



Identification and Evaluation of the Expansin Genes in *Salix purpurea* Genome

Yang R.X. *, Liu X.R. *, Lan B.L., Wang H., Liu X., Xu J.C. ✉

National Engineering Laboratory for Tree Breeding, Beijing Forestry University, Beijing 100083, China

* These authors contributed equally to this work.

✉ Corresponding author email: jcxu282@sina.com

Tree Genetics and Molecular Breeding, 2020, Vol.10, No.1 doi: [10.5376/tgmb.2020.10.0001](https://doi.org/10.5376/tgmb.2020.10.0001)

Received: 11 May, 2020

Accepted: 14 May, 2020

Published: 14 May, 2020

Copyright © 2020 Yang et al., This article was first published in Molecular Plant Breeding in Chinese, and here was authorized to translate and publish the paper in English under the terms of Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preferred citation for this article:

Yang R.X., Liu X.R., Lan B.L., Wang H., Liu X., and Xu J.C., 2020, Identification and evaluation of the expansin genes in *Salix purpurea* genome, Tree Genetics and Molecular Breeding, 10(1): 1-13 (doi: [10.5376/tgmb.2020.10.0001](https://doi.org/10.5376/tgmb.2020.10.0001))

Abstract Expansin is an indispensable cell wall protein, involved in plant growth and resistance to stress. Based on the sequencing genome data, a total of 34 expansin genes were identified on 14 chromosomes in *Salix purpurea*, which were grouped into 4 subfamilies. The analysis of the physicochemical properties showed that most expansin proteins were in weakly alkaline, had a higher aliphatic index and more stable protein structure. The gene sequence was more conservative among the members in the same subfamily with the amino acid identity between 34.2% and 99.4%, while that was between 15.0% and 39.8% among the members in the different subfamilies. Also, the same subfamily genes had similar exon-intron structures, but the different subfamily genes showed that much differently. The codon usage analysis revealed that there are 5 high-frequency codons in *Salix purpurea* expansin genes. The codon usage bias was much affected by natural selection. The synteny analysis showed that the expansin genes divided in correspondence to whole-genome duplication events in willow. The results here would help us to understand well the role and status of expansin genes in the growth and development of *Salix purpurea*. It will be helpful for further study on the expansins functions.

Keywords *Salix purpurea*; Expansin; Gene family; Codon bias; Gene evolution

Background

Expansins are important components of plant cell wall that can loosen plant cell wall and increase cell flexibility by breaking the hydrogen bonding between cellulose microfibrils and hemicellulose. Therefore, expansin plays an important role in various growth processes of plants, such as seed germination (Yan et al., 2014), root elongation (Noh et al., 2013), fruit ripening (Jiang et al., 2019), drought resistance (Chen et al., 2016), heavy metals toxicity (Zhang et al., 2018a), and salt tolerance (Lu et al., 2013).

Typical expansin in plants consists of three domains, namely the N-terminal signal peptide, the catalysis domain in the middle, and the C-terminal cellulose-binding domain. According to the differences of gene sequence and structural composition, expansins can be divided into four subfamilies: EXPA, EXPB, EXLA and EXLB. In recent years, along with the development of genome sequencing, the expansin genes in more plant species have been identified, such as grape (Dal santo et al., 2013) and poplar (Li et al., 2014) of woody plants, and *Arabidopsis* (Sampedro et al., 2006), maize (Zhang et al., 2014), rice (Shi et al., 2014) of herbaceous plants. Clearly, the expansin genes are not only different in quantity in different species, but also various in gene sequence composition and codon usage pattern.

Willow is an important woody tree species for ecological conservation with rich variety resources. The study here supposed to identify the expansin genes based on the sequenced genome of willow variety *Salix purpurea*, and to understand the characteristics of the willow expansin genes via bioinformatics analysis. The final results would provide some useful information for further exploring the expansin gene function and the effect of the expansin genes on willow growth, development and resistance.

1.2 Analysis of the expansin gene composition in *Salix purpurea*

Statistics showed that the total exons of each expansin gene was in length of 663-861bp, with an average of 774 bp. And the total length of introns varied greatly from 185 bp to 1558 bp. Mostly, the *SpEXPA* subfamily genes contain 2 introns, while *SpEXPB*, *SpEXLA* and *SpEXLB* subfamily genes contain 3, 4 and 4 introns. All the expansin genes had a signal peptide with 17-36 amino acids (Table 1). Thus, the structure of willow expansin genes was subfamily characteristically, as same as the other species.

