Genomic Analysis of Sugar Transporter Gene Family in Cultivated Peanut (Arachis Hypogaea): Evolution, Expression Profiles and Gene Variation

zhijun Xu
Zhanjiang experiment station/South subtropical crop reaearch institute https://orcid.org/0000-0002-9829-8444

Sheng Zhao
Shenzhen agriculture genome research institute

Xiaowen Hu
Zhanjiang experiment station/South subtropical crop research institute

Xiaoli Wu
Zhanjiang academy of agriculture science

Yang Liu (lyfull@163.com)
Zhanjiang Experiment Station / South Subtropical Crop Research Institute

Research article

Keywords: cultivated peanut, STP gene family, evolution, expression patterns, gene variation

Posted Date: July 16th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-41476/v1

Abstract

Background

Sugar transporter (STP) gene family, belonging to the major facilitator superfamily, plays significant roles in monosaccharide distribution and many aspects of physiological processes. However, little information was available about the STP genes in cultivated peanut (*Arachis hypogaea*), an important edible and oil crop. The recent release of the whole-genome sequence of cultivated peanut allowed us to perform a genome-wide investigation into the phylogeny and expression profiling of peanut STP genes.

Results

A total of thirty-six STP genes containing the Sugar_tr conserved motifs were identified from the *A. hypogaea* genome and all the genes were renamed on the basis of their respective chromosome distribution. According to their phylogenetic features, the STPs were classified into four groups. The structure of the STP genes and their encoded proteins were examined. Synteny analysis and phylogenetic comparison of the STP genes provided deep insight into the evolutionary characteristics of peanut STP genes. The segmental duplication events played a major role in the expansion of the peanut STP gene family and homeologous chromosomes rearrangement may lead to the exchange of STP genes between the A and B sub-genome. Expression profiles derived from transcriptome data exhibited distinct expression patterns of AhSTP genes in various tissues. Among them, four AhSTP genes, *AhSTP3*, *AhSTP9*, *AhSTP19*, *AhSTP28*, exhibited high expression in early stage of pod formulate in subsp. *hypogaea*, and in developing phase of seed in subsp. *fastigiate*, suggesting these genes may be involved in the development of pod and seed. Gene variation analysis of the four genes utilizing the subspecies genome sequences indicated multiple variations occurred in gene sequence or promoter region, and provided valuable clues for the different expression profiles of the four genes during seed development in subsp. *hypogaea* and subsp. *fastigiate*.

Conclusion

Thirty-six STP genes were identified in cultivated peanut and their protein character,structure,evolution characteristics, expression patterns and gene variation were analyzed.This study provided a foundation for further functional characterization of STP genes with an aim of cultivated peanut crop improvement.

Background

Sugar transporter proteins are regarded as H^+/sugar symporters and members of the major facilitator superfamily, commonly with 12 transmembrane domains [1]. STP proteins are mostly characterized by plasma-membrane-location and exhibit acquisition of broad monosaccharide substrates, such as glucose, fructose, galactose, pentose xylose (except ribose), mannose and non-metabolized 3-0-methylglucose (3-OMG) [2]. Since the first STP gene was reported in Chlorella [3], over one hundred of its homologs had been identified and characterized in a wide range of plant species, such as rabidopsis

(*Arabidopsis thaliana*, 14 members) [2], rice (*Oryza sativa*, 29 members) [4, 5], tomato (*Solanum lycopersicum*, 18 members) [6], pear (*Pyrus bretschneideri*, 20 members) [7, 8], cassava (Manihot esculenta, 20 members) [8], woodland strawberry (Fragaria vesca, 24 members) [9], Cabbage (Brassica oleracea, 22 members) [10], etc.

Substantial evidence indicates that the STPs are involved in the uptake of hexose hydrolyzed from sucrose in the apoplast and sugar unloading which contribute to carbon allocation, yield, and environmental adaptation in crops [1]. In Arabidopsis, for example, after wounding, *AtSTP3* and *AtSTP4* transcript levels transiently rise [11, 12]; pathogen attack can induce the expression of *AtSTP4* [12]; *AtSTP1* and *AtSTP14* show circadian oscillations and are strongly dark-inducible [13, 14].

Previous studies have shown that STP genes have different expression patterns during plant growth and respond to different stress conditions. In Arabidopsis, most of the AtSTPs show sink-specific expression [2]. For example, *AtSTP1* is expressed in germinating seeds, roots and guard cells [13, 15, 16]. *AtSTP2*, *AtSTP4*, *AtSTP6*, *AtSTP9* and *AtSTP11*, are expressed in pollen during different developmental stages [17–20]. *AtSTP3* is expressed in the source tissues (green leaves) under normal condition [12]. *AtSTP4* show expression in root tips [11], *AtSTP5* in silique and whole seedling [21], *AtSTP13* in the vascular tissue of petals [22]. *AtSTP14* is expressed most prominently in leaves and siliques [23]. As the newly report, *AtSTP7* and *AtSTP12* can be detected at whole open flowers, leaves, stems, roots, whole seedlings, and siliques; while *AtSTP8* is expressed exclusively in floral tissues [21]. Although previous studies have been explored in the expression patterns and functional analysis of STP genes, the extensive expression profiles of the STP genes in subterranean pod crops, such as peanut, remain poorly understood.

Cultivated peanut is an important edible and oil crop planted in the tropics and semi-arid tropics of the world, providing high-quality vegetable oil and protein. Genetic, cytogenetic, phylogeographic and molecular evidences have suggested that cultivated peanut is an allotetraploid (AABB, 2n = 4x = 40) formed through hybridization between *A. duranensis* (AA, 2n = 2x = 20) and *A. ipaensis* (BB, 2n = 2x = 20) [24–29]. It's specific in flower stalks (peg) elongation and developing pods under the ground, which contains sugar transportation, accumulation and transition processes [30]. But until now, little information on STP gene family expression in peg elongation, pod development and other tissues of peanut is available for us. In this study, we took advantage of the peanut genomes and transcriptome data to identify the complete set of STP gene family in peanut. Their sequence characteristics and gene structures were detailly analyzed. Segmental and tandem duplication of peanut STPs were investigated by synteny analysis. The expression profiles of AhSTP genes in different tissues, including peg and pod, were analyzed using the publicly available transcriptome data, in an attempt to understand their possible roles in peg and pod development. In addition, the cause for specific AhSTP genes was explored through gene variation analysis, possessing different expression profiles during seed development in peanut subspecies.

Results

Identification and protein character analysis of STPs in A. hypogaea, A. duranensis **and** A. ipaënsis

A total of 36 STP genes was identified in peanut genome, while, 15 and 16 STP genes were identified in the haploid ancestors of cultivated peanut, *A. duranensis* and *A. ipaënsis*, respectively (Table S1 and Table S2). These STP genes were named *AhSTP1* to *AhSTP36* in *A. hypogaea*, *AdSTP1* to *AdSTP15* in *A. duranensis*, and *AiSTP1* to *AiSTP16* in *A. ipaënsis*, respectively, according to their order on the chromosomes. The STP genes were unevenly located on the three genomes, and no STP gene was found to distribute on chromosome A07 of *A. duranensis*, B08 of *A. ipaënsis* or *A. hypogaea* (Table S1).

Protein characteristics, including the length of the protein sequence, the molecular weight (MW), theoretical isoelectric point (pI), aliphatic index (AI), grand average of hydropathicity (GRAVY), protein transmembrane domains (TMD) and STP domain location were analyzed (Fig. 1 and Table S1). The protein sequence length ranged from 397 (AhSTP21) to 1208 amino acids (aa) (AdSTP1), with the predicted MW from 44.01 (AhSTP21) to 135.36 KD (AdSTP1). The pI, AI and GRAVY ranged from 5.92 to 9.72, 93.7 to 111.67, and 0.14 to 0.67. In addition, 60 0f the 67 STPs contained 10 ~ 12 conserved TMDs, 6 STPs harbored 9 conserved TMDs and AhSTP26 carried only 8 TMDs. While 3 STPS (AdSTP1, AiSTP1 and AiSTP16) contained excessive TMD, and AiSTP16 contained double sugar_tr domain. The conserved STP domain of STPs fluctuated from 376 to 476 aa. Subcellular localization prediction indicated that all peanut STP proteins were localized in the cell membrane.

Phylogenetic analysis of peanut STPs

A phylogenetic tree was constructed by neighbor-joining method, using 14 AtSTP, 15 AdSTP, 16 AiSTP, 36 AhSTP and 30 GmSTP full-length protein sequences (Fig. 2). Phylogenetic analysis revealed that the STP proteins could be classified into four groups with AtSTP, AdSTP, AiSTP, AhSTP and GmSTP in each group. In group I, there were 32 STP, including 2 AtSTP, 5 AdSTP, 6 AiSTP, 12 AhSTP and 7 GmSTP. Group II, III, IV contained 26, 31 and 22 members with 10, 8 and 6 AhSTPs in each group, respectively. Interestingly, all AhSTPs were composed homologous pairs with AdSTP or AiSTP in the phylogenetic tree.

Gene structure and motif composition of peanut STPs

The exon-intron organizations of all the identified STP genes were analyzed to get more insight into the evolution of the STP gene family in peanut. As shown in Fig. 3B, in *A. hypogaea*, the exon of AhSTPs fluctuated from 1 to 11, of which 20 AhSTP genes contained 3 exons and 7 with 4 exons. In *A. duranensis*, all the STP genes had only 1 exon to 8 exons. In *A.ipaënsis*, all the STP genes possessed 1 exon to 6 exons. The majority of the AhSTP had a similar structure with its ancestor STP (AdSTP or AiSTP) at the bottom branch of phylogenetic tree (Fig. 3). Furthermore, in *A. hypogaea*, only 3 AhSTP genes had no untranslated sequence (UTR), while, *A. duranensis and A. ipaënsis* both had 6 STP genes with no UTR.

Conserved motif analysis of all the peanut STP proteins was performed using the MEME program. The results showed that 12 motifs were identified with their length ranging from 21 to 49 aa, 9 motifs were

distributed on all STP proteins except motif 3, 4 and 9 (Fig. 3C). Most STP proteins had similar motif architecture at the bottom branch of phylogenetic tree. For example, AhSTP20 and AhSTP24, with 12 motifs, had the same motifs distribution in the same branch of group I, as well as AdSTP6, AdSTP7, AhSTP25 and AiSTP7 in the same branch of group III. Comparing to the other STPs, the motifs distribution on AiSTP16 was unique, since that motif 1/7/9 distributed 3 times, motif 2/3/4/5/8/10/11 distributed 2 times, and motif 6/12 distributed once (Fig. 3B). The unique motifs distribution on AiSTP16 might indicate its unique evolution and function. To further understand the character of the motifs, the motif sequence was predicted by Pfam program, and motifs 1, 2, 3, 5, 6, 7, 8 were part of the putative sugar_tr domain (Table 1). Overall, the similar gene structures and conserved motif compositions of the STPs in the same branch, together with the phylogenetic analysis results, could strongly support the reliability of the group classifications.

Table 1
Twelve different motifs commonly observed in peanut STP proteins

Motif ID	Length	Protein squences	Pfam Domain
1	41	ICJYVAGFAWSWGPLGWLVPSEIFPLEIRSAGQSINVSVNM	Sugar_tr
2	29	VGFANQSVPJYLSEMAPYKYRGALNIGFQ	Sugar_tr
3	47	RQYEGKVTLFVIITCIVAAMGGLJFGYDJGISGGVTSMDPFLKKFFP	Sugar_tr
4	46	NQYCKFDSQVLTLFTSSLYLAALVASLFASTVTRAFGRRLSMILGG	
5	40	AAVPAILJCIGAJFLPDTPNSLIERGQHEKAKKMLQKIRG	Sugar_tr
6	41	LSMLCHFKFGLFIFFAAWVLVMTIFIYFLLPETKGVPIEEM	Sugar_tr
7	41	TLVSIFTVDKFGRRKLFLEGGAQMFICQIVIGAAIGSKFGD	Sugar_tr
8	49	DEEFQDLVDASEAASKVKHPWKNJLKRRYRPQLVMAIFIPFFQQLTGIN	Sugar_tr
9	29	FYAPVLFTTJGFGSBASLMSAVITGGVNV	
10	29	JFFLVGALLNGFAQNVAMLIIGRILLGFG	
11	21	ITIGILVANLJNYFTAKIKNG	
12	21	NYVWKSHWFWNKFVPSDSVLV	

Synteny analysis of peanut STP genes

According to the study of Yu [31], segmental duplications were characterized as multiple genes of one family occurring through polyploidy followed by chromosome rearrangements. In *A. hypogaea*, 24 AhSTP genes were clustered into 18 segmental duplication events in 18 chromosomes except A02 and B08 (Fig. 4A). Intriguingly, *AhSTP9*, *AhSTP15*, *AhSTP28*, and *AhSTP33* exhibited segmental duplication with each other on chromosome A05, A09, B05 and B09. The same phenomenon was also observed in

AhSTP13, *AhSTP14* and *AhSTP32* on chromosome A07, A08 and B07 (Fig. 4A). Moreover, *AhSTP13* and *AhSTP31* were identified as segmental duplication genes. Besides the segmental duplication events, a tandem duplication event region harboring 2 genes (*AhSTP35/36*) was identified on chromosome B10. The results indicated that gene duplication might generate some AhSTP genes and segmental duplication played a key role in driving the AhSTP evolution.

Two comparative syntenic maps of peanut with four species, including the peanut ancestors (*A. duranensis* and *A. ipaënsis*) and two other dicots (Arabidopsis and soybean), were constructed to further reveal the phylogenetic mechanisms of peanut STP gene family (Fig. 4B and Fig. 4C). Twenty-eight AhSTP genes showed syntenic relationship with 14 AdSTP genes in *A. duranensis*, followed by *A. ipaënsis* (28 AhSTP/15 AiSTP), Arabidopsis (6 AhSTP/7 AtSTP) and soybean (17 AhSTP /17 GmSTP). Orthologous pairs between peanut and the four species (*A. duranensis*, *A. ipaënsis*, Arabidopsis and soybean) were 28, 28, 7, 44. Some AhSTP genes exhibited to associate with more than one gene pair, especially between peanut and soybean STP genes. For instance, both *AhSTP14* and *AhSTP32* were associated with four GmSTP genes. These genes might play an important role in the evolution of STP gene family.

Six (*AhSTP5/16/18/20/26/36*) of the 36 AhSTP genes were not found collinear STP genes in its ancestors *A. duranensis* and *A. ipaënsis*, which may indicate that these genes diverged after the formation of cultivated peanut. And some STP genes in *A. duranensis* and *A. ipaënsis*, such as *AdSTP7/14* and *AiSTP9*, were not found collinear STP genes in *A. hypogaea*, implied that gene loss may happen during *A. hypogaea* evolution. In addition, 6 AhSTP genes (*AhSTP3/15/19/32/33/35*) were identified collinear genes among all the four species, indicating that these orthologous pairs may exist before the species divergence.

Ka and Ks calculation of orthologous pairs

To better understand the evolution of peanut STP gene family, the Ka/Ks ratios of the STP gene pairs were calculated. All segmental and tandem duplicated AhSTP gene pairs, and most of the orthologous STP pairs between peanut and its ancestors, had Ka/Ks < 1, indicating that the peanut STP gene family might have experienced strong purifying selective pressure during evolution (Table 2, Table 3 and Table S3).

Table 2
Non-synonymous (Ka) and synonymous substitution rate (Ks) of orthologous STP gene pairs in *A. hypogaea*

Orthologous pairs		Ks	Ka	Ka/Ks	Divergence time (MYA)
AhSTP1	*AhSTP18*	0.0403	0.0021	0.0521	2.4543
AhSTP3	*AhSTP19*	0.1619	0.0062	0.0383	9.8599
AhSTP6	*AhSTP22*	0.1634	0.0127	0.0777	9.9513
AhSTP7	*AhSTP25*	0.0276	0.0042	0.1522	1.6809
AhSTP9	*AhSTP15*	1.6317	0.2230	0.1367	99.3727
AhSTP9	*AhSTP28*	0.0332	0.0021	0.0633	2.0219
AhSTP9	*AhSTP33*	1.6955	0.2183	0.1288	103.2582
AhSTP10	*AhSTP27*	0.0478	0.0063	0.1318	2.9111
AhSTP11	*AhSTP30*	0.0655	0.0062	0.0947	3.9890
AhSTP12	*AhSTP29*	0.1133	0.0104	0.0918	6.9001
AhSTP13	*AhSTP14*	1.5043	0.2242	0.1490	91.6139
AhSTP13	*AhSTP31*	0.0326	0.0107	0.3282	1.9854
AhSTP13	*AhSTP32*	1.5451	0.2225	0.1440	94.0987
AhSTP14	*AhSTP32*	0.0541	0.0085	0.1571	3.2948
AhSTP15	*AhSTP28*	1.5299	0.2258	0.1476	93.1730
AhSTP15	*AhSTP33*	0.0134	0.0042	0.3134	0.8161
AhSTP17	*AhSTP34*	0.0597	0.0021	0.0352	3.6358
AhSTP28	*AhSTP33*	1.5847	0.2210	0.1395	96.5104
AhSTP35	*AhSTP36*	0.1501	0.0199	0.1326	9.1413

Table 3
Non-synonymous (Ka) and synonymous substitution rate (Ks) of nearest orthologous STP gene pairs between *A. hypogaea* and its ancestor

Genome	orthologous pairs		Ks	Ka	Ka/Ks	Nearest divergence time (MYA)
AA	*AdSTP1*	*AhSTP1*	0.0000	0.0000	-	0.0000
	AdSTP2	*AhSTP2*	0.0286	0.0021	0.0734	1.7418
	AdSTP3	*AhSTP3*	0.0140	0.0021	0.1500	0.8526
	AdSTP4	*AhSTP6*	0.0203	0.0000	0.0000	1.2363
	AdSTP5	*AhSTP4*	0.0000	0.0000	-	0.0000
	AdSTP6	*AhSTP7*	0.0000	0.0000	-	0.0000
	AdSTP6	*AhSTP25*	0.0276	0.0042	0.1522	1.6809
	AdSTP8	*AhSTP9*	0.0263	0.0021	0.0798	1.6017
	AdSTP8	*AhSTP28*	0.0197	0.0042	0.2132	1.1998
	AdSTP9	*AhSTP10*	0.0202	0.0021	0.1040	1.2302
	AdSTP10	*AhSTP11*	0.0201	0.0066	0.3284	1.2241
	AdSTP11	*AhSTP12*	0.0286	0.0041	0.1434	1.7418
	AdSTP12	*AhSTP14*	0.0000	0.0000	-	0.0000
	AdSTP13	*AhSTP15*	0.0000	0.0000	-	0.0000
	AdSTP13	*AhSTP33*	0.0134	0.0042	0.3134	0.8161
	AdSTP15	*AhSTP17*	0.0441	0.0021	0.0476	2.6857
BB	*AiSTP3*	*AhSTP19*	0.0000	0.0000	-	0.0000
	AiSTP4	*AhSTP22*	0.0000	0.0000	-	0.0000
	AiSTP5	*AhSTP21*	0.0000	0.0000	-	0.0000
	AiSTP6	*AhSTP23*	0.0000	0.0000	-	0.0000
	AiSTP7	*AhSTP25*	0.0000	0.0000	-	0.0000
	AiSTP8	*AhSTP27*	0.0000	0.0000	-	0.0000
	AiSTP10	*AhSTP30*	0.0284	0.0041	0.1444	1.7296
	AiSTP11	*AhSTP29*	0.0070	0.0021	0.3000	0.4263
	AiSTP12	*AhSTP31*	0.0000	0.0000	-	0.0000
	AiSTP12	*AhSTP13*	0.0326	0.0107	0.3282	1.9854

Genome	orthologous pairs		Ks	Ka	Ka/Ks	Nearest divergence time (MYA)
	AiSTP13	*AhSTP32*	0.0000	0.0000	-	0.0000
	AiSTP14	*AhSTP15*	0.0134	0.0042	0.3134	0.8161
	AiSTP14	*AhSTP33*	0.0000	0.0000	-	0.0000
	AiSTP15	*AhSTP34*	0.0000	0.0000	-	0.0000
	AiSTP16	*AhSTP35*	0.1501	0.0199	0.1326	9.1413

Further, the divergence time of the STP gene pairs was also estimated to characterize the formation of the peanut STP genes. The divergence time of the orthologous STP pairs between peanut and its ancestors showed that most of the genes were diverged after the formation of A and B genome, according to the divergence time of A and B genome by Bertioli [32] (Table 3 and Table S3). All the Ks values of 5 orthologous STP pairs in A genome and 10 orthologous STP pairs in B genome were zero (Table 3), which revealed these genes kept conservation during the evolution of allotetraploid peanut formed from its wild diploid ancestors. These also suggested that the evolution of AhSTP genes was asymmetrically in A and B sub-genome and AhSTP genes in B sub-genome were possible to be more conservative than in A sub-genome. Moreover, *AhSTP25* (B04) and *AhSTP33* (B09) in B sub-genome showed nearest divergence time with orthologous STP pairs *AdSTP6/AhSTP7* (A04) and *AdSTP13*/*AhSTP15* (A09) in A genome, respectively (Table 3, Fig. 5). Similarly, there were no STP genes on chromosome A07 in *A. duranensis*, while, in *A. hypogaea*, *AhSTP13*, on chromosome A07, showed nearest divergence time with orthologous STP pairs *AiSTP12/ AhSTP31*(B07) in B genome (Table 3, Fig. 5). This result implied the location conversion of these AhSTP genes might happen through genetic exchange (or homologous chromosome rearrangements) between A and B sub-genome in *A. hypogaea* after the polyploidy event. The divergence time of tandem duplicated gene pair AhSTP35/36 was much earlier than that of the A and B genome as estimated in previous studies [33, 34]. And the divergence time of orthologous pairs *AiSTP16/ AhSTP35* was also much earlier than that of the other nearest orthologous pairs showing the complicated formation of these genes.

Gene expression Profile Analysis

The transcriptome data, derived from different developmental stages of 22 peanut tissues [35] (Table S4), were used to investigate the expression patterns of all 36 AhSTP genes. All the 36 AhSTP genes were expressed with different expression patterns among development stage of peanut (Fig. 6). Eleven AhSTP genes were expressed in all the 22 tissues (FPKM > 0) and 4 genes (*AhSTP1/3/19/28*) showed constitutive expression with FPKM > 1 in all the tissues. Based on the FPKM value, the transcripts of AhSTP genes were exhibited a most abundance in the tissue 13 (Peg tip to fruit Pattee 1). *AhSTP3* showed the highest total transcript abundance of all the 22 tissues, and also showed the highest transcript abundance in tissue 1–4, 7, 11, 13–17 and 22. *AhSTP4* was the highest expressed gene in tissue 5; while *AhSTP* 19 was the highest expressed gene in tissue 6, 8, 10, 12; *AhSTP 28* was the highest expressed gene in tissue 9, *AhSTP 9* was the highest expressed gene in tissue 18–19; *AhSTP* 18 was the

most expressed gene in tissue 20–21(Fig. 6). Furthermore, some genes also exhibited preferential expression in the detected tissues. 13 genes in pistil (tissue 10), 3 genes (*AhSTP2/4/21*) in repr shoot (tissue 5), 2 genes (AhSTP6/22) in nodule (tissue 7) and 2 genes (*AhSTP13/31*) in seed pat 5 and 6 (tissue 18 and 19) showed relatively highest transcript abundances (Fig. 6).

The transcriptional level of 36 AhSTP genes was further analyzed to understand the gene expression in response to peg elongation and pod development. Four genes (*AhSTP 3/9/19/28*), were found significantly high expression in the stage of peg tip to fruit Pattee (tissue 13) (Fig. 7A), indicating these genes may participate the process of fertilized egg development in Tifrunner. The three constitutive expression genes (*AhSTP3/19/28*), had relative lower expression from development tissue 13 to 14, and kept up-regulated expression from tissue 14 to 16, while *AhSTP9*, an induced expression gene, kept relatively high expression in the stage of fruit pat 1(tissue 14) to pericarp Pat 5 (tissue 16). However, no significant expression change of the genes was observed in seed developmental stage of Tifrunner. Moreover, using the gene expression data of cultivated peanut subsp. *fastigiata* ICGV 91114 [35], the four gene expression were analyzed in seed developmental stage of 5, 15 and 25 days after pegging. The four genes were found significantly high expression in seed_25, a seed developing phase of seed coat epidermal cell differentiation, fatty acid and defense proteins synthesis [36] (Fig. 7B). Overall, these results suggested that some AhSTP genes may participate in early pod formulation and seed development, and have different expression profile in subspecies of cultivated peanut.

Gene variation analysis

To explore the cause of different expression profiles of *AhSTP3/9/19/28* during seed development in peanut subspecies, genic-SSR in the four genes were identified with no SSR in *AhSTP3*, three SSR in both *AhSTP9* and *AhSTP19* and four SSR in *AhSTP28* (Table S5). SSR markers were successfully developed for all the SSR loci and used for genotyping the 37 diverse cultivars (Table S6). Unfortunately, no allelic variation or substantial structure variation in the three genes was detected in the different cultivars of subsp. *hypogaea* or subsp. *fastigiata*.

Secondly, gene sequence of the four genes was extracted from the subsp. *hypogaea* and subsp. *fastigiata* genomes. Sequence analysis showed that *AhSTP19* kept consistent in gene level of the genomes, eight-base deletion and single-base insertion were detected in the first intron of *AhSTP9* and *AhSTP28* respectively, and three single-base substitution were detected in the coding region sequence (CDS) of *AhSTP3* (Fig.S1A and Fig.S1B). Two of the three substitutions in *AhSTP3* were non-synonymous substitutions (Fig.S1B and Fig.S1C), occurring in 804 and 993 bp of the CDS, resulted in glutamic acid replaced by lysine and aspartic acid transformed to glycine in 268 and 328 aa of the protein from subsp. *hypogaea* to subsp. *fastigiata*. Protein secondary structure prediction of AhSTP3 indicated that the extended strand increased while the beta turn and random coil decreased (Fig.S2), and the protein binding site increased two sites in subsp. *fastigiata* (Fig.S3).

Thirdly, the promoter region of the genes, the initiator codon upstream 2000 bp sequences, was extracted from the genomes. Sequence alignment analysis indicated that no variation was detected in the promoter

region of *AhSTP9* and *AhSTP28*, while, single-base substitution or insertion/deletion was observed in *AhSTP3* and *AhSTP19* respectively (Fig.S1E). The single-base substitution in *AhSTP3* promoter region caused the AT1-motif deletion in subsp. *fastigiata*, which is an important part of light responsive module (Fig.S1D, Table S7). The single-base substitution, four-base insertion and deletion in *AhSTP19* promoter region resulted in changes of cis-acting regulatory elements (CREs) in the subspecies, such as deletion of CAAT-box, TATA-box and AT ~ TATA-box (Table S7).

The above results indicated that multiple variations were discovered in the four genes or promoter region of the two subspecies, and were potential causes for the different expression profiles during seed development.

Discussion

The release of peanut genomes [32, 34, 37] provided an availability to explore the peanut STP gene family and discover its potential function in pod formulation and development. In this study, we first analyzed protein properties, gene phylogenetic, structure, evolution, expression profiles and gene variation of STP gene family in cultivated peanut. A total of 36 AhSTPs was identified in cultivated peanut genome through HMMER and BLAST search, and named *AhSTP1* to *AhSTP36*, with 17 AhSTPs in A sub-genome and 19 AhSTPs in B sub-genome. To better understand the evolution of AhSTP genes, STP gene family in the two diploid wild peanut genomes was identified and named in the same way as cultivated peanut, with *AdSTP1* to *AdSTP15* in A genome and *AiSTP1* to *AiSTP16* in B genome.

STP proteins contain a common structure with 12 transmembrane domains in plants [1]. As observed in tomato [6], Cabbage [10] and cassava [8], the loss of TMD, ranging from one to four TMDs, were also occurring in AhSTPs, AdSTPs and AiSTPs. Unlike other species, 3 STP genes (*AdSTP1* and *AiSTP1/16*) in the diploid genomes contained excessive TMDs. *AiSTP16* contained double sugar_tr domain, and all of the conserved motifs were doubled at least as well, while *AhSTP35*, the orthologous gene of *AiSTP16*, contained only 1 sugar_tr domain. These results indicated that the domain gain and loss may have happened during evolution from *A.ipaënsis* to *A.hypogaea* as happened in other gene families [38].

In this study, comparison of STP genes in cultivated peanut with other dicot plants had revealed that cultivated peanut contained more STP genes and the STP gene number in sub-genome remained comparatively stable through evolution from its ancestors. Gene duplication is a primary driving force leading to functional speciation and diversification in evolution [39, 40]. Tandem and segmental duplication events play a critical role in the expansion of STP gene family [8–10]. Here, segmental duplication events were the major force driving the expansion of AhSTP genes. Evidence of STP Gene loss and new STP gene gain were also observed during the polyploid of cultivated peanut as in cabbage [10] and cassava [8]. In cultivated peanut, no AhSTP gene was identified collinear with *AdSTP7/14* and *AiSTP9* in its ancestor genomes, while, 6 AhSTP were not found collinear STP genes from its ancestor genomes. These results indicated that AhSTP gene family was generated by different patterns of evolution events, and the 6 genes may have new features.

Moreover, the Ks analysis of orthologous STP genes pairs provided more clues about the evolution of STP gene family in cultivated peanut. Previous studies have showed that after polyploid formation of allotetraploid peanut, A and B sub-genomes were subjected to asymmetric homoeologous sequence exchanges (or homeologous chromosomes rearrangement), gene family expansion and contraction, homoeolog expression divergence, and selection [33, 34, 37]. In this study, Ks value or divergence time of orthologous STP gene pairs between peanut and its ancestors', indicated that STP genes in cultivated peanut B sub-genome were more conservation than in A sub-genome. This may due to the human domestication, which has larger effects on homoeologous structural variation genes of A sub-genome in cultivated peanut, according to the study of Yin [33]. In addition, genetic exchange (or homeologous chromosomes rearrangement) between the sub-genomes was also observed in this study, the divergence time of *AhSTP25* (B04), *AhSTP33* (B09) and *AhSTP13* (A07) implied these genes were likely to come from homeologous chromosome A04, A09, and B07. All these results suggested that the AhSTP genes were more advanced in terms of evolution in A sub-genome, seemed to have played important roles in plant adaption, and explained why A sub-genome was more different to its ancestors *A. duranensis* from the perspective of a gene family.

Phylogenetic analysis showed 36 AhSTPs were clustered into four groups with the other STP proteins and AhSTPs grouped with AtSTP, GmSTP, AdSTP and AiSTP in each group. A close relationship was observed between AhSTPs and its ancestor STPs (AdSTP and AiSTP), then the GmSTPs, and the relationship between STPs are consistent with the species divergence [34, 37]. Homologous proteins clustered with the characterized proteins in the same group possibly possessing similar biochemical properties [8]. Group I contained AtSTP1/12, group II contained AtSTP4/9/10/11, group III contained AtSTP2/6/7/8/13/14, and group IV contained AtSTP3/5. Previous studies about AtSTP genes provide valuable clues about the functional role of AhSTP genes that involved in the specific peanut physiological process. For example, in group I, *AtSTP1* is a guard cell-specific localization gene involved in carbon acquisition and plays a possible role in osmoregulation [16, 41], while, *AtSTP12* is highly expressed in reproductive organs, and its protein product might contribute to sugar uptake into the pollen tube and the embryo sac [41].

The functional roles of some AhSTP genes involved in specific peanut physiological process were inferred through valuable clues. For example, *AhSTP3* and *AhSTP19*, a segmental gene pair, exhibited high expression in the stage peg tip developed to form a pod, which is a stage embryo cell division facilitating geocarpic pod [30]. While their orthologs in Arabidopsis, *AtSTP12*, located in the inner integument, could uptake sugar to supply the embryo development [41], indicating that *AhSTP3/19* may share the similar function in cultivated peanut. *AhSTP35* was preferentially expressed in the pistil, and with no expression in other detected tissues. Interestingly, its orthologs in Arabidopsis, *AtSTP11* was found to be the most prominent AtSTP in germinating pollen and in the growing pollen tube [42, 43], participating in the allocation of sugars to growing pollen tubes. *AhSTP15* and *AhSTP33* were orthologs of *AtSTP6*, a pollen-specific H^+-monosaccharide symporter, and their function should be further studied. *AhSTP32* exhibited expression in all the tissues detected, while its ortholog, AtSTP5 might be a

pseudogene with a nonfunctional protein product [21], indicating *AhSTP32* may have no function in cultivated peanut.

Peg is an important yield related trait, with the capacity for embryo positive gravitropism transportation and penetrating the soil to form subterranean pods [30]. The development stage from peg to pod is a biological process participated by phytohormone, signals, energy, et al. [30], while monosaccharides (e.g., glucose, fructose, and mannose) are important to generate energy and synthesize cellulose and starch [4, 21]. Therefore, it is important to investigate the expression of AhSTP genes in the peg and pod of cultivated peanut and dissect the potential roles of these genes in sugar distribution. Previous studies in Arabidopsis have showed that STP genes are always expressed coupling with cell wall invertase genes [44], and after the sugar cleavage by cell wall invertases, the resulting monosaccharides could be imported into the inner integument by AtSTP12 to supply the early development of embryo [21]. Using the transcriptome data of subsp. *hypogaea* Tifrunner [35], four AhSTP genes were found high expressed in the stage of pod formulation, and two genes were orthologs of *AtSTP12*, indicating these genes may have important function in early development of embryo. While, the embryo of cultivated peanut is unique in positive gravitropism transportation under soil through the tube tissue peg, then the embryo development to a pod, suggesting that the tissue these genes performing function perhaps had altered. Meaningful, using the transcriptome data of ICGV 91114 [36], a subsp. *fastigiata* cultivated peanut, we found these genes also high expressed in the developing seed, while in the subsp. *hypogaea* no significant expression change were observed in these genes during seed developing, indicating these genes may associate with seed development in subsp. *fastigiate* and the gene function need to study further. Therefore, AhSTP genes exhibiting high expression in the pod at the early growth stage or in the seed at the mature stage may be involved in pod and seed development.

Gene expression profile is a complex biological process regulated by a variety of factors, such as alternative splicing, epigenetic modifications, etc.[45], and gene variation is probably to be the most direct cause for altering gene expression pattern. The expansion or contraction of SSR, the transition or transversion of SNP and the insertion or deletion of InDel in gene sequence or promoter region, all can lead to changes in gene expression pattern and function via mutation in CDS, UTR, intron or promoter region[46–48]. Gene sequence of *AhSTP19* was found conservation in the two subspecies, but SNP and InDel in the promoter region caused altering of CREs. None of genic-SSR, CDS or promoter variation was observed in AhSTP9/28, but small altering was found in the first intron of the genes. For these genes, more evidences are needed to verify the effect on gene expression patterns and functions. While, for AhSTP3, farther study is essential to evaluate the effect of two non-synonymous substitutions in the CDS on protein structure and function. Overall, the above findings provide valuable clues for the different expression profiles of *AhSTP3/9/19/28* during seed development in peanut subspecies.

Conclusions

This study presents the first genome-wide analysis of STP gene family in cultivated peanut genome. Thirty-six AhSTP genes were characterized and classified into four groups, with high similar exon-intron

structures and motif compositions at the bottom branch. Synteny analysis and phylogenetic comparison of STP genes from its ancestor and two other dicots provided valuable clues for the evolutionary characteristics of peanut STP genes. Segmental duplication events were the major force driving the expansion of AhSTP genes, and homeologous chromosomes rearrangement may lead the exchange of STP genes between the A and B sub-genome. The transcriptome data exhibited distinct expression patterns of AhSTP genes in various tissues, and four AhSTP genes may involve in the development of pod and seed. Gene variation was speculated to be potential causes of the four genes with different expression profiles during seed developing phase in peanut subspecies. This study provided a foundation for further understanding the biological roles of STP genes in cultivated peanut.

Materials And Methods

Identification of STP genes in Arachis hypogaea **and its ancestors** Arachis duranensis, Arachis ipaënsis **genomes**

The genome sequence and annotated gene models of *A. hypogaea* (subsp. *hypogaea* Tifrunner), *A. duranensis* (*Aradu*.V14167, AA, 2n = 20) and *A. ipaënsis* (*Araip*.K30076, BB, 2n = 20) were downloaded from Peanutbase (https://peanutbase.org/). The sugar_tr domain (PF00083), download from Pfam (http://pfam.xfam.org/), was used to search sugar transporter proteins in the three genomes by HMMER (Version 3.0, https://hmmer.janelia.org/) and the best match sequence were found using a blast search with 14 AtSTPs and 29 OsSTPs [5]. To confirm the results, all protein sequences were checked using the Pfam and SMART program (http://smart.embl.de). The STP protein sequences of Arabidopsis (AtSTP), rice (OsSTP), soybean (GmSTP) were obtained from TAIR (http://www.Arabidopsis.org), RGAP7 (http://rice.plantbiology.msu.edu/) and Soybase (https://www.soybase.org/).

Protein properties and phylogenetic analysis

The properties of peanut STP, including protein sequence length, MW, pI, GRAVY and AI, were analyzed using the ProtParam tool (https://web.expasy.org/protparam/). TMD and subcellular localization were predicted using TMHMM tool (https://services.healthtech.dtu.dk/service.php?TMHMM-2.0) and Plant-mPLoc (http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/), respectively. Multiple alignment of peanut, its ancestors, soybean and Arabidopsis STP sequences, was performed by ClustalW (Table S8). The neighbor-joining (NJ) phylogenetic trees were constructed using MEGA7 software with the bootstrap of 1000 replicates.

Gene structure and conserved motif analysis

Peanut STP gene structures were analyzed using TBtools software [33] via comparing the cDNA to the gene sequence. Conserved motifs in peanut STP proteins were analyzed using MEME programs (http://meme-suite.org/tools/meme) with the following parameters: optimum width, 20–60; number of repetitions, any; maximum number of motifs, 12.

Synteny analysis

Multiple Collinearity Scan toolkit (MCScanX) was adopted to analyze the STP gene duplication events, with the default parameters. The circos map of peanut STP genes was constructed to reveal the synteny relationship between cultivated peanut STP genes by TBtool software [49]. The syntenic maps were constructed using Tbtool software to reveal the synteny relationship of STP genes between cultivated peanut and selected plants. Non-synonymous (Ka) and synonymous (Ks) substitution of each duplicated STP gene pairs were calculated using DnaSP 6.0 [50]. The divergence time was estimated according to the published method: T = Ks/2*m, in which m means molecular clock [32, 33]. The average mutation rate for *Arachis* was considered as 8.12×10^{-9} mutations per base per year, according to the previous work [32, 33].

Gene expression Profile Analysis

To explore the expression profiles of AhSTPs in 22 different peanut tissues of Tifrunner, the FPKM (fragments per kilobase per million fragments mapped) values of the AhSTP genes were download from the peanut developmental transcriptome map dataset in Peanutbase (https://peanutbase.org/) (BioProject ID: PRJNA291488) submitted by Clevenger [35]. The 22 peanut tissues, including leaf, root, floral organ, peg, pod, were described in Table S4. The heatmap of AhSTPs expression profiles was performed using the Tbtools software [49] based on the normalization method of Z-score standardization. The gene expression atlas for *fastigiata* subspecies of cultivated groundnut were download from National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database (BioProject ID: PRJNA484860) submitted by Sinha [36]. And the genes expression data of the seeds_5, seeds_15 and seeds_25 were used to analyze AhSTPs expression profiles in seed developmental stage.

Plant materials and gene variation analysis

Thirty-seven diverse peanut cultivars were used as plant material to detect the gene sequence variation, with 14 cultivars belonging to subsp. *hypogaea* and 23 cultivars belonging to subsp. *fastigiata* respectively. Detailed information for the 37 peanut cultivars was listed in Table S6. Genomic DNA was extracted from the young leaves of these cultivars using a modified cetyltrimethyl ammonium bromide (CTAB) method. Gene associated simple sequence repeats (genic-SSR) were identified using MISA with the default parameters and primers of the identified SSRs were designed using Primer 3 software described by Zhao [51]. PCR reactions were performed as previously described by Huang [52].

The peanut genome sequence and annotated gene models of subsp. *fastigiata* Shitouqi were downloaded from Peanut genome resource (http://peanutgr.fafu.edu.cn/). Gene sequence pairwise alignment was performed by ClustalW. Protein secondary structure was predicted using SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopm a.html) and PredictProtein (https://predictprotein.org/home). Gene promoter and CRE were predicted using PlantCare (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) and NEW PLACE (https://www.dna.affrc.go.jp/PLACE/?action=newplace).

Abbreviations

AhSTP
peanut (*Arachis hypogaea*) STP; AdSTP:*Arachis duranensis* STP; AiSTP:*Arachis ipaënsis* STP; AtSTP:*Arabidopsis thaliana* STP; GmSTP:soybean (*Glycine max*) STP; OsSTP:rice (*Oryza sativa*) STP; SNP:single nucleotide polymorphism; InDel:insertion-deletion.

Declarations

Availability of data and materials

All data utilizedin this study are included in this article and itsAdditional files.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Guangdong Basic and Applied Basic Research Foundation (2020A1515010636) and the Central Public-interest Scientific Institution Basal Research Fund for Chinese Academy of Tropical Agricultural Sciences (No. 1630102017002).

Acknowledgements

The authors thank to lab members for assistance.

Authors' contributions

ZJX performed the experiments. ZJX, SZ, XWH and XLW analyzed the data. ZJX and SZ wrote the manuscript. ZJX and YL designed the study. All authors have read and approved the final manuscript.

References

1. Yan N. Structural advances for the major facilitator superfamily (MFS) transporters. Trends Biochem Sci. 2013;38(3):151–9.
2. Buttner M. The Arabidopsis sugar transporter (AtSTP) family: an update. Plant Biol (Stuttg). 2010;12(Suppl 1):35–41.
3. Sauer N, Tanner W. The hexose carrier from Chlorella. cDNA cloning of a eucaryotic H+-cotransporter. FEBS Lett. 1989;259(1):43–6.
4. Toyofuku K, Kasahara M, Yamaguchi J. Characterization and expression of monosaccharide transporters (osMSTs) in rice. Plant Cell Physiol. 2000;41(8):940–7.

5. Johnson DA, Thomas MA. The monosaccharide transporter gene family in Arabidopsis and rice: a history of duplications, adaptive evolution, and functional divergence. Mol Biol Evol. 2007;24(11):2412–23.
6. Reuscher S, Akiyama M, Yasuda T, Makino H, Aoki K, Shibata D, Shiratake K. The sugar transporter inventory of tomato: genome-wide identification and expression analysis. Plant Cell Physiol. 2014;55(6):1123–41.
7. Li JM, Zheng DM, Li LT, Qiao X, Wei SW, Bai B, Zhang SL, Wu J. Genome-Wide Function, Evolutionary Characterization and Expression Analysis of Sugar Transporter Family Genes in Pear (*Pyrus bretschneideri* Rehd). Plant Cell Physiol. 2015;56(9):1721–37.
8. Liu Q, Dang H, Chen Z, Wu J, Chen Y, Chen S, Luo L. Genome-Wide Identification, Expression, and Functional Analysis of the Sugar Transporter Gene Family in Cassava (*Manihot esculenta*). Int J Mol Sci. 2018;19(4):987.
9. Jiu S, Haider MS, Kurjogi MM, Zhang K, Zhu X, Fang J. Genome-wide Characterization and Expression Analysis of Sugar Transporter Family Genes in Woodland Strawberry. Plant Genome. 2018;11(3):1–16.
10. Zhang W, Wang S, Yu F, Tang J, Yu L, Wang H, Li J. Genome-Wide Identification and Expression Profiling of Sugar Transporter Protein (STP) Family Genes in Cabbage (*Brassica oleracea* var. capitata L.) Reveals their Involvement in Clubroot Disease Responses. Genes (Basel). 2019;10(1):71.
11. Truernit E, Schmid J, Epple P, Illig J, Sauer N. The sink-specific and stress-regulated Arabidopsis STP4 gene: enhanced expression of a gene encoding a monosaccharide transporter by wounding, elicitors, and pathogen challenge. Plant Cell. 1996;8(12):2169–82.
12. Büttner M, Truernit E, Baier K, Scholz-Starke J, Sontheim M, Lauterbach C, Huss VAR, Sauer N. AtSTP3, a green leaf-specific, low affinity monosaccharide-H + symporter of *Arabidopsis thaliana*. Plant Cell Environ. 2000;23(2):175–84.
13. Harmer SL, Hogenesch JB, Straume M, Chang HS, Han B, Zhu T, Wang X, Kreps JA, Kay SA. Orchestrated transcription of key pathways in Arabidopsis by the circadian clock. Science. 2000;290(5499):2110–3.
14. Stadler R, Büttner M, Ache P, Hedrich R, Ivashikina N, Melzer M, Shearson SM, Smith SM, Sauer N. Diurnal and light-regulated expression of AtSTP1 in guard cells of Arabidopsis. Plant Physiol. 2003;133(2):528–37.
15. Sauer N, Friedländer K, Gräml-Wicke U. Primary structure, genomic organization and heterologous expression of a glucose transporter from *Arabidopsis thaliana*. Embo j. 1990;9(10):3045–50.
16. Sherson SM, Hemmann G, Wallace G, Forbes S, Germain V, Stadler R, Bechtold N, Sauer N, Smith SM. Monosaccharide/proton symporter AtSTP1 plays a major role in uptake and response of Arabidopsis seeds and seedlings to sugars. Plant J. 2000;24(6):849–57.
17. Truernit E, Stadler R, Baier K, Sauer N. A male gametophyte-specific monosaccharide transporter in Arabidopsis. Plant J. 1999;17(2):191–201.

18. Schneidereit A, Scholz-Starke J, Buttner M. Functional characterization and expression analyses of the glucose-specific AtSTP9 monosaccharide transporter in pollen of Arabidopsis. Plant Physiol. 2003;133(1):182–90.
19. Scholz-Starke J, Buttner M, Sauer N. AtSTP6, a new pollen-specific H+-monosaccharide symporter from Arabidopsis. Plant Physiol. 2003;131(1):70–7.
20. Schneidereit A, Scholz-Starke J, Sauer N, Buttner M. AtSTP11, a pollen tube-specific monosaccharide transporter in Arabidopsis. Planta. 2005;221(1):48–55.
21. Rottmann T, Klebl F, Schneider S, Kischka D, Ruscher D, Sauer N, Stadler R. Sugar Transporter STP7 Specificity for l-Arabinose and d-Xylose Contrasts with the Typical Hexose Transporters STP8 and STP12. Plant Physiol. 2018;176(3):2330–50.
22. Norholm MH, Nour-Eldin HH, Brodersen P, Mundy J, Halkier BA. Expression of the Arabidopsis high-affinity hexose transporter STP13 correlates with programmed cell death. FEBS Lett. 2006;580(9):2381–7.
23. Poschet G, Hannich B, Buttner M. Identification and characterization of AtSTP14, a novel galactose transporter from Arabidopsis. Plant Cell Physiol. 2010;51(9):1571–80.
24. Robledo G, Lavia GI, Seijo G. Species relations among wild Arachis species with the A genome as revealed by FISH mapping of rDNA loci and heterochromatin detection. Theor Appl Genet. 2009;118(7):1295–307.
25. Grabiele M, Chalup L, Robledo G, Seijo GJPS, Evolution. Genetic and geographic origin of domesticated peanut as evidenced by 5S rDNA and chloroplast DNA sequences. Plant Syst Evol. 2012;298(6):1151–65.
26. Seijo G, Lavia GI, Fernandez A, Krapovickas A, Ducasse DA, Bertioli DJ, Moscone EAJAJoB. Genomic relationships between the cultivated peanut (*Arachis hypogaea*, Leguminosae) and its close relatives revealed by double GISH. Am J Bot. 2007;94(12):1963–71.
27. Ramos ML, Fleming G, Chu Y, Akiyama Y, Gallo M, Oziasakins P. Chromosomal and phylogenetic context for conglutin genes in Arachis based on genomic sequence. Mol Genet Genomics. 2006;275(6):578–92.
28. Samoluk SS, Robledo G, Podio M, Chalup L, Ortiz JPA, Pessino SC, Seijo JG. First insight into divergence, representation and chromosome distribution of reverse transcriptase fragments from L1 retrotransposons in peanut and wild relative species. Genetica. 2015;143(1):113–25.
29. Favero AP, Simpson CE, Valls JFM, Vello NA. Study of the evolution of cultivated peanut through crossability studies among *Arachis ipaënsis*, *A. duranensis*, and *A. hypogaea*. Crop Sci. 2006; 46(4):1546–1552.
30. Kumar R, Pandey MK, Roychoudhry S, Nayyar H, Kepinski S, Varshney RK. Peg Biology: Deciphering the Molecular Regulations Involved During Peanut Peg Development. Front Plant Sci. 2019;10:1289.
31. Yu J, Wang J, Lin W, Li S, Li H, Zhou J, Ni P, Dong W, Hu S, Zeng C, Zhang J, Zhang Y, Li R, Xu Z, Li S, Li X, Zheng H, Cong L, Lin L, Yin J, Geng J, Li G, Shi J, Liu J, Lv H, Li J, Wang J, Deng Y, Ran L, Shi X, Wang X, Wu Q, Li C, Ren X, Wang J, Wang X, Li D, Liu D, Zhang X, Ji Z, Zhao W, Sun Y, Zhang Z, Bao J,

Han Y, Dong L, Ji J, Chen P, Wu S, Liu J, Xiao Y, Bu D, Tan J, Yang L, Ye C, Zhang J, Xu J, Zhou Y, Yu Y, Zhang B, Zhuang S, Wei H, Liu B, Lei M, Yu H, Li Y, Xu H, Wei S, He X, Fang L, Zhang Z, Zhang Y, Huang X, Su Z, Tong W, Li J, Tong Z, Li S, Ye J, Wang L, Fang L, Lei T, Chen C, Chen H, Xu Z, Li H, Huang H, Zhang F, Xu H, Li N, Zhao C, Li S, Dong L, Huang Y, Li L, Xi Y, Qi Q, Li W, Zhang B, Hu W, Zhang Y, Tian X, Jiao Y, Liang X, Jin J, Gao L, Zheng W, Hao B, Liu S, Wang W, Yuan L, Cao M, McDermott J, Samudrala R, Wang J, Wong GK, Yang H. The Genomes of *Oryza sativa*: A History of Duplications. PLoS Biol. 2005;3(2):e38.

32. Bertioli DJ, Cannon SB, Froenicke L, Huang G, Farmer AD, Cannon EKS, Liu X, Gao D, Clevenger J, Dash S, Ren L, Moretzsohn MC, Shirasawa K, Huang W, Vidigal B, Abernathy B, Chu Y, Niederhuth CE, Umale P, Araújo AC, Kozik A, Kim KD, Burow MD, Varshney RK, Wang X, Zhang X, Barkley N, Guimarães PM, Isobe S, Guo B, Liao B, Stalker HT, Schmitz RJ, Scheffler BE, Leal-Bertioli SC, Xun X, Jackson SA, Michelmore R, Ozias-Akins P. The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. Nat Genet. 2016;48(4):438–46.
33. Yin D, Ji C, Song Q, Zhang W, Zhang X, Zhao K, Chen CY, Wang C, He G, Liang Z, Ma X, Li Z, Tang Y, Wang Y, Li K, Ning L, Zhang H, Zhao K, Li X, Yu H, Lei Y, Wang M, Ma L, Zheng H, Zhang Y, Zhang J, Hu W, Chen ZJ. Comparison of Arachis monticola with Diploid and Cultivated Tetraploid Genomes Reveals Asymmetric Subgenome Evolution and Improvement of Peanut. Adv Sci (Weinh). 2020;7(4):1901672.
34. Zhuang W, Chen H, Yang M, Wang J, Pandey MK, Zhang C, Chang WC, Zhang L, Zhang X, Tang R, Garg V, Wang X, Tang H, Chow CN, Wang J, Deng Y, Wang D, Khan AW, Yang Q, Cai T, Bajaj P, Wu K, Guo B, Zhang X, Li J, Liang F, Hu J, Liao B, Liu S, Chitikineni A, Yan H, Zheng Y, Shan S, Liu Q, Xie D, Wang Z, Khan SA, Ali N, Zhao C, Li X, Luo Z, Zhang S, Zhuang R, Peng Z, Wang S, Mamadou G, Zhuang Y, Zhao Z, Yu W, Xiong F, Quan W, Yuan M, Li Y, Zou H, Xia H, Zha L, Fan J, Yu J, Xie W, Yuan J, Chen K, Zhao S, Chu W, Chen Y, Sun P, Meng F, Zhuo T, Zhao Y, Li C, He G, Zhao Y, Wang C, Kavikishor PB, Pan RL, Paterson AH, Wang X, Ming R, Varshney RK. The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. Nat Genet. 2019;51(5):865–76.
35. Clevenger J, Chu Y, Scheffler BE, Oziasakins P. A Developmental Transcriptome Map for Allotetraploid *Arachis hypogaea*. Front Plant Sci. 2016;7:1446.
36. Sinha P, Bajaj P, Pazhamala LT, Nayak SN, Pandey MK, Chitikineni A, Huai D, Khan AW, Desai A, Jiang H, Zhuang W, Guo B, Liao B, Varshney RK. Arachis hypogea gene expression atlas (AhGEA) for *fastigiata* subspecies of cultivated groundnut to accelerate functional and translational genomics applications. Plant Biotechnol J. 2020;10: pbi.13374.
37. Bertioli DJ, Jenkins J, Clevenger J, Dudchenko O, Gao D, Seijo G, Lealbertioli SCM, Ren L, Farmer AD, Pandey MK, Samoluk SS, Abernathy B, Agarwal G, Ballén-Taborda C, Cameron C, Campbell J, Chavarro C, Chitikineni A, Chu Y, Dash S, El Baidouri M, Guo B, Huang W, Kim KD, Korani W, Lanciano S, Lui CG, Mirouze M, Moretzsohn MC, Pham M, Shin JH, Shirasawa K, Sinharoy S, Sreedasyam A, Weeks NT, Zhang X, Zheng Z, Sun Z, Froenicke L, Aiden EL, Michelmore R, Varshney RK, Holbrook CC, Cannon EKS, Scheffler BE, Grimwood J, Ozias-Akins P, Cannon SB, Jackson SA, Schmutz J. The

genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. Nat Genet. 2019;51(5):877–84.

38. Xie T, Chen C, Li C, Liu J, Liu C, He Y. Genome-wide investigation of WRKY gene family in pineapple: evolution and expression profiles during development and stress. BMC Genom. 2018;19(1):490.
39. Lynch M, Conery JSJS. The Evolutionary Fate and Consequences of Duplicate Genes. 2000; 290(5494):1151–1155.
40. Moore RC, Purugganan MD. The early stages of duplicate gene evolution. Proc Natl Acad Sci U S A. 2003;100(26):15682–7.
41. Slewinski TL. Diverse Functional Roles of Monosaccharide Transporters and their Homologs in Vascular Plants: A Physiological Perspective. Mol Plant. 2011;4(4):641–62.
42. Wang Y, Zhang W, Song L, Zou J, Su Z, Wu W. Transcriptome Analyses Show Changes in Gene Expression to Accompany Pollen Germination and Tube Growth in Arabidopsis. Plant Physiol. 2008;148(3):1201–11.
43. Qin Y, Leydon AR, Manziello A, Pandey R, Mount DB, Denic SZ, Vasic B, Johnson MA, Palanivelu R. Penetration of the stigma and style elicits a novel transcriptome in pollen tubes, pointing to genes critical for growth in a pistil. PLoS Genet. 2009;5(8):e1000621.
44. Rennie EA, Turgeon R. A comprehensive picture of phloem loading strategies. Proc Natl Acad Sci U S A. 2009;106(33):14162–7.
45. Mack KL, Nachman MW. Gene Regulation and Speciation. Trends Genet. 2017;33(1):68–80.
46. Li YC, Korol AB, Fahima T, Nevo E. Microsatellites within genes: structure, function, and evolution. Mol Biol Evol. 2004;21(6):991–1007.
47. Shastry BS. SNPs: impact on gene function and phenotype. Methods Mol Biol. 2009;578:3–22.
48. Taylor CM, Kamphuis LG, Zhang W, Garg G, Berger JD, Mousavi-Derazmahalleh M, Bayer PE, Edwards D, Singh KB, Cowling WA, Nelson MN. INDEL variation in the regulatory region of the major flowering time gene LanFTc1 is associated with vernalization response and flowering time in narrow-leafed lupin (*Lupinus angustifolius* L.). Plant Cell Environ. 2019;42(1):174–87.
49. Chen C, Xia R, Chen H, He Y. TBtools, a Toolkit for Biologists integrating various HTS-data handling tools with a user-friendly interface. bioRxiv. 2018:289660.
50. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets
Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. Mol Biol Evol. 2017; 34(12):3299–302.
51. Zhao C, Qiu J, Agarwal G, Wang J, Ren X, Xia H, Guo B, Ma C, Wan S, Bertioli DJ, Varshney RK, Pandey MK, Wang X. Genome-Wide Discovery of Microsatellite Markers from Diploid Progenitor Species, Arachis duranensis and A. ipaensis, and Their Application in Cultivated Peanut (*A. hypogaea*). Front Plant Sci. 2017;8:1209.

52. Huang L, Ren X, Wu B, Li X, Chen W, Zhou X, Chen Y, Pandey MK, Jiao Y, Luo H, Lei Y, Varshney RK, Liao B, Jiang H. Development and deployment of a high-density linkage map identified quantitative trait loci for plant height in peanut (*Arachis hypogaea* L.). Sci Rep. 2016;6:39478.

Figures

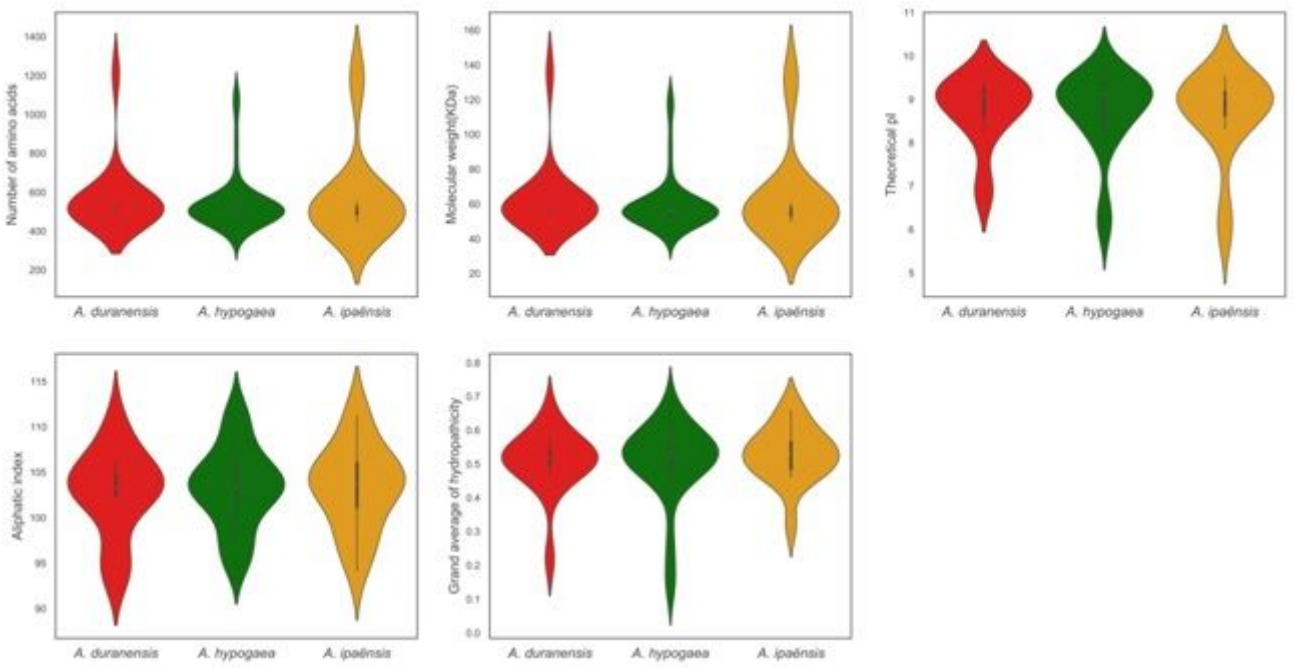

Figure 1

DistributionofSTP protein characteristics in A.hypogaea, A.duranensis and A.ipaënsis.

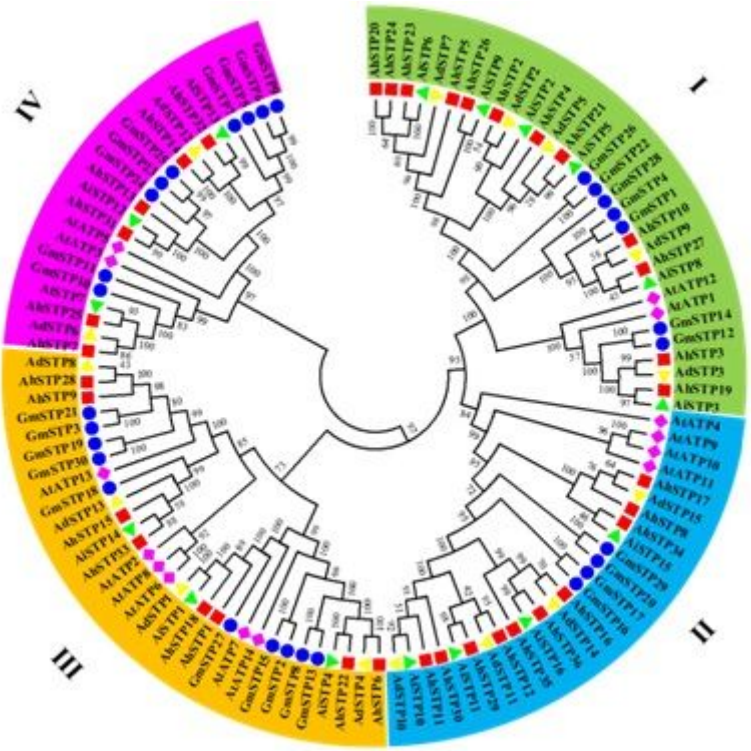

Figure 2

Phylogenetic relationship of STP proteins in five plant species. Phylogenetic analysis of 111 STP proteins from A.hypogaea (36), A.duranensis(15),A.ipaënsis (16), A.thaliana (14) and G.max(30) are marked with red squares, yellow triangles, green triangles, Purple diamonds and blue circles, respectively. Four groups were marked with different background colors.

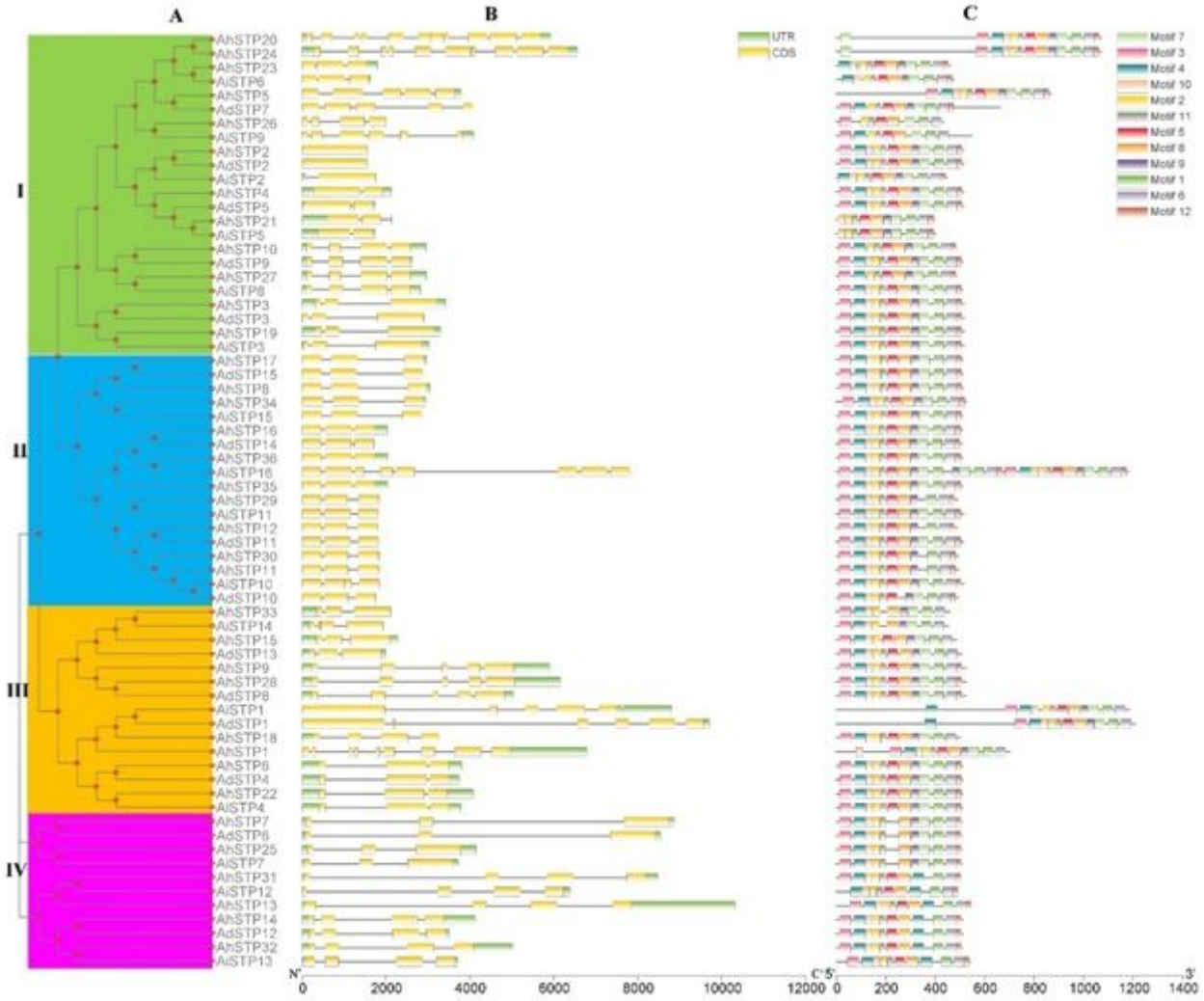

Figure 3

Gene structure and motif composition of peanut STP. (A) An unrooted phylogenetic tree was constructed using full-length peanut STP proteins; (B) Motif composition of STP proteins and the motifs are represented by different color blocks as shown at the offside of the figure and the same color block in different proteinsindicates thesame motif. (C) Gene exon-intron structure of peanut STP proteins.

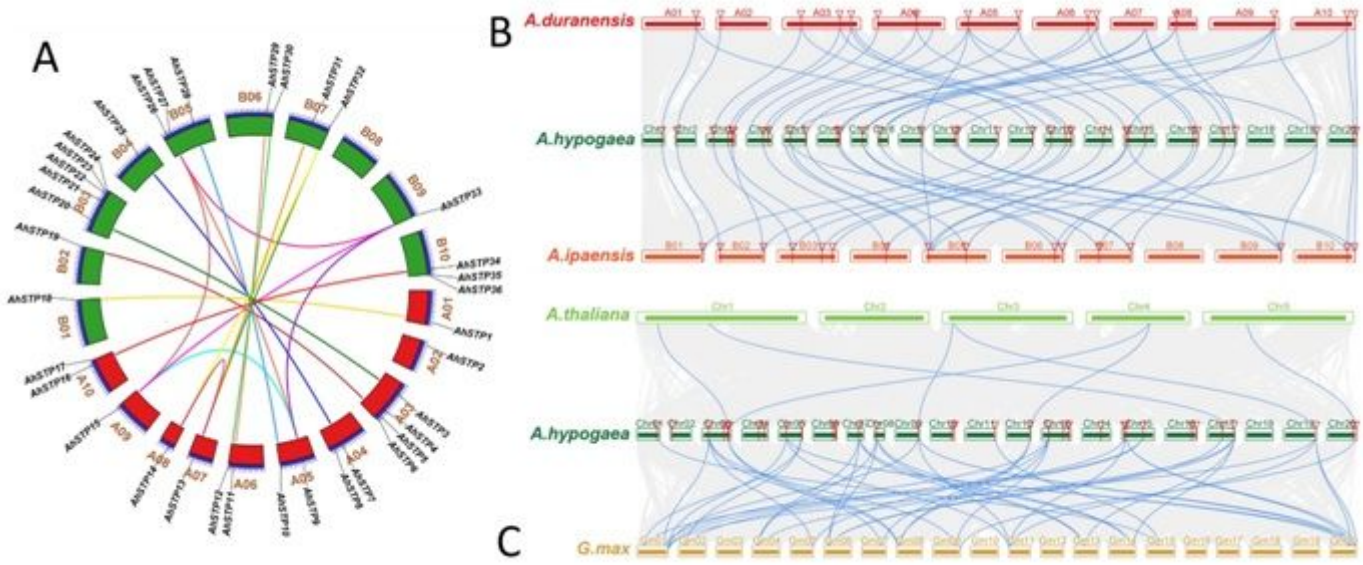

Figure 4

Synteny analysis of peanut STP genes. (A) The chromosomal distribution and segmental relationships of peanutSTP genes. (B) Synteny analysis of STP genes between peanut and its ancestors. (C) Synteny analysis of STP genes between peanut andArabidopsis thaliana, Glycine max. In B and C, Gray lines in the background indicate the collinearblocks within peanut and other plant genomes, while the blue lines highlight the syntenic STP gene pairs.

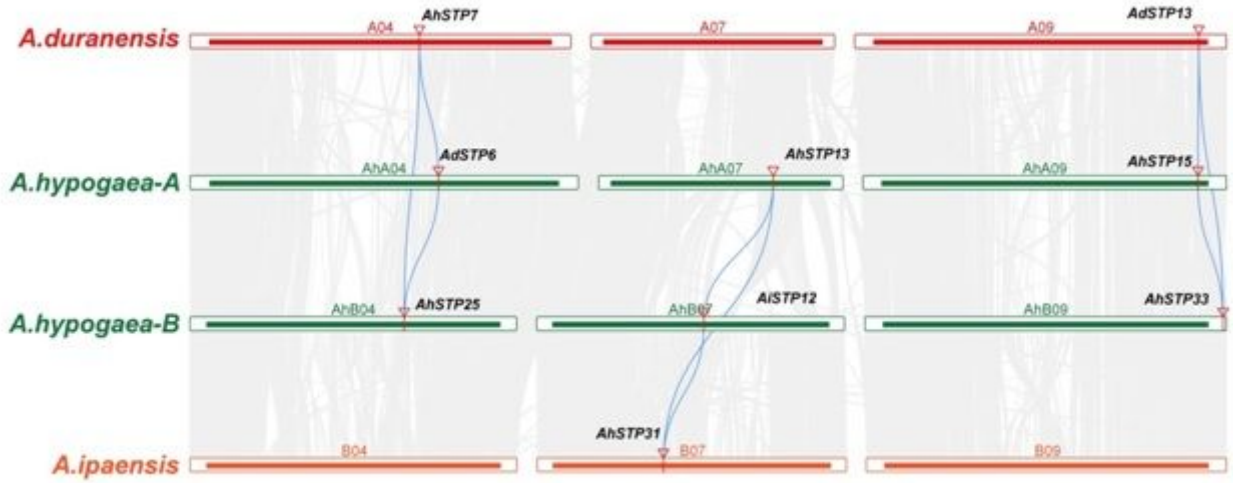

Figure 5

Syntenyanalysis of genetic exchange (or homeologous chromosomes rearrangement)between A and B sub-genome in peanut STP gene family. Gray lines in the background indicate the collinear blocks within peanut sub-A chromesomes, sub-B chromsomes and its ancestors, while the blue lines highlight the syntenic STP gene pairs.

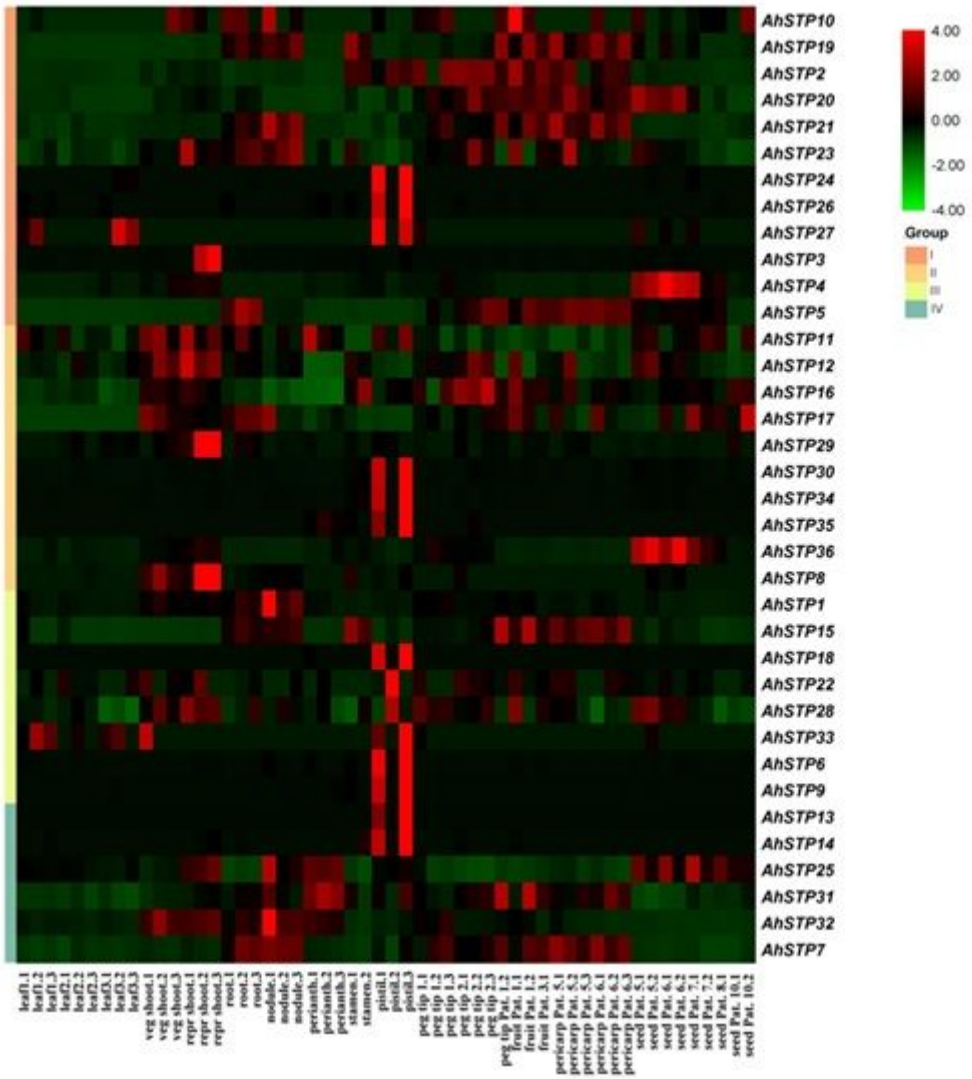

Figure 6

Expression profiles of the peanut STP genes. Expression profiles of AhSTPgenes in 22tissues of different development stages asdescribed inTable S3.

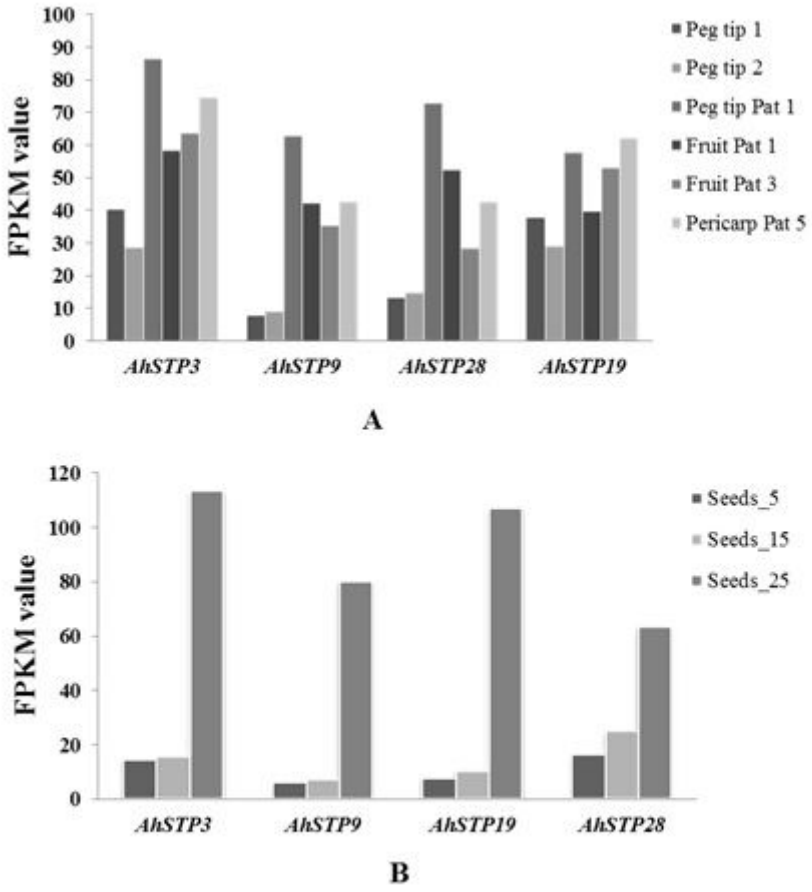

Figure 7

Expression analysis of AhSTP3, AhSTP9, AhSTP19 and AhSTP28 in pod and seed using the transcriptome data. (A) Expression of the four genes in pod early formulate stage from peg tip to fruit Pattee in subsp. hypogaeaTifrunner. (B) Expression of the four genes in three stage of seeds in subsp. fastigiata ICGV 91114.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AdditionfilesFig.S1Fig.S3.pptx
- AdditionfilesFigS1S3.docx
- AdditionfilesTableS1toS8.xlsx