, Shinozaki K, Yamaguchi-Shinozaki K (2002). DNA-binding specificity of the ERF/AP2 domain of *Arabidopsis* DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. *Biochem. Biophys. Res. Commun.* 290:998-1009.
- Sakuma Y, Maruyama K, Osakabe Y, Qin F, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2006). Functional analysis of an *Arabidopsis* transcription factor, DREB2A, involved in drought-responsive gene expression. *Plant Cell* 18:1292-1309.
- Schranz ME, Quijada P, Sung SB, Lukens L, Amasino RM, Osborn TC (2002). Characterization and effects of the replicated time gene *FLC* in *Brassica rapa*. *Genet.* 162:1457-1468.
- Seki M, Kamei A, Yamaguchi-Shinozaki K, Shinozaki K (2003). Molecular responses to drought, salinity and frost: Common and different paths for plant protection. *Cr. Rev. Plant Sci.* 14:194-199.
- Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y, Kamiya A, Nakajima M, Enju A, Sakurai T, Satou M, Akiyama K, Taji T, Yamaguchi-Shinozaki K, Carninci P, Kawai J, Hayashizaki Y, Shinozaki K (2002). Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *Plant J.* 31:279-292.
- Shinozaki K, Yamaguchi-Shinozaki K, Seki M (2003). Regulatory network of gene expression in the drought and cold stress responses. *Curr. Opin. Plant Biol.* 6:410-417.
- Soares-Cavalcanti NM, Belarmino LC, Kido EA, Wanderley-Nogueira AC, Bezerra-Neto JP, Cavalcanti-Lira R, Pandolfi V, Nepomuceno AL, Abdelnoor RV, Nascimento LC, Benko-Iseppon AM (2012). In silico identification of known osmotic stress responsive genes from *Arabidopsis* in soybean and Medicago. *Genet. Mol. Biol.* 35:315-321.
- Solankey SS, Singh RK, Baranwal DK, Singh DK (2014). Genetic expression of tomato for heat and drought stress tolerance: An overview. *Int. J. Veg. Sci.* 21(5):496-515.
- Suwabe K, Tsukazaki H, Iketani H, Hatakeyama K, Kondo M, Fujimura M, Nunome T, Fukuoka H, Hirai M, Matsumoto S (2006). Simple sequence repeat-based comparative genomics between *Brassica rapa* and *Arabidopsis thaliana*: the genetic origin of clubroot resistance. *Genet.* 173:309-319.
- Tamura K, Stecher G, Peterson D, Filipksi A, Kumar S (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.* 30:2725-2729.
- Vasquez-Robinet C, Mane SP, Ulanov AV, Watkinson JI, Stromberg VK, De Koeber D, et al. (2008). Physiological and molecular adaptations to drought in Andean potato genotypes. *J. Exp. Bot.* 59:2109-2123.
- Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Hered.* 93:77-78.
- Wang L, Yu S, Tong C, Zhao Y, Liu Y, Song C, Zhang Y, Zhang X, Wang Y, Hua W, Li D, Li D, Li F, Yu J, Xu C, Han X, Huang S, Tai S, Wang J, Xu X, Li Y, Liu S, Varshney RK, Wang J, Zhang X (2014). Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. *Genome Biol.* 15:R39.
- Wei X, Liu K, Zhang Y, Feng Q, Wang L, Zhao Y, Li D, Zhao Q, Zhu X, Zhu X, Li W, Fan D, Gao Y, Lu Y, Zhang X, Tang X, Zhou C, Zhu C, Liu L, Zhong R, Tian Q, Wen Z, Weng Q, Han B, Huang X, Zhang X (2015). Genetic discovery for oil production and quality in sesame. *Nat. Com.* 6:8609.
- Wei X, Wang L, Yu J, Zhang Y, Li D, Zhang X (2015). Genome-wide identification and analysis of the MADS-box gene family in sesame. *Gene* 569:66-76.
- Windsor AJ, Mitchell-Olds T (2006). Comparative genomics as a tool for gene discovery. *Curr. Opin. Biotech.* 17:161-167.
- Witcombe JR, Hollington PA, Howarth CJ, Reader S, Steele KA (2007). Breeding for abiotic stresses for sustainable agriculture. *Philos. Trans. R. Soc. B.* 363:703-716.
- Zhang Y, Wang L, Xin H, Li D, Ma C, Ding X, Hong W, Zhang X (2013). Construction of a high-density genetic map for sesame based on large scale marker development by specific length amplified fragment (SLAF) sequencing. *BMC Plant Biol.* 13:141.