Table 1 Information of expansin genes in *Salix purpurea* genome

Gene	Length of exons (bp)	Length of introns (bp)	Number of introns	Signal peptide length (bp)	
Phytozome ID	Gene name				
SapurV1A.0041s0110.1	<i>SpEXPA1</i>	789	1 126	2	25
SapurV1A.0222s0210.1	<i>SpEXPA2</i>	754	239	2	23
SapurV1A.0458s0030.1	<i>SpEXPA3</i>	663	808	3	20
SapurV1A.0801s0010.1	<i>SpEXPA4</i>	750	280	2	20
SapurV1A.0241s0270.1	<i>SpEXPA5</i>	717	770	2	26
SapurV1A.0451s0080.1	<i>SpEXPA7</i>	777	801	2	21
SapurV1A.0147s0050.1	<i>SpEXPA8</i>	843	1047	1	31
SapurV1A.0438s0090.1	<i>SpEXPA9</i>	726	214	2	20
SapurV1A.0081s0690.1	<i>SpEXPA10a</i>	747	228	2	21
SapurV1A.0699s0280.1	<i>SpEXPA10b</i>	759	364	3	25
SapurV1A.0732s0010.1	<i>SpEXPA11</i>	807	385	1	20
SapurV1A.0245s0010.1	<i>SpEXPA12</i>	753	304	2	23
SapurV1A.0065s0390.1	<i>SpEXPA13</i>	768	281	2	28
SapurV1A.0281s0020.1	<i>SpEXPA15</i>	777	518	2	21
SapurV1A.0001s1750.1	<i>SpEXPA16</i>	783	771	2	23
SapurV1A.0016s1280.1	<i>SpEXPA17</i>	771	185	2	24
SapurV1A.0384s0260.1	<i>SpEXPA18</i>	771	185	2	24
SapurV1A.0003s0080.1	<i>SpEXPA19</i>	771	248	2	22
SapurV1A.1125s0060.1	<i>SpEXPA20</i>	765	395	2	28
SapurV1A.1212s0130.1	<i>SpEXPA21</i>	753	235	2	21
SapurV1A.0125s0570.1	<i>SpEXPA22</i>	753	209	2	21
SapurV1A.0549s0190.1	<i>SpEXPA23</i>	840	1069	1	36
SapurV1A.0282s0280.1	<i>SpEXPA24</i>	810	372	2	30
SapurV1A.1559s0010.1	<i>SpEXPA25</i>	861	439	2	24
SapurV1A.0584s0100.1	<i>SpEXPA26</i>	861	396	2	24
SapurV1A.1509s0020.1	<i>SpEXPA27</i>	672	279	1	29
SapurV1A.0248s0210.1	<i>SpEXPB1</i>	819	555	3	26
SapurV1A.1037s0080.1	<i>SpEXPB2</i>	789	1144	3	20
SapurV1A.0048s0820.1	<i>SpEXPB3</i>	789	1558	3	22
SapurV1A.0080s0500.1	<i>SpEXLA1</i>	780	906	4	17
SapurV1A.0218s0040.1	<i>SpEXLA2</i>	825	925	4	32
SapurV1A.0107s0210.1	<i>SpEXLB1</i>	723	577	3	24
SapurV1A.0132s0090.1	<i>SpEXLB3</i>	774	985	4	25
SapurV1A.0885s0100.1	<i>SpEXLB4</i>	771	687	4	24

The sequence identity in nucleotides among the willow expansin genes was in range of 31.6%~99.0%, which was really higher than that in amino acids (15.5%~99.4%). Of them, the sequence identity in nucleotides among the subfamily members was from 43.9% to 99.0%, and that in amino acids was from 34.2% to 99.4%. While, the sequence identity in nucleotides and in amino acids among different subfamily members were in range of 31.6%~54.6% and 15.0%~39.8%, respectively (Figure 3A). It also showed the expansin gene subfamily characteristically. Further phylogenetic tree analysis showed that the subfamily genes all clustered together

(Figure 3B). Therefore, sequence alignment can be used for classification of expansin gene subfamily and identification of homologous genes in different species, and promote the understanding of the evolution history of expansin genes via further study of the gene function diversity among different subfamily genes.

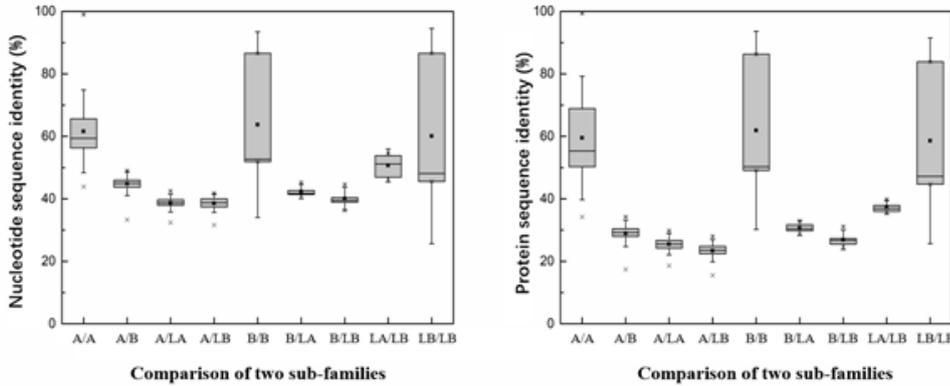


Figure 3 Sequence identity of *Salix purpurea* expansin genes in nucleotides (A) or amino acids (B)

Note: The horizontal lines in the box are the median; The black points in the box are the average; × is the maximum value; × is the minimum value

1.3 Structural analysis of the expansin genes in *Salix purpurea*

Of the 26 *EXPA* subfamily genes, 20 genes contained 2 introns (17 genes had the intron splicing mode of AB, 3 genes were in AF, CB and CF), 4 genes contained 1 intron (in B or C splicing mode), and 2 genes had 3 introns (in ACB or ACF splicing mode). All *EXPB* subfamily genes contained 3 introns (in ABF splicing mode excepting for *SpEXPB1* in ACB). *SpEXLA* and *SpEXLB* subfamily genes generally contained 4 introns (in ACBF splicing mode excepting for *SpEXLB1* in ACF). Obviously, the subfamily expansin genes of willow not only had similar exon-intron structure but also had the similar intron splicing mode, which also really showed subfamily characteristicly (Figure 4). Several other genes displayed a little variation, probably suffered from different evolution pressure.

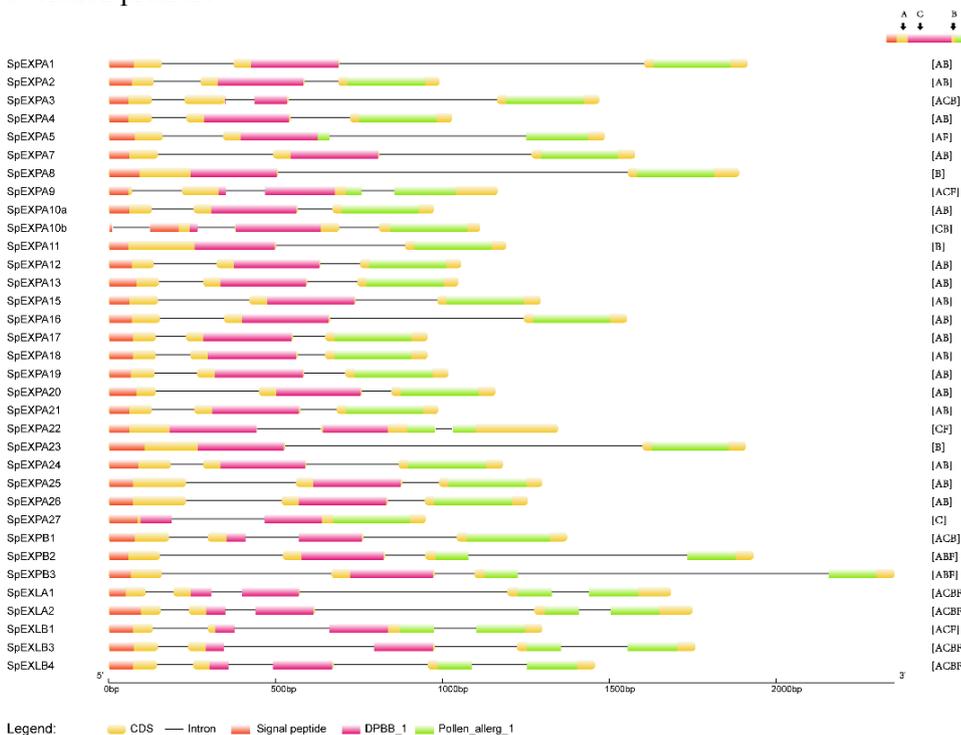


Figure 4 Exon-intron structures of expansin genes in *Salix purpurea*

1.4 Codon usage pattern of *Salix purpurea* expansin genes

A total of 34 willow expansin genes contain 8773 codons. The relative frequency of synonymous codon (RFSC) of some amino acids was really different (Table 2). For example, the RFSC values of the three synonymous codons AUU/AUC/AUA encoding Ile were 45.25%, 37.00% and 17.75%, respectively. Further, 5 high frequency codons (synonymous codons with frequency higher than 60% or 1.5 times higher than the average frequency (Lin et al., 2002) were determined in willow expansin genes, which were CCU (coding Pro), GCU (coding Ala), GAU (coding Asp), AGA and AGG (coding Arg). The less amounts of high frequency codons indicated the expansin genes expressed weakly in willow. The probable modification of the codons might be expected in future to advance the gene expression efficiency.

Table 2 RFSC value of expansin genes in *Salix purpurea*

Amino acid	Codon	Amount of codon	RFSC (%)	Amino acid	Codon	Amount of codon	RFSC (%)
Phe	UUU	190	41.58	Tyr	UAU	168	50.30
	UUC	267	58.42		UAC	166	49.70
Leu	UUA	51	8.70	TER	UAA	13	38.24
	UUG	113	19.28	UAG	15	44.12	
	CUU	105	17.92	His	CAU	86	52.44
	CUC	135	23.04		CAC	78	47.56
	CUA	58	9.90	Gln	CAA	178	49.04
	CUG	124	21.16		CAG	185	50.96
Ile	AUU	181	45.25	Asn	AAU	280	53.74
	AUC	148	37.00		AAC	241	46.26
	AUA	71	17.75	Lys	AAA	150	45.45
Met	AUG	199	100.00		AAG	180	54.55
	Val	GUU	184	33.39	Asp	GAU	208
GUC		126	22.87	GAC		112	35.00
GUA		67	12.16	Glu	GAA	93	52.84
GUG		174	31.58		GAG	83	47.16
Ser	UCU	169	22.41	Cys	UGU	140	46.51
	UCC	125	16.58		UGC	161	53.49
	UCA	133	17.64	TER	UGA	6	17.65
	UCG	47	6.23	Trp	UGG	250	100.00
Pro	CCU	162	42.19	Arg	CGU	44	11.17
	CCC	63	16.41		CGC	34	8.63
	CCA	120	31.25		CGA	32	8.12
	CCG	39	10.16		CGG	24	6.09
Thr	ACU	174	33.08	Ser	AGU	109	14.46
	ACC	149	28.33		AGC	171	22.68
	ACA	156	29.66	Arg	AGA	148	37.56
	ACG	47	8.94		AGG	112	28.43
Ala	GCU	290	39.67	Gly	GGU	252	25.25
	GCC	163	22.30		GGC	210	21.04
	GCA	224	30.64		GGA	335	33.57
	GCG	54	7.39		GGG	201	20.14

Note: The boxed and bolded codons are high frequency codons

The correlation analysis between GC12 and GC3 of codons (Figure 5A) showed that the expansin genes loci were evenly distributed at both sides of the standard curve, and the correlation was not significant (P=0.282) by two-tailed test. The regression coefficient was 0.089 1, indicating a significant difference of the base composition

was present at the first and second site of codon comparing with the third site. The mutation pressure (ϵ) accounts for 8.91% of the variation, and the natural selection pressure ($1-\epsilon$) contributes 91.09%. Thus, natural selection had a great influence on codon usage bias.

The ENC ratio of the willow expansin genes ranged from -0.03 to 0.24. Of them, the genes with -0.05~0.05 accounts for 30.00% and the genes with 0.10~0.20 accounts for 47.06%, indicating a great difference was present between the observed ENC value and the expected value. Correlation analysis between ENC value and GC3s displayed that most of the genes were distributed farther under the standard curve. It indicated that natural selection was the main influencing element for gene codon bias in evolution. A few genes distributed on the standard curve indicated that the mutation effect could partly cause the effect on codon bias (Figure 5B).

A Parity Rule 2 (PR2)-bias plot was used to evaluate the selection of pyrimidine (C and T) and purine (A and G) at the third base of codon. The vector dots were unevenly distributed in four regions, majorly at the down-left of the plan. It indicated that the frequency of T at the 3rd base of codons was higher than that of A, and the frequency of C was higher than that of G. In other words, the pyrimidine was more frequently used than purine. Natural selection had great influence on codon usage bias, and played an important role in evolution as some other factors (Figure 5C).

Based on the correspondence analysis of the relative synonymous codon usage (RSCU, the relative probability of a particular codon among the synonymous codons for the amino acid that could avoid the effect of amino acid composition on codon usage), the 1st axis contributed the largest variation of 13.43%, followed by the 2nd, 3rd, and 4th vector axes with the variation of 12.05%, 9.32% and 8.46%, respectively. A significant negative correlation was defined for the 1st axis with CAI, CBI and FOP with the correlation coefficient of -0.556, -0.617 and -0.576, respectively. No significant correlation was revealed for the 1st axis with ENC and GC3 with the correlation coefficient of 0.22, 0.088, respectively. Obviously, CAI, CBI and FOP had great influence on the codon bias of willow expansin genes, along with the other factors accordingly (Figure 5D).

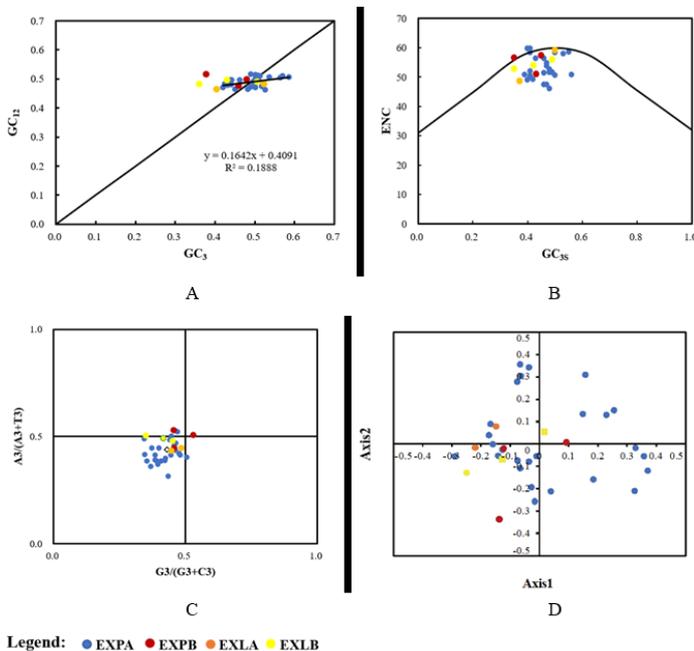


Figure 5 Codon usage analysis of expansin genes

Note: A: Neutrality plots (GC12: the average GC content at codons 1 and 2; GC3: The GC content at codon 3); B: ENC plot (ENC: The number of effective codons; GC3S: The GC content at the 3rd position of the synonymous codon); C: PR2-bias plots (\diamond : The average distribution position of all points); D: Correspondence analysis of the RSCU values of the chloroplast genes (ENC ratio = (expected ENC value - observed ENC value) / expected ENC value)

1.5 The synteny analysis of expansin genes in *Salix purpurea*

Based on the BLASTX analysis, 17 pairs of willow expansin genes were detected having a syntenic relationship (including 4 pairs of tandem duplications), even the cross-synteny among several expansin genes such as *SpEXPA1/SpEXPA7/SpEXPA15*, and *SpEXPA2/SpEXPA12/SpEXPA13* (Figure 6). They were mostly distributed on chromosome 4, 8, 10, 16 and 19 and displayed as paralogs or orthologs, where the duplicate fragments probably existed.

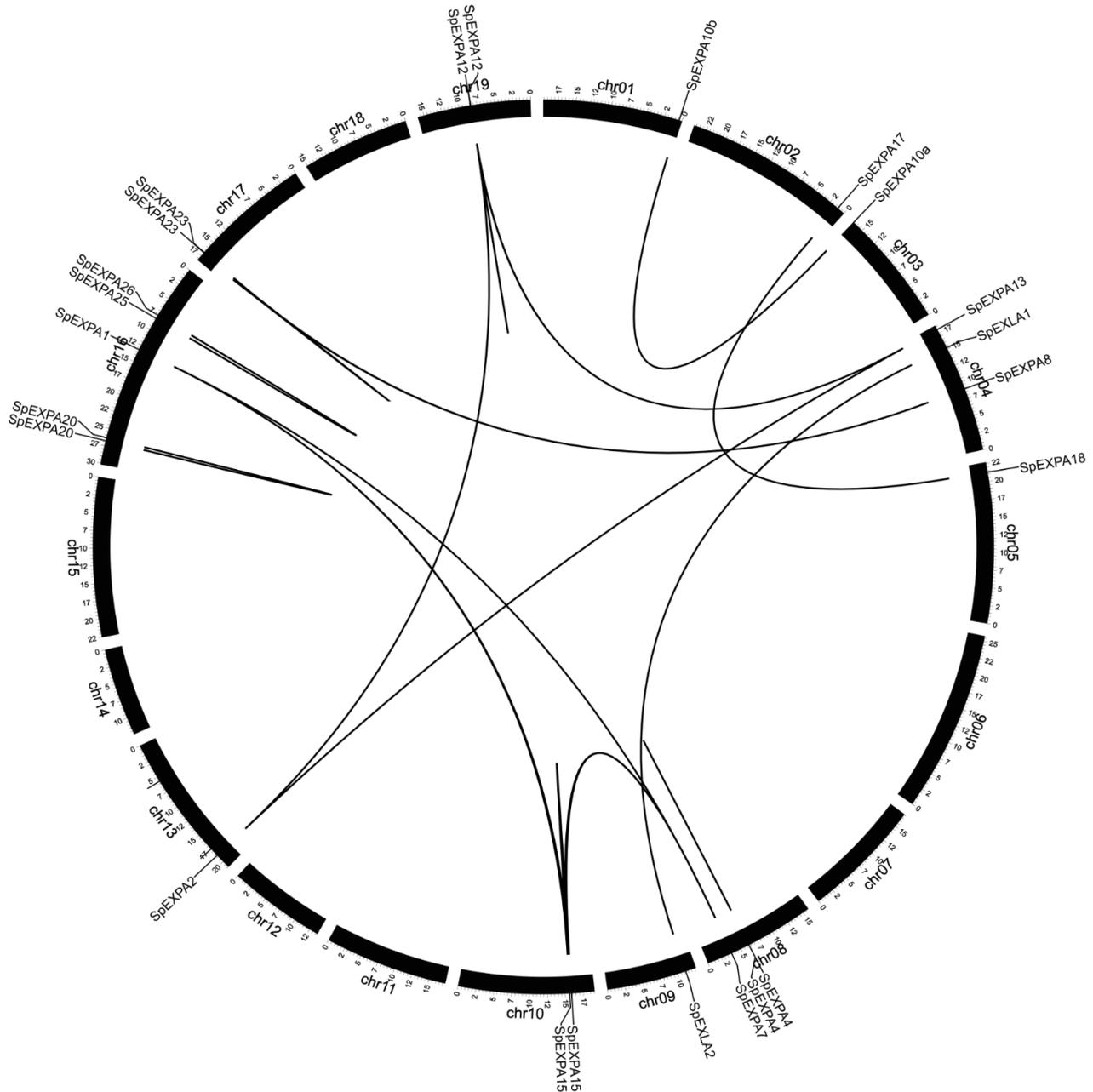


Figure 6 The synteny analysis of expansin genes in *Salix purpurea*

Furthermore, the Ka/Ks ratio of the paralogous pairs was calculated to explore the evolutionary model of the expansin gene pairs in *Salix purpurea*. The results showed that the Ka/Ks ratio of all the expansin gene pairs in *Salix purpurea* were less than 1, indicating that they had experienced a strong purification selection in evolution (Ka/Ks>1 indicates positive selection, Ka/Ks=1 indicates neutral selection, Ka/Ks<1 indicates purification selection). Based on the estimation of the fragment duplication events, the pairs of the expansin genes in *Salix*

purpurea might be divergent in 2.18 and 66.40 Mya (Table 3). Probably they had experienced a gradual evolution process via several times of gene replication and expansion.

Table 3 The Ka/Ks ratio between the paralogous expansin gene pairs in *Salix purpurea*

Paralogous pairs	Ka	Ks	Ka/Ks	Divergence time (Mya)
<i>SpEXPA1</i> & <i>SpEXPA7</i>	0.10	1.15	0.08	38.33
<i>SpEXPA1</i> & <i>SpEXPA15</i>	0.10	1.88	0.05	62.72
<i>SpEXPA7</i> & <i>SpEXPA15</i>	0.06	0.39	0.16	13.00
<i>SpEXPA2</i> & <i>SpEXPA12</i>	0.08	1.68	0.05	56.07
<i>SpEXPA2</i> & <i>SpEXPA13</i>	0.13	1.57	0.08	52.32
<i>SpEXPA12</i> & <i>SpEXPA13</i>	0.12	1.99	0.06	66.40
<i>SpEXPA8</i> & <i>SpEXPA23</i>	0.95	1.15	0.82	38.43
<i>SpEXPA10a</i> & <i>SpEXPA10b</i>	0.06	0.37	0.15	12.23
<i>SpEXPA17</i> & <i>SpEXPA18</i>	0.05	0.47	0.11	15.61
<i>SpEXPA25</i> & <i>SpEXPA26</i>	0.04	0.07	0.67	2.18
<i>SpEXPB2</i> & <i>SpEXPB3</i>	0.09	0.46	0.19	15.29
<i>SpEXPLA2</i> & <i>SpEXPLA1</i>	0.07	0.53	0.12	17.69
<i>SpEXPLB3</i> & <i>SpEXPLB4</i>	0.11	0.38	0.30	12.71

1.6 Physicochemical properties of expansin proteins in *Salix purpurea*

The molecular weight of the expansins in *Salix purpurea* was predicted in range of 23.61-31.70kD with an average value of 27.91. The most expansins showed slightly alkaline and had the pI value >7.0 excepting for SpEXPB1, SpEXLB3, and SpEXLB4, with an average of 8.66. The aliphatic index of the proteins ranged from 61.72 to 82.51, with an average value of 70.61. The relatively high aliphatic index ensured them function normally in various environment conditions, especially SpEXPA23 belonged to thermophilic protein (the aliphatic index was 82.51). The grand average of hydropathicity of willow expansins was from -0.403 to 0.051, indicating them as amphiphilic proteins (positive value means hydrophobicity while negative value indicates hydrophilicity). And, the instability index was between 22.93 and 46.40 with an average value of 34.51 (more than 40 was unstable). Thus, 85.29% of the expansins had the stable protein structure, which was necessary for their function in plant growth and resistance (Table 4).

2 Discussion

In this study, 34 expansin genes belong to four subfamilies were identified based on the sequenced genome of *Salix purpurea*. The genes in a subfamily had a conservative intron-exon structure and high sequence identity. Codon analysis showed the willow expansin genes having low codon usage bias, mainly resulted from the natural selection. And, codon composition also displayed characteristically by species and genus, In addition, the formation of willow expansin gene family might be in relation to several duplication events of the genome. All the information here partly revealed the evolution history and background of the willow genome. Similarly, Harikrishnan detected a large amount of duplication fragments in willow genome through RNA sequencing (Harikrishnan et al., 2015), with an average Ka/Ks (nonsynonymous substitution/synonymous substitution) of 0.23. Probably the willow genome had experienced strong purification selection process in evolution, which was consistent with our results based on a single gene family analysis here.

More studies on the expansin gene function had been carried out in several laboratories via gene expression (Castillo et al., 2018) and gene transformation experiment (Zenoni et al., 2011). The results really confirmed the important role of expansin genes in plant growth and development and environmental adaptation process from different perspectives. Bioinformatics analysis also provided the evidences, including the expansin genes in *Salix purpurea*, that the sequence identity of the subfamily genes was generally over 80%, and the proteins had three typical domains. The synteny analysis of the willow expansin genes showed their Ka/Ks value all less than 1, indicating that the willow expansin genes had suffered from a serious purification selection pressure in evolution and the gene with nonsynonymous mutation were really eliminated. Also, the structure stability and conservation

of the expansin genes would be necessary for their function process, as same as the performance of the indispensable housekeeping genes in plants.

Table 4 Physicochemical properties of expansin proteins in *Salix purpurea*

Protein name	Molecular weight (kD)	Theoretical pI	Aliphatic index	Grand average of hydropathicity	Instability index
SpEXPA1	28.46	9.72	65.95	-0.096	46.40
SpEXPA2	26.53	8.04	65.24	-0.065	37.79
SpEXPA3	23.61	9.69	66.45	-0.135	35.11
SpEXPA4	26.52	9.50	68.51	-0.129	41.12
SpEXPA5	25.67	8.75	63.61	-0.044	42.62
SpEXPA7	27.58	9.45	71.09	-0.029	30.24
SpEXPA8	30.53	9.12	73.54	-0.111	35.53
SpEXPA9	26.00	9.07	67.59	-0.073	39.08
SpEXPA10a	26.61	8.88	68.06	-0.125	22.93
SpEXPA10b	26.98	9.39	67.74	-0.163	27.88
SpEXPA11	28.68	9.15	68.10	-0.054	37.75
SpEXPA12	26.49	8.06	61.72	-0.142	31.33
SpEXPA13	27.67	8.94	71.53	-0.062	29.68
SpEXPA15	28.12	9.33	71.01	-0.057	35.80
SpEXPA16	28.08	9.62	71.27	-0.047	31.74
SpEXPA17	27.74	9.26	65.66	-0.024	28.62
SpEXPA18	27.65	9.17	64.49	-0.049	33.09
SpEXPA19	27.79	9.26	68.63	-0.123	29.39
SpEXPA20	28.15	10.13	67.20	-0.124	44.10
SpEXPA21	27.89	8.08	75.68	-0.143	39.41
SpEXPA22	27.59	8.37	74.12	-0.140	32.68
SpEXPA23	30.39	8.89	82.51	-0.072	35.74
SpEXPA24	29.14	9.54	63.90	0.051	34.21
SpEXPA25	31.70	9.44	68.60	-0.403	33.87
SpEXPA26	31.53	9.34	68.60	-0.390	30.16
SpEXPA27	24.74	9.07	76.55	-0.064	28.82
SpEXPB1	28.81	5.52	72.06	-0.059	39.95
SpEXPB2	28.67	8.64	79.24	-0.139	28.17
SpEXPB3	29.05	8.06	75.81	-0.224	30.26
SpEXLA1	28.27	8.60	76.49	-0.063	35.68
SpEXLA2	29.82	8.69	72.30	-0.016	30.76
SpEXLB1	26.59	7.52	75.25	-0.173	33.57
SpEXLB3	27.90	5.49	78.95	-0.179	41.26
SpEXLB4	28.06	4.59	73.16	-0.144	38.43
Average value	27.91	8.66	70.61	-0.112	34.51

The expansin genes obviously display species and genus characteristics. Based on the 11 sequenced plant genomes, more expansin genes were identified in monocots plants of rice (58) (Shi et al., 2014) and maize (88) (Zhang et al., 2014b), while less gene members were detected in dicots plants of willow (34), poplar (36) (Li et al., 2014), Arabidopsis (36) (Sampedro and Cosgrove, 2005), potato (36) (Chen et al., In 2019), grape (29) (Dalsanto et al., 2013), jujube (30) (Hou et al., 2019), apple (36) (Zhang et al., 2014a), soybean(65) (Zhu et al., 2014), and cucumber (35) (Hao et al., 2015). Possibly, the monocots are at the top of evolution, and had experienced several times of duplication events in evolution that might cause more expansin genes retained. And the enriched amount of expansin genes is consistent with the evolution of species and advanced performance of plants, which really determined the important status of expansin genes in plant genome. In addition, the frequency of synonymous codons usage in expansin genes also varied by species. The amount of high-frequency codons in monocots of maize and rice were 24 and 26, respectively, while 5, 6, 5 and 9 in dicots of willow, poplar, grape, and Arabidopsis

(Zhang et al., 2018b). Meanwhile, 4 of 5 high-frequency codons (CCU, GCU, AGA and AGG) were shared in *Salix purpurea* and *Populus trichocarpa*, 3 of them (CCU, GCU, AGA) had the RFSC values close to the threshold values of high-frequency codons in dicots of grape and Arabidopsis but significantly reduced in monocots of maize and rice. On the contrary, some high-frequency codons in monocots were used less in dicots. Some reports demonstrated that more high-frequency codons might be improvable on gene expression (Liu, et al., 2004). Therefore, codon usage bias of the expansin genes in monocots and dicots also pretends the gene expression variation even the status in genome.

Of the all expansin genes in plant genome, the proportion of *EXPB* subfamily members was small in several dicots species (ranging from 8.57% to 16.67%), while that in monocots was relatively high such as 32.76% in rice and 54.55% in maize (Li et al., 2016). Considering of the mechanism, the expansins could bind then break the covalent bonding between cellulose and hemicellulose in cell wall, and finally loosen cell wall components (Cosgrove, 2015). Some experiments reported that the cell wall components in monocots and dicots varied greatly. The hemicellulose in dicots mainly consists of xyloglucan, while the hemicellulose in monocots mainly consists of glucurono-arabinoxylan polysaccharide that was more likely to be the target of *EXPB* (Cosgrove, 2000). Considering of the evolution process, all *EXPB* genes in monocot of rice were involved in segmental or tandem replication but only a few *EXPB* genes in dicots of soybean and Arabidopsis were involved (Zhu et al., 2014). The difference in duplication and expansion in evolution could also lead to the divergency of expansin genes varied in species and genera (Kong et al., 2007). In terms of function, the *EXPB* genes were more involved in reproductive-related pollen tube elongation (Cosgrove et al., 1997) and fruit ripening (Brummell et al, 1999), while the *EXPA* genes were mostly involved in growth and development of plants. They might partly explained that the woody dicots plants need high proportion of *EXPA* genes for wood production, while poaceae monocots plants need high proportion of *EXPB* genes for seed harvest. It would be interested to be further discussed in future.

3 Materials and methods

3.1 Genes and genome sequences

The genomic sequence of *Salix purpurea* was downloaded from the phytozome database, and the expansin genes sequences of *Populus trichocarpa* was obtained from the EXPANSIN website.

3.2 Identification of the expansin genes in *Salix purpurea*

The nucleotide sequence of the expansin genes in *Populus trichocarpa* was used as a probe to search the sequenced genome library of *Salix purpurea*, and defined the candidate expansin genes with E-value 0.001 of BlastP alignment. The gene domain of Pollen_allerg_1 (PF03330) and DPBB_1 (PF01357) were verified by using the hidden Markov model of Pfam, NCBI conserved domain database (CDD) and SMART. The willow expansin genes were finally named according to the sequence identity with that in *Populus trichocarpa*. Based on the physical information of sequenced genome, each expansin gene was individually mapped on the chromosomes by the TBtools software. The Neighbor-Joining (NJ) method of MEGA7 was used to construct the phylogenetic tree based on the amino acid sequence of the expansin genes in *Salix purpurea* and *Populus trichocarpa*, and bootstrap was used to evaluate the phylogenetic tