International Journal of Ecology and Environmental Sciences 50: 645-657, 2024
ISSN: 2320-5199 (Online); https://doi.org/10.55863/ijees.2024.0316

Comparative Analysis of Sequence, Chromosomal Arrangement and *Cis*-Regulatory Elements of *SQS* Genes across Brassicaceae

RAJKUMAR BANGKIM, KOMAL CHAUDHARY AND PRIYA PANJABI*
Department of Botany, University of Delhi, Delhi, 110007, India
E-mail: bangkim3@gmail.com, komal4chaudhary@gmail.com, ppanjabi09@gmail.com
***Author for correspondence**

ABSTRACT
Plant sterols are known to be involved in plant growth and development and also in plants' response to stress conditions. Squalene synthase (SQS) is an important enzyme for the biosynthesis of sterols and triterpenoids. Owing to their crucial role in plants, *SQS* genes have been identified and characterized in several plants, however not much is known about their *cis*-regulatory regions. In our previous study, we identified Brassicaceae specific duplications of the gene, providing an ideal opportunity to explore the diversification of the regulatory regions across the gene copies. In this study, we identified 49 *SQS* genes from 12 species of Brassicaceae and analyzed their genomic distribution, physicochemical properties, subcellular localization, *cis*-regulatory element (CRE) landscape and sequence evolution (Ka/Ks ratios). The SQS proteins were predicted to be highly stable, dominated by aliphatic amino acids, hydrophilic in nature and localized within the endoplasmic reticulum. *In-silico* promoter analysis identified 50 types of CREs with the predominance of light-responsive (22), followed by hormone-responsive (9), stress-responsive (7) and plant growth and development (9) elements. We identified differences both in the type and distribution of CREs across the SQS duplicates suggestive of regulatory divergence of the paralogs. The Ka/Ks analysis depicted that all the *SQS* genes are under purifying selection pressure. The information generated in this study would provide a useful resource for further investigations in *SQS* genes in Brassicaceae.

Key words: Brassicaceae, *Cis*-regulatory elements, Evolutionary analysis, Gene duplication, Phytosterols, Squalene synthase (SQS), Environmental stress regulation

INTRODUCTION

Phytosterols, or plant sterols, are important components of plant cell membranes and play a crucial role in diverse physiological processes during plant growth and development. They are also known to be involved in plant responses towards abiotic (cold, salt, and heat) as well as biotic (pests and pathogens) stress (Wang et al. 2012a, b, Chalbi et al. 2015, Singh et al. 2015, Chen et al. 2020, Wei et al. 2020, Rossi and Huang 2022). Moreover, phytosterols are also important components of human diet owing to their cholesterol reduction, anti-cancer, anti-diabetic, antioxidant, and anti-inflammatory properties (López-García et al. 2017, 2020, Poulose et al. 2021, Essadek et al. 2023, Liu et al. 2023).

Squalene synthase (SQS), a membrane-bound enzyme involved in phytosterol biosynthesis, plays a central role in catalyzing the conversion of farnesyl diphosphate (FPP) to squalene, an intermediate in the biosynthesis of sterols (Rilling and Epstein 1969, Poulter et al. 1990). *SQS* genes have been functionally characterized from a wide variety of organisms, including bacteria, fungi, animals, green algae, lycophyte and seed plants (Jennings et al. 1991, McKenzie et al. 1992, Summers et al. 1993, Devarenne et al. 1998, Okada et al. 2000, Busquets et al. 2008, Lee and Poulter 2008, Jiang et al. 2015).

Several studies have shown the involvement of *SQS* genes in regulating plant responses to various abiotic and biotic stresses. For example, *SQS* gene expression was reported to increase in *Glycyrrhiza glabra* and *Trigonella foenum-graecum* when subjected to salt and temperature stress conditions respectively (Shirazi et al. 2019, Sheikhi et al. 2023). Transgenic *Arabidopsis* lines overexpressing *SQS* gene from *Torreya grandis* were found to exhibit enhanced drought tolerance (Nasrollahi et al. 2014). Further, silencing of this gene in *Nicotiana benthamiana*, *A. thaliana* and *Withania somnifera* was found to increase susceptibility towards various pathogens including bacteria, fungi and insects (Wang et al. 2012b, Singh et al. 2015).

SQS genes have also been observed to be responsive to phytohormones. Significant enhancement in the gene expression within roots has been reported upon treatment with phytohormones specially methyl jasmonate (MeJA) in plant species, *Panax ginseng* and *Bupleurum falcatum* (Kim et al. 2011a, b). In a few isolated studies, these findings have been corroborated with the presence of stress responsive elements in the promoter regions of *SQS* genes. The promoter regions of *SQS* genes from S*anghuangporus baumii, P. ginseng* and *W. somnifera* were found to contain *cis*-regulatory elements (CREs) involved in light, hormone (abscisic acid and MeJA) and stress responsiveness (Kim et al. 2011a, Bhat et al. 2012, Wang et al. 2022).

In a recent study (Hazra et al. 2023), the copy number of the *SQS* gene was found to be highly variable (1-10) across plants, particularly within the family Brassicaceae. In a parallel study dealing with phylogenetic analysis of *SQS* genes across land plants, we identified family specific duplication events that occurred in the ancestor leading to Brassicaceae. Owing to these duplications the *SQS* genes within Brassicaceae were found to cluster into three clades; SQS1, SQS2 and SQS3 (unpublished). Gene duplication events provide raw material for the evolutionary processes. The newly formed duplicates can over a period of time acquire new functions (neo-functionalisation) or have altered expression patterns (sub-functionalisation). This divergence can be owing to accumulation of mutations in either the coding or the upstream (regulatory) sequences (Arsovski et al. 2015). The family Brassicaceae is marked by whole genome duplication events and thereby an excellent model for studying polyploidy and evolutionary processes (Kagale et al. 2014).

Although, a high level of conservation is observed across the coding region of *SQS* genes, which corresponds to its presence across all the domains of life (Robinson et al. 1993, Rong et al. 2016, Zhang et al. 2018), very little is known about the *cis*-regulatory regions which can provide a glimpse into the regulation of these gene copies within and across species. Comparative studies focusing on the CRE landscapes of the duplicate gene copies enables insights into the evolution of gene expression patterns.

In this study, we identified 49 *SQS* genes from twelve Brassicaceae species. We identified the *cis*-regulatory elements of the retrieved *SQS* genes and analyzed the evolutionary selection pressure (Ka/Ks) and physicochemical properties of the SQS protein sequences. *SQS* gene upstream sequences were found to contain CREs for plant growth and development, light regulation, promoter binding sites, hormone and stress responsiveness. Comparative analyses of the CRE landscape across the three Brassicaceace specific *SQS* paralogs identified some paralog specific CRE distribution patterns that are suggestive of regulatory divergence of the gene copies. This, to the best of our knowledge, is the first comprehensive analysis of the CRE landscape of *SQS* genes. The Ka/Ks analysis revealed that the *SQS* genes are under purifying selection pressure. The results reported in this study will benefit further studies on the evolutionary and functional analysis of *SQS* genes.

MATERIALS AND METHODS

Identification, genomic location and physicochemical characteristics of SQS homologs

SQS protein sequence from *A. thaliana* (*AtSQS1*) was obtained from TAIR10 (The *Arabidopsis* Information Resource, https://www.arabidopsis.org, Rhee et al. 2003). Using *AtSQS1* protein sequence as query, SQS homologs from 11 species of Brassicaceae (*A. thaliana*, *A. halleri*, *A. lyrata*, *Capsella rubella*, *Thellungiella halophila*, *Camelina sativa, Brassica rapa*, *B. oleracea*, *B. napus*, *B. nigra, B. carinata*) were identified from databases: Phytozome-version13 (https://phytozome.jgi.doe. gov/pz/portal.html, Goodstein et al. 2012), NCBI (National Centre for Biotechnology Information, https://www.ncbi.nlm.nih.gov) and BRAD (http://brassicadb.cn, Chen et al. 2022) through BLASTP searches (at default parameters). The physical and chemical characteristics of the retrieved SQS proteins were estimated using the Expasy Prot-Param tool (http://weB.expasy.org/protparam/), including sequence length, molecular weight, theoretical isoelectric point, instability index, aliphatic index, grand average of hydropathicity. The subcellular localization of SQS proteins were predicted using the online ProtComp v. 9.0 program (http://www.softberry.com).

The genomic location of each *SQS* gene was obtained from the respective genome databases from where they were retrieved. Subsequently, the MG2C

program (http://mg2c.iask.in/mg2c_v2.0/, Chao et al. 2021) was employed to map the distribution of *SQS* genes on the respective chromosomes.

Cis-regulatory element analysis

For *cis*-regulatory element identification, 1500 bp 5' upstream sequences from the start codon were retrieved from each *SQS* gene of the Brassicaceae species. The online program PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html, Lescot et al. 2002) was then used to identify the types and number of CREs retrieved. The results were visualized with TB-tools software (Chen et al. 2020).

Estimation of evolutionary selection pressure

The ratios of nonsynonymous (Ka) to synonymous (Ks) substitution rate (Ka/Ks) for all coding sequences of *SQS* genes were calculated with the simple Ka/Ks calculator in TBtools software (Chen et al. 2020). These comparisons were made in three ways. Firstly, the Ka/Ks ratios were calculated between tandemly duplicated gene pairs for each species. Secondly, the Ka/Ks values were calculated between each of the *SQS* genes of the Brassicaceae species and the *A. thaliana* genes; *AtSQS1* and *AtSQS2*, used as a reference one at a time. Lastly, pairwise Ka/Ks ratios were also calculated for genes belonging to each of the three clades (SQS1, SQS2 and SQS3) separately.

RESULTS AND DISCUSSION

Identification and physicochemical properties of SQS proteins

A total of 47 SQS protein homologs were identified from eleven Brassicaceae species which included *A. thaliana*, *A. lyrata*, *A. halleri*, *C. rubella*, *C. sativa*, *T. halophila*, *B. rapa*, *B. nigra*, *B. oleracea*, *B. napus and B. carinata* (Supplementary Table 1**)**. In addition, two *B. juncea* SQS homologs (present on B07 and A08 chromosomes) retrieved from the NCBI were also included. Details of the species names and the total number of SQS orthologs identified have been provided in Table 1.

For each of the genes, the genomic location, length of the coding sequence and protein properties [amino acid lengths, molecular weight, isoelectric point, instability index, aliphatic index, grand average of hydropathicity (GRAVY) and subcellular locations] were estimated and have been provided in Table 1. One of the genes, *Bol026669*, was truncated and predicted to code for 168 amino acid long protein with a molecular weight of 18.68 kDa. The length of coding sequence (CDS) of the remaining 48 *SQS* genes ranged from 963-1308 bp. The estimated molecular weight ranged from 36.84 to 50.48 kDa with amino acid lengths of 322-435 and the isoelectric point of 5.39 to 8.18. The instability index ranged from 30.16 to 44.19 with majority having values less than 40 suggestive of their being stable proteins. In addition, the aliphatic indices ranged from 91.42 to 108.36, indicating that most SQSs contained large amounts of aliphatic amino acids. GRAVY values of 89% of SQS proteins were less than 0, indicating their hydrophilic nature. All SQS proteins were predicted to be localized within the endoplasmic reticulum.

Genomic distribution of *SQS* genes across Brassicaceae

Most Brassicaceae species were found to have more than two copies of the *SQS* gene. The genomic distribution of each identified *SQS* gene was mapped to the specific chromosomes/scaffolds by using MG2C online tool (http://mg2c.iask.in/mg2c_v2.0, Chao et al. 2021) (Fig. 1). *A. thaliana* is known to contain two copies of the gene present in tandem on chromosome 4 (Chr 4) (Busquets et al. 2008). In *A. lyrata,* 3 genes were identified on two scaffolds, while *A. halleri* contains four genes, three on Chr 7 and one on Chr 3.

In our previous study (unpublished results), we had identified the presence of tandemly duplicated *SQS* genes in most Brassicaceae species. We observed this pattern of duplication in all the additional Brassicaceae species analyzed in this study. Both *C. rubella* and *T. halophila* have a pair of tandemly placed *SQS* genes on scaffold 7 and scaffold 1, respectively. *C. sativa* has three pairs of tandem *SQS* genes located on three chromosomes; Chr 10, Chr 11 and Chr 12.

The diploid species, *B. oleracea*, *B. rapa* and *B. nigra* contain 3, 4 and 6 copies of the gene distributed on 2, 2 and 4 chromosomes each respectively (Fig. 1). The polyploid species analyzed contained more copies of the *SQS* gene as expected. The genome of *B. napus* contains nine *SQS* genes distributed across

Table 1. List and characteristics of *SQS* genes identified from Brassicaceae species. For each species, the total number of genes identified, gene ID, genomic location and protein properties [length of coding region (CDS), total amino acids (aa) in the protein, molecular weight (MW), isoelectric point (pI), instability index (II), aliphatic index (AI) and grand average of hydropathicity (GRAVY)] have been provided.

Species	Total genes	Gene ID	Chromosomal location	CDS (bp)	aa	MW (kDa)	pI	II	AI	GRAVY
Arabidopsis thaliana	2	AT4G34640	Chr4:16538281..16541926 (+)	1233	410	47.14	6.18	43.28	93.93	-0.13
		AT4G34650	Chr4:16542376..16545121 (+)	1242	413	47.18	8.13	35.87	101.5	-0.16
Arabidopsis halleri	4	Ah7G04360	Chr7:2038090..2041711 (-)	1239	412	47.32	6.57	40.27	94.17	-0.14
		Ah7G04350	Chr7:2031126..2034312 (-)	1242	413	47.07	7.12	34.39	101.72	-0.14
		Ah7G04370	Chr7:2035201..2037603 (-)	1128	375	42.89	6.24	35.27	98.27	-0.14
		Ah3G24480	Chr3:9483767..9486460 (-)	1308	435	50.46	6.89	35.87	96.53	-0.14
Arabidopsis lyrata	3	AL7G16560	Scaffold7:2687881..2691426 (-)	1239	412	47.28	6.23	39.45	95.12	-0.14
		AL7G16550	Scaffold7:2684594..2687666 (-)	1242	413	47.02	7.13	35.48	101.96	-0.14
		AL3G33820	Scaffold3:8803195..8806230 (-)	1308	435	50.48	7.93	37.09	96.99	-0.12
Capsella rubella	2	Carub.0007s0564	Scaffold7:2122676..2126268 (-)	1245	414	47.49	6.18	39.94	93.96	-0.12
		Carub.0007s0563	Scaffold7:2119645..2122138 (-)	1242	413	47.36	6.83	35.75	101.48	-0.12
Thellungiella halophila	2	Thhalv10025338m.g	Scaffold_1:2285479..2289293 (-)	1230	409	46.94	6.6	42.08	95.35	-0.12
		Thhalv10025392m.g	Scaffold_1:2282612..2285105 (-)	1191	396	45.24	7.57	36.01	104.12	-0.11
Camelina sativa	6	LOC104716666	Chr10: 2699163..2702001 (-)	1242	413	47.23	6.61	38.45	104.55	-0.11
		LOC104716668	Chr10: 2702399..2706473 (-)	1233	410	46.96	6.18	40.04	95.83	-0.11
		LOC104721325	Chr11: 3045672..3048571 (-)	1233	410	46.9	7.88	36.65	102	-0.11
		LOC104729750	Chr11: 3048944..3053111 (-)	1260	419	47.83	5.91	39.82	99.12	-0.10
		LOC104729749	Chr12: 2827481..2830547 (-)	1242	413	47.21	6.64	35.4	103.61	-0.10
		LOC109124813	Chr12: 2830650..2835267 (-)	1233	410	46.97	6.04	39.17	95.59	-0.10
Brassica rapa	4	Brara.H01190	ChrA08:13438743..13442092 (-)	1233	410	47.09	6.6	40.88	94.88	-0.10
		Brara.A00359	ChrA01:1829666..1833681 (-)	1230	409	46.97	6.57	42.52	95.11	-0.10
		Brara.K00336	Scaffold14751:213148..215580 (-)	1254	417	47.78	6.64	43.73	94.92	-0.10
		Brara.A00358	ChrA01:1825852..1829196 (-)	1257	418	47.59	6.09	34.9	104.04	-0.10
Brassica oleracea	3	Bol026669	ChrC03:23542260..23543581 (+)	495	168	18.68	7.0	37.25	100.43	-0.09
		Bol034675	ChrC01:33965574..33968500 (+)	1221	406	46.22	6.24	43.34	98.45	-0.09
		Bol013683	ChrC01:2208563..2215565 (-)	1269	422	48.02	5.39	42.13	108.36	-0.09
Brassica napus	9	LOC111208288	ChrC07: 53397062-53400258 (+)	1251	416	47.57	6.47	37.2	103.1	-0.09
		LOC106428793	ChrC07: 53393903-53396884 (+)	1254	417	47.72	6.47	43.17	93.29	-0.09
		LOC106361022	ChrA08: 13328923-13325352 (-)	1233	410	47.09	6.6	41.72	94.88	-0.08
		LOC106353554	ChrA01: 1933854..1937728 (-)	1230	409	46.97	6.57	42.52	95.11	-0.07
		LOC106353454	ChrA01: 1929825..1932692 (-)	1257	418	47.62	6.2	35.87	103.8	-0.07
		LOC111212694	ChrC01: 2235351..2239195 (-)	1230	409	47.02	6.1	43.29	94.4	-0.06
		LOC111212693	ChrC01: 2231578..2234676 (-)	1287	428	48.84	6.86	34.93	107.97	-0.06
		LOC111214348	ChrA03: 30618189-30621186 (-)	1254	417	47.78	6.47	44.19	94.92	-0.06
		LOC106371265	ChrC03: 67613161-67609853 (-)	1233	410	47.11	7.16	38.99	93.93	-0.04
Brassica nigra	6	BniB02g087140.2N.1	ChrB02:69594447..69596905 (+)	1215	404	46.26	8.18	41.31	96.04	-0.11
		BniB02g087140.2N.2	ChrB02:69596909..69599796 (+)	960	321	36.78	5.96	34.24	101.12	-0.08
		BniB03g020760.2N	ChrB03:9176668..9179936 (+)	1233	410	47.18	8.16	38.76	96.78	-0.03
		BniB05g003660.2N	ChrB05:2000701..2003303 (-)	1248	415	47.26	7.93	35.45	104.07	-0.03
		BniB05g003670.2N	ChrB05:2004126..2007325 (-)	1233	410	47.07	6.61	41.41	95.59	-0.03
		BniB07g031030.2N	ChrB07:41565390..41569964 (+)	1260	419	47.94	7.18	36.85	94.22	-0.02
Brassica carinata	6	BcaB01g03586	ChrB01:54956901..54972571 (+)	1257	418	48.06	6.65	30.16	99.76	-0.02
		BcaB03g11822	ChrB03:1729804..1733032 (-)	1233	410	47.18	8.16	38.47	96.54	-0.02
		BcaB04g19083	ChrB04:41795543..41799409 (+)	1059	352	39.93	7.52	34.75	91.42	-0.01
		BcaC01g00888	ChrC01:10158588..10162576 (-)	1233	410	47.06	6.89	39.57	93.93	0.01
		BcaNung00679	Contig324:658318..660890 (-)	1188	395	44.94	6.1	34.72	103.16	0.01
		BcaNung00680	Contig324:661893..665238 (-)	1230	409	47	6.1	43.68	95.11	0.02
Brassica juncea	2	BjSQS_A08	ChrA08:18527199..18533733 (-)	1233	410	47.09	6.6	41.72	94.88	0.04
		BjSQS_B07	ChrB07:49418848..49422584 (-)	1233	410	47.06	7.13	39.56	96.54	0.09

The prediction of sub-cellular localization indicated that all SQS proteins are localized in the endoplasmic reticulum

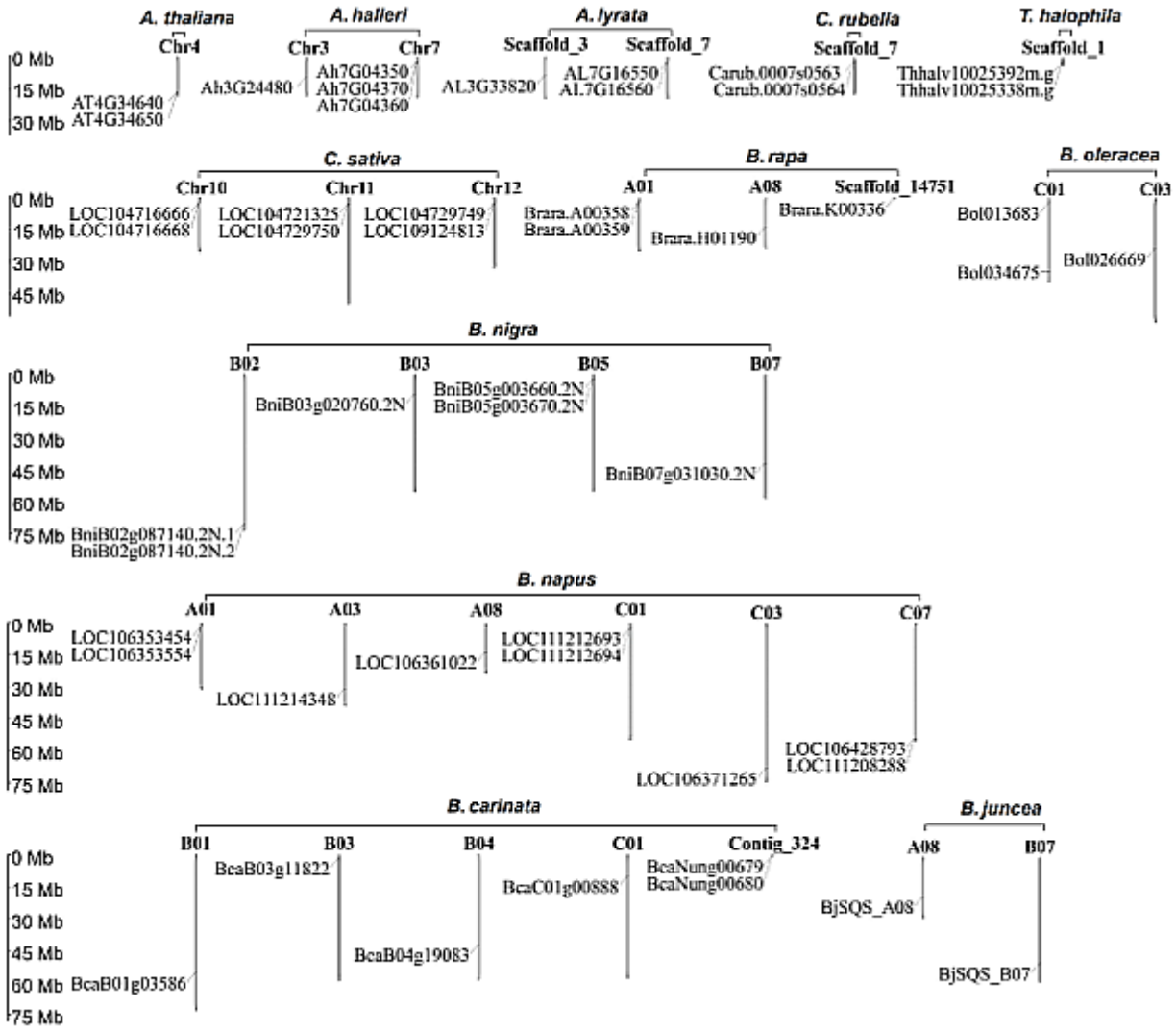

Figure 1. Genomic distribution of *SQS* genes across Brassicaceae. The genomic coordinates obtained from databases for each gene were used to map them on chromosomes/scaffolds using MG2C online tool (http://mg2c.iask.in/mg2c_v2.0/, Chao et al. 2021). The chromosome/scaffold number and length are provided on top and left of the figure, respectively

six chromosomes. Four of these chromosomes are similar to those of progenitor species; *B. rapa* (A01, A08) and *B. oleracea* (C01, C03). Of the six, three chromosomes contain a pair of tandem *SQS* genes each. In *B. carinata,* the six *SQS* genes were present on four chromosomes and a contig, two of the chromosomes were similar to progenitor species; *B. nigra* (B03) and *B. oleracea* (C01).

Evolutionary selection pressure analysis of *SQS* genes across Brassicaceae

To determine the selection pressure on *SQS* genes of Brassicaceae species, Ka/Ks values were calculated. In general, the Ka/Ks ratios lower than 1 indicates the presence of purifying selection; however, Ka/Ks ratio greater than 1 indicates positive selection, and Ka/Ks =1 indicates neutral selection (Hurst 2002). In our study, a total of 17 tandemly duplicated gene pairs in eleven Brassicaceae species were identified. The Ka/Ks ratios for each of the duplicated gene pairs were less than 0.5 and ranged from 0.17 to 0.38 with an average of 0.22, indicating that these gene pairs are under purification selection (Fig. 2a). Next, we determined the Ka/Ks values for orthologous gene pairs between *A. thaliana* (used as a reference genome) and other Brassicaceae species. This involved comparing gene pairs between *AtSQS1* with all other Brassicaceae orthologs, and applying a

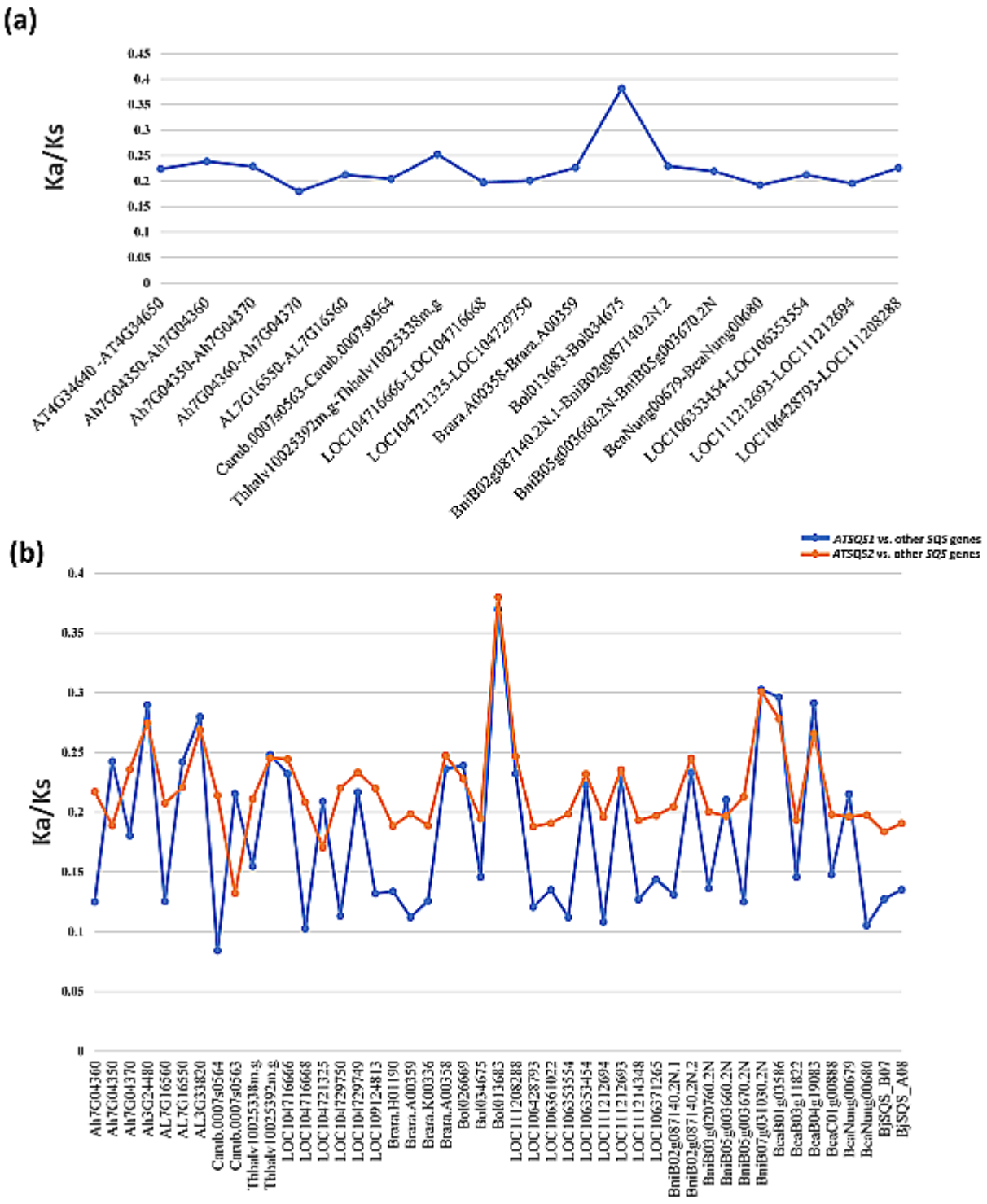

Figure 2. Comparative analysis of Ka/Ks values of *SQS* genes. **(a)** Ka/Ks ratio between the tandemly duplicated *SQS* gene pairs in each Brassicaceae species. The x- and y- axes represent the Ka/Ks values and gene pair names respectively. **(b)** Ka/Ks ratios between *SQS* genes from *Arabidopsis* and the rest of Brassicaceae species. The Ka/Ks ratios between each of *SQS* genes from Brassicaceae species and *A. thaliana SQS1* (AT4G34640; in blue) and *SQS2* (AT4G34650; in red) are shown. X-axis represents the Ka/Ks values; y-axis represents the *SQS* gene names from Brassicaceae species used for analysis with *Arabidopsis* genes

parallel approach to *AtSQS2*. The results indicated Ka/Ks values ranging from 0.08 to 0.37, with a mean value of 0.20, suggesting purifying selection (Fig. 2b). Additionally, all pairwise comparisons of genes within each of the three Brassicaceae SQS clades also exhibited Ka/Ks ratios of <1 (data not shown).

Cis-regulatory element analysis

The *cis*-regulatory elements were predicted in the sequences upstream (~1500 bp) to the translation start site of all the 49 *SQS* genes using the PlantCARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html, Lescot et al. 2002). In total,

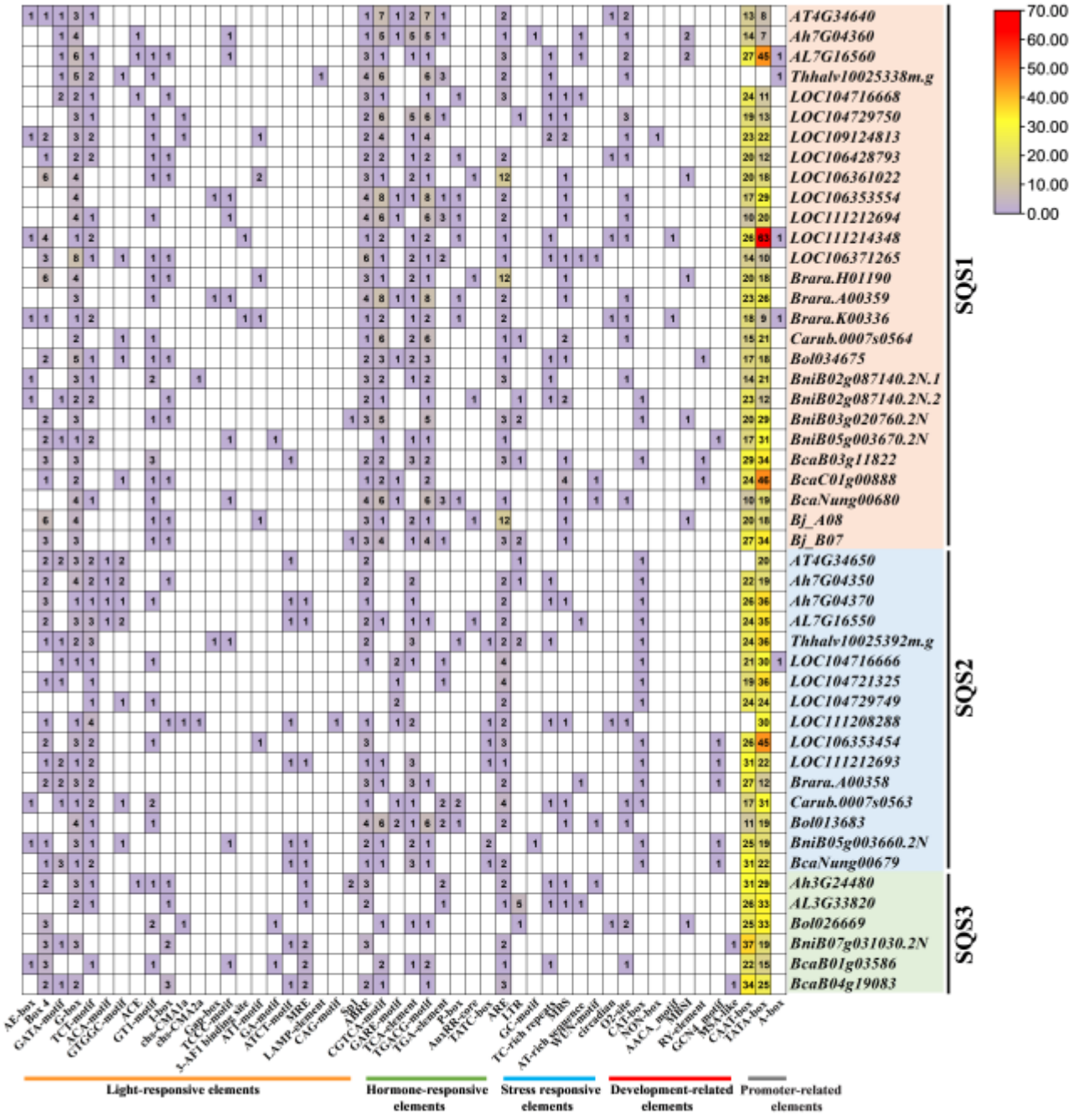

Figure 3. The *cis*-regulatory elements (CREs) predicted in *SQS* gene promoters. *SQS* gene names and the clade they belong to (unpublished results) have been shown on the right. The different types of CREs, categorized into five groups (indicated by colour bars) and their number (in boxes) for each *SQS* gene have been provided

50 types of CREs with known functions were identified across the *SQS* gene promoters analyzed. In addition, several elements of unknown function were also predicted. The fifty types of CREs could be categorized into five groups: light-responsive (22 CREs), hormone-responsive (9), stress-responsive (7), development-related elements (9) and promoter-related elements (3). Almost all the *SQS* genes contained at least one to many types of CREs belonging to each of the five categories (Fig. 3).

The promoter-related elements (TATA-box and CAAT-box) constituted the bulk of CREs identified. Maximum number (22) of CRE types were predicted in the light-responsive category. Among them, G-box was the most prevalent element predicted and was present in 91.8% of sequences followed by TCT-motif (71.4%), Box 4 (65.63%) and GT1-motif (57.1%). The remaining elements identified varied in their prevalence that ranged from 36.7% (I-box) to as low as 2.0% (LAMP-element and CAG-motif) of the sequences (Fig. 4a).

Of the nine types of hormone-responsive elements predicted, abscisic acid responsive (ABRE), methyl-jasmonate responsive (CGTCA-motif, TGACG-motif), salicylic acid responsive (TCA-element), were the most prevalent; present in 89.8, 73.5, 71.4, and 69.4% of the total sequences, respectively. The remaining five elements were predicted in 30.6 to 10.2% of the sequences. These included- auxin-responsive (TGA-element; 30.6%), gibberellin-

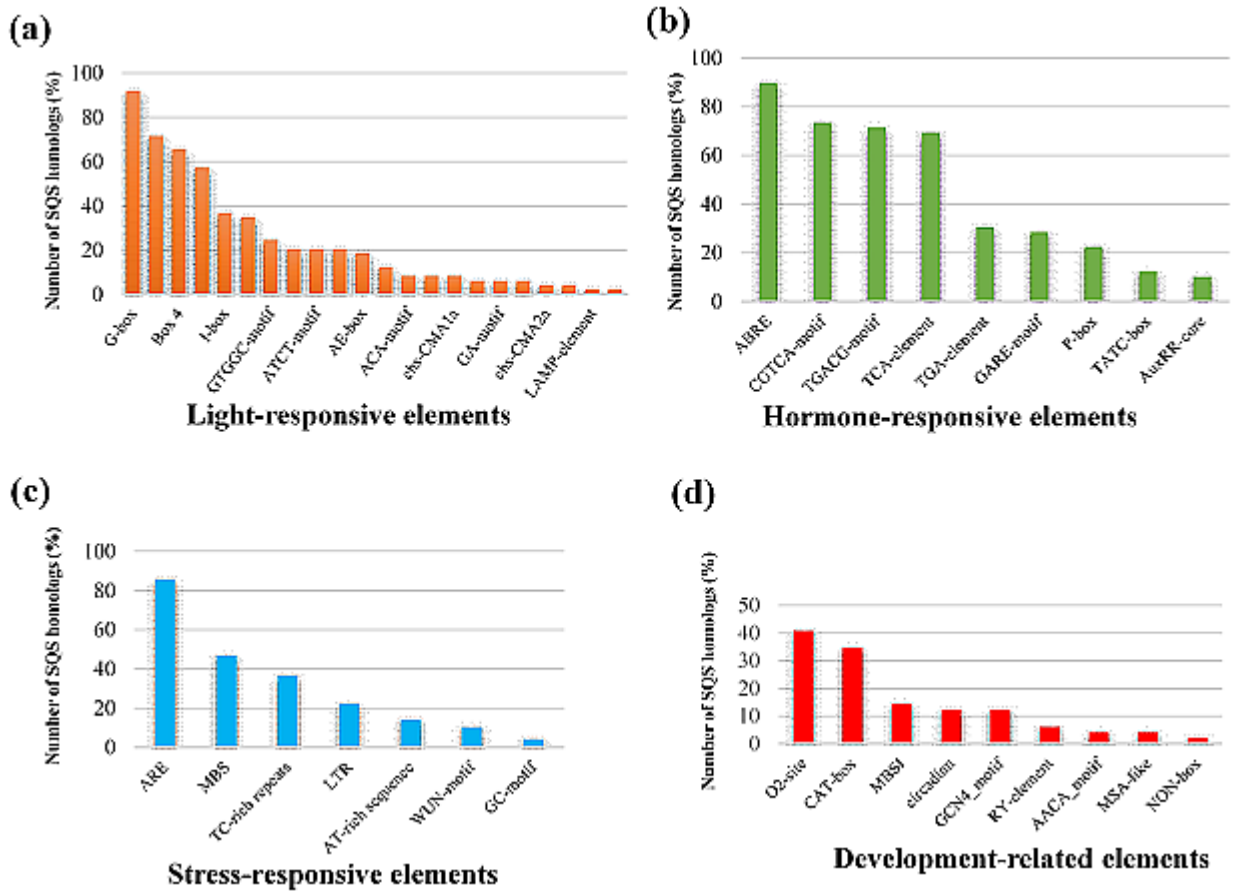

Figure 4. Distribution of *cis*-regulatory elements (CREs) in the sequences upstream to *SQS* genes. The type and distribution of CREs belonging to the four categories: **(a)** light-responsive elements, **(b)** hormone-responsive elements, **(c)** stress-responsive elements and **(d)** development-related elements, identified in the 49 Brassicaceae *SQS* genes has been shown. X-axis: the total number (%) of *SQS* genes having a particular CRE; Y-axis: type of CRE.

responsive (GARE-motif; 28.6%), P-box (22.4%), TATC-box (12.2%), and AuxRR-core (10.2%) (Fig. 4b).

In the category of stress responsive elements, anaerobic induction responsive (ARE) element was the most predominant and present in 85.7% of the total sequences analyzed and the remaining six elements such as drought responsive (MBS), defense and stress responsive (TC-rich repeats), low-temperature responsive (LTR), elicitor responsive (AT-rich sequence), wound-responsive (WUN-motif), anoxic specific (GC-motif) were less predominant and were present in 46.9, 36.7, 22.4, 14.3, 10.2 and 4.1% of the sequences, respectively (Fig. 4c).

Unlike the other three categories of CREs described above, none of the nine plant development-related elements were found to be predominantly present across the sequences. The highest frequency of occurrence was 40.8 and 34.7% for the zein metabolism regulation element (O2-site) and meristem expression element (CAT-box) respectively. The remaining seven elements; flavonoid biosynthetic genes regulation (MBSI), circadian control element (circadian), endosperm expression (GCN4_motif), seed-specific regulation (RY-element), endosperm-specific element (AACA_motif), cell cycle regulation (MSA-like), and meristem specific activation (NON-box), were identified in very few (14.3 to 2.0%) sequences (Fig. 4d).

We also compared the CRE landscape across the three SQS clades to get an insight into the divergence of these elements (in terms of the both type and frequency) during the evolution of the gene within Brassicaceae. There were some obvious patterns observed despite the fact that the total number of sequences of SQS1 and SQS2 analyzed for this comparison were not equal (Fig. 3). Two hormone-responsive elements (CGTCA- and TGACG-motifs), were present in all SQS1 homologs analyzed but were less prevalent (30-37%) in SQS2 homologs. Similarly, the O2-site (plant development element), was prevalent (55.5%) in SQS1, but rarely represented in the SQS2 sequences. Another plant development element, the CAT-box, however was predominant (14 of the 16 sequences) in the SQS2 homologs, and present very rarely (3 of the 27) in the SQS1 clade sequences. CAT-box has a role in meristem-specific activation. TCT-box was present in 100% of SQS2, while only in 59.2% of SQS1 sequences.

As revealed by the *in-silico* promoter analysis, *SQS* genes are potentially regulated through multiple *cis*-regulatory elements. Of these, CREs belonging to the light- and hormone- responsive categories were most prevalent across members of all the three clades signifying their role in regulating the *SQS* genes. However, none of the plant development-related CREs were present predominantly. Differences in the *cis*-acting element sequences between the three clades is suggestive of divergence that might be involved in the spatial/temporal expression differences of the paralogs. Variable expression patterns of *SQS* gene copies have been previously reported in Arabidopsis, *Glycine max* and *P. ginseng* (Busquets et al. 2008, Kim et al. 2011a, Nguyen et al. 2013). *Cis*-regulatory elements including light, hormone (abscisic acid, MeJA-responsiveness, gibberellin, auxin), and stress (low-temperature, heat and drought) have been reported previously, although only in *SQS* genes from few plant species including *S. baumii*, *W. somnifera* and *P. ginseng* (Kim et al. 2011a, Bhat et al. 2012, Wang et al. 2022). This, to the best of our knowledge, is the first extensive *in-silico* prediction and comparison of CREs in *SQS* promoters across Brassicaceae. The results obtained in this study will facilitate future studies on elucidating the gene expression patterns of *SQS* paralogs in Brassicaceae.

CONCLUSIONS

In this study, we identified 49 *SQS* genes from Brassicaceae species and compared their genomic location, physicochemical properties, sequence evolutionary rates and *cis*-regulatory elements. The copy number of *SQS* genes was found to range from 2 - 9. Most SQS proteins were predicted to be stable, hydrophilic proteins, rich in aliphatic amino acids and localized within the endoplasmic reticulum and under purifying selection pressure. In the SQS upstream regulatory regions, CREs responsible for light, hormone and stress-responsiveness were most prevalent. In each of these categories, we identified certain elements that were uniformly predicted in most *SQS* genes indicating their significance in regulating the genes. Additionally, certain paralog

specific CREs were also identified suggestive of regulatory diversification of the gene copies. These findings offer insights into the molecular evolution of *SQS* genes within the Brassicaceae family.

ACKNOWLEDGEMENTS

The study was supported by Research and Development (R&D) grants from the University of Delhi to PP. RB is grateful to University Grants Commission (UGC) for providing fellowship (JRF/SRF). KC is thankful to the Principal, Bhaskaracharya College of Applied Sciences, University of Delhi for providing encouragement to pursue this study.

Author's contributions: RB, KC and PP conceptualized the project and contributed to the methodology. RB wrote the original draft, conducted the formal analyses and put together the figures. PP contributed to the analysis, writing and critical review of the manuscript. All authors read and approved the final manuscript.

Conflict of interest: The authors have no competing interests to declare that are relevant to the content of this article.

REFERENCES

Arsovski, A.A., Pradinuk, J., Guo, X.Q., Wang, S. and Adams, K.L. 2015. Evolution of *cis*-regulatory elements and regulatory networks in duplicated genes of Arabidopsis. Plant Physiology, 169(4), 2982-2991. https://doi.org/10.1104/pp.15.00717

Bhat, W.W., Lattoo, S.K., Razdan, S., Dhar, N., Rana, S., Dhar, R.S., Khan, S. and Vishwakarma, R.A. 2012. Molecular cloning, bacterial expression and promoter analysis of squalene synthase from *Withania somnifera* (L.) Dunal. Gene, 499(1), 25-36. https://doi.org/10.1016/j.gene.2012.03.004

Busquets, A., Keim, V., Closa, M., Del Arco, A., Boronat, A., Arró, M. and Ferrer, A. 2008. *Arabidopsis thaliana* contains a single gene encoding squalene synthase. Plant Molecular Biology, 67, 25-36. https://doi. org/10.1007/s11103-008-9299-3

Chalbi, N., Martínez-Ballesta, M.C., Youssef, N.B. and Carvajal, M. 2015. Intrinsic stability of Brassicaceae plasma membrane in relation to changes in proteins and lipids as a response to salinity. Journal of Plant Physiology, 175, 148-156. http:// doi.org/10.1016/j.jplph.2014.12.003

Chao, J., Li, Z., Sun, Y., Aluko, O.O., Wu, X., Wang, Q. and Liu, G. 2021. MG2C: A user-friendly online tool for drawing genetic maps. Molecular Horticulture, 1, 1-4. https://doi.org/10.1186/s43897-021-00020-x

Chen, C., Chen, H., Zhang, Y., Thomas, H.R., Frank, M.H., He, Y. and Xia, R. 2020. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Molecular Plant, 13(8), 1194-1202. http://doi.org/10.1016/j.molp.2020.06.009

Chen, H., Wang, T., He, X., Cai, X., Lin, R., Liang, J., Wu, J., King, G. and Wang, X. 2022. BRAD V3. 0: an upgraded Brassicaceae database. Nucleic Acids Research, 50(D1), D1432-D1441. https://doi.org/10.1093/nar/gkab1057

Devarenne, T.P., Shin, D.H., Back, K., Yin, S. and Chappell, J. 1998. Molecular characterization of tobacco squalene synthase and regulation in response to fungal elicitor. Archives of Biochemistry and Biophysics, 349(2), 205-215. https://doi.org/10.1002/pld3.63

Essadek, S., Gondcaille, C., Savary, S., Samadi, M., Vamecq, J., Lizard, G., El Kebbaj, R., Latruffe, N., Benani, A., Nasser, B. and Cherkaoui-Malki, M. 2023. Two Argan Oil Phytosterols, Schottenol and Spinasterol, Attenuate Oxidative Stress and Restore LPS-Dysregulated Peroxisomal Functions in Acox1"/" and Wild-Type BV-2 Microglial Cells. Antioxidants, 12(1), 168. https://doi.org/10.3390/antiox12010168

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N. and Rokhsar, D.S. 2012. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research, 40(D1), D1178-D1186. https://doi.org/10.1093/nar/gkr944

Hazra, A., Dutta, M., Dutta, R., Bhattacharya, E., Bose, R. and Biswas, S.M. 2023. Squalene synthase in plants–Functional intricacy and evolutionary divergence while retaining a core catalytic structure. Plant Gene, 33, 100403. https://doi.org/10.1016/j.plgene.2023.100403

Hurst, L.D. 2002. The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends in Genetics, 18(9), 486-487. https://doi.org/10.1016/s0168-9525(02)02722-1

Jennings, S.M., Tsay, Y.H., Fisch, T.M. and Robinson, G.W. 1991. Molecular cloning and characterization of the yeast gene for squalene synthetase. Proceedings of the National Academy of Sciences, 88(14), 6038-6042. https://doi.org/10.1073/pnas.88.14.6038

Jiang, Y., Chen, H., Chen, X., Köllner, T.G., Jia, Q., Wymore, T.W., Wang, F. and Chen, F. 2015. Volatile squalene from a nonseed plant *Selaginella moellendorffii*: emission and biosynthesis. Plant Physiology and Biochemistry, 96, 1-8. https://doi.org/10.1016/j.plaphy.2015.07.010

Kagale, S., Robinson, S.J., Nixon, J., Xiao, R., Huebert, T., Condie, J., Kessler, D., Clarke, W.E., Edger, P.P., Links, M.G. and Sharpe, A.G. 2014. Polyploid evolution of the Brassicaceae during the Cenozoic era. The Plant Cell, 26(7), 2777-2791. https://doi.org/10.1105/tpc.114.126391

Kim, T.D., Han, J.Y., Huh, G.H. and Choi, Y.E. 2011a. Expression and functional characterization of three squalene synthase genes associated with saponin biosynthesis in *Panax ginseng*. Plant and Cell Physiology, 52(1), 125-137. https://doi.org/10.1093/pcp/pcq179

Kim, Y.S., Cho, J.H., Park, S., Han, J.Y., Back, K. and Choi, Y.E. 2011b. Gene regulation patterns in triterpene biosynthetic pathway driven by overexpression of squalene synthase and methyl jasmonate elicitation in *Bupleurum falcatum*. Planta, 233, 343-355. https://doi.org//10.1007/s00425-010-1292-9

Lee, S. and Poulter, C.D. 2008. Cloning, solubilization, and characterization of squalene synthase from *Thermosynechococcus elongatus* BP-1. Journal of Bacteriology, 190(11), 3808-3816. https://doi.org/10.1128/JB.01939-07

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouzé, P. and Rombauts, S. 2002. PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Rresearch, 30(1), 325-327. https://doi.org/10.1093/nar/30.1.325

Liu, D., Pi, J., Zhang, B., Zeng, H., Li, C., Xiao, Z., Fang, F., Liu, M., Deng, N. and Wang, J. 2023. Phytosterol of lotus seed core powder alleviates hypercholesterolemia by regulating gut microbiota in high-cholesterol diet-induced C57BL/6J mice. Food Bioscience, 51, 102279. https://doi.org/10.1016/j.fbio.2022.102279

Lopez-Garcia, G., Cilla, A., Barberá, R. and Alegria, A. 2017. Antiproliferative effect of plant sterols at colonic concentrations on Caco-2 cells. Journal of Functional Foods, 39, 84-90. https://doi.org/10.1016/j.jff.2017.10.006

Loìpez-Garciìa, G., Cilla, A., Barberaì, R. and Alegría, A. 2020. Anti-inflammatory and cytoprotective effect of plant sterol and galactooligosaccharides-enriched beverages in caco-2 cells. Journal of Agricultural and Food Chemistry, 68(7), 1862-1870. https://doi.org/10.1021/acs.jafc.9b03025

McKenzie, T.L., Jiang, G.U.O.J.I.A.N., Straubhaar, J.R., Conrad, D.G. and Shechter, I. 1992. Molecular cloning, expression, and characterization of the cDNA for the rat hepatic squalene synthase. Journal of Biological Chemistry, 267(30), 21368-21374. https://doi.org/10.1016/S0021-9258(19)36619-0

Nasrollahi, V., Mirzaie-Asl, A., Piri, K., Nazeri, S. and Mehrabi, R. 2014. The effect of drought stress on the expression of key genes involved in the biosynthesis of triterpenoid saponins in liquorice (*Glycyrrhiza glabra*). Phytochemistry, 103, 32-37. https://doi.org/10.1016/j.phytochem.2014.03.004

Nguyen, H.T., Neelakadan, A.K., Quach, T.N., Valliyodan, B., Kumar, R., Zhang, Z. and Nguyen, H.T. 2013. Molecular characterization of Glycine max squalene synthase genes in seed phytosterol biosynthesis. Plant Physiology and Biochemistry, 73, 23-32. https://doi.org/10.1016/j.plaphy.2013.07.018

Okada, S., Devarenne, T.P. and Chappell, J. 2000. Molecular characterization of squalene synthase from the green microalga *Botryococcus braunii*, race B. Archives of Biochemistry and Biophysics, 373(2), 307-317. https://doi.org/10.1006/abbi.1999.1568

Poulose, N., Sajayan, A., Ravindran, A., Chandran, A., Priyadharshini, G.B., Selvin, J. and Kiran, G.S. 2021. Anti-diabetic potential of a stigmasterol from the seaweed *Gelidium spinosum* and its application in the formulation of nanoemulsion conjugate for the development of functional biscuits. Frontiers in Nutrition, 8, 694362. https://doi.org/10.3389/fnut.2021.694362

Poulter, C.D. 1990. Biosynthesis of non-head-to-tail terpenes. Formation of 1'-1 and 1'-3 linkages. Accounts of Chemical Research, 23(3), 70-77. https://doi.org/10.1021/ar00171a003

Rhee, S.Y., Beavis, W., Berardini, T.Z., Chen, G., Dixon, D., Doyle, A., Garcia-Hernandez, M., Huala, E., Lander, G., Montoya, M. and Miller, N. 2003. The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. Nucleic Acids Research, 31(1), 224-228. https://doi.org/10.1093/nar/gkg076

Rilling, H.C. and Epstein, W.W. 1969. Mechanism of squalene biosynthesis. Presqualene, a pyrophosphorylated precursor to squalene. Journal of the American Chemical Society, 91(4), 1041-1042. https://doi.org/10.1021/ja01032a052

Robinson, G.W., Tsay, Y.H., Kienzle, B.K., Smith-Monroy, C.A. and Bishop, R.W. 1993. Conservation between human and fungal squalene synthetases: similarities in structure, function, and regulation. Molecular and Cellular Biology, 13(5), 2706-2717. https://doi.org/10.1128/mcb.13.5.2706

Rong, Q., Jiang, D., Chen, Y., Shen, Y., Yuan, Q., Lin, H., Zha, L., Zhang, Y. and Huang, L. 2016. Molecular cloning and functional analysis of squalene synthase 2 (SQS2) in *Salvia miltiorrhiza* Bunge. Frontiers in Plant Science, 7, 1274. https://doi.org/10.3389/fpls.2016.01274

Rossi, S. and Huang, B. 2022. Sitosterol-mediated antioxidant regulation to enhance heat tolerance in creeping bentgrass. Journal of the American Society for Horticultural Science, 147(1), 18-24. https://doi.org/10.21273/JASHS05107-21

Sheikhi, S., Ebrahimi, A., Heidari, P., Amerian, M.R., Rashidi-Monfared, S. and Alipour, H. 2023. Exogenous 24-epibrassinolide ameliorates tolerance to high-temperature by adjusting the biosynthesis of pigments, enzymatic, non-enzymatic antioxidants, and diosgenin content in fenugreek. Scientific Reports, 13(1), 6661. https://doi.org/10.1038/s41598-023-33913-6

Shirazi, Z., Aalami, A., Tohidfar, M. and Sohani, M.M. 2019. Triterpenoid gene expression and phytochemical content in Iranian licorice under salinity stress. Protoplasma, 256, 827-837. https://doi.org/10.1007/s00709-018-01340-4

Singh, A.K., Dwivedi, V., Rai, A., Pal, S., Reddy, S.G.E., Rao, D.K.V., Shasany, A.K. and Nagegowda, D.A. 2015. Virusinduced gene silencing of *Withania somnifera* squalene synthase negatively regulates sterol and defencerelated genes resulting in reduced withanolides and biotic stress tolerance. Plant Biotechnology Journal, 13(9), 1287-1299. https://doi.org /10.1111/pbi.12347

Summers, C., Karst, F. and Charles, A.D. 1993. Cloning, expression and characterisation of the cDNA encoding human hepatic squalene synthase, and its relationship to phytoene synthase. Gene, 136(1-2), 185-192. https://doi.org/10.1016/0378-1119(93)90462-c.

Wang, H., Nagegowda, D.A., Rawat, R., BouvierNavé, P., Guo,

D., Bach, T.J. and Chye, M.L. 2012a. Overexpression of *Brassica juncea* wildtype and mutant HMGCoA synthase 1 in Arabidopsis upregulates genes in sterol biosynthesis and enhances sterol production and stress tolerance. Plant Biotechnology Journal, 10(1), 31-42. http://doi.org/10.1111/j.1467-7652.2011.00631.x

Wang, K., Senthil-Kumar, M., Ryu, C.M., Kang, L. and Mysore, K.S. 2012b. Phytosterols play a key role in plant innate immunity against bacterial pathogens by regulating nutrient efflux into the apoplast. Plant Physiology, 158(4), 789-1802. http://doi.org/10.1104/pp.111.189217

Wang, S., Wang, X., Liu, Z., Sun, T. and Zou, L. 2022. Expression analysis of squalene synthase gene in mevalonate pathway of *Sanghuangporus baumii*. Biotechnology & Biotechnological Equipment, 36(1), 176-183. https://doi.org/10.1080/13102818.2022.2060759

Wei, S., Yang, X., Huo, G., Ge, G., Liu, H., Luo, L., Hu, J., Huang, D. and Long, P. 2020. Distinct metabolome changes during seed germination of lettuce (*Lactuca sativa* L.) in response to thermal stress as revealed by untargeted metabolomics analysis. International Journal of Molecular Sciences, 21(4), 1481. https://doi.org/10.3390/ijms21041481

Zhang, B., Liu, Y., Chen, M., Feng, J., Ma, Z., Zhang, X. and Zhu, C. 2018. Cloning, expression analysis and functional characterization of squalene synthase (SQS) from *Tripterygium wilfordii*. Molecules, 23(2), 269. https://doi.org/10.3390/molecules23020269

Received: 8th May 2024
Accepted: 6th June 2024

Supplementary Table 1. List of 49 *SQS* genes identified from twelve Brassicaceae species were provided, along with their accession no. and the databases from which they were retrieved

S.No.	Species	Accession No.	Database
1	*Arabidopsis thaliana*	AT4G34640	PHYTOZOME
2		AT4G34650	
3	*Arabidopsis halleri*	Ah7G04360	PHYTOZOME
4		Ah7G04350	
5		Ah7G04370	
6		Ah3G24480	
7	*Arabidopsis lyrata*	AL7G16560	PHYTOZOME
8		AL7G16550	
9		AL3G33820	
10	*Capsella rubella*	Carub.0007s0564	PHYTOZOME
11		Carub.0007s0563	
12	*Thellungiella halophila*	Thhalv10025338m.g	BRAD
13		Thhalv10025392m.g	
14	*Camelina sativa*	LOC104716666	NCBI
15		LOC104716668	
16		LOC104721325	
17		LOC104729750	
18		LOC104729749	
19		LOC109124813	
20	*Brassica rapa*	Brara.H01190	PHYTOZOME
21		Brara.A00359	
22		Brara.K00336	
23		Brara.A00358	
24	*Brassica oleracea*	Bol026669	PHYTOZOME
25		Bol034675	
26		Bol013683	
27	*Brassica napus*	LOC111208288	NCBI
28		LOC106428793	
29		LOC106361022	
30		LOC106353554	
31		LOC106353454	
32		LOC111212694	
33		LOC111212693	
34		LOC111214348	
35		LOC106371265	
36	*Brassica nigra*	BniB02g087140.2N.1	BRAD
37		BniB02g087140.2N.2	
38		BniB03g020760.2N	
39		BniB05g003660.2N	
40		BniB05g003670.2N	
41		BniB07g031030.2N	
42	*Brassica carinata*	BcaB01g03586	BRAD
43		BcaB03g11822	
44		BcaB04g19083	
45		BcaC01g00888	
46		BcaNung00679	
47		BcaNung00680	
48	*Brassica juncea*	BjSQS_B07	NCBI
49		BjSQS_A08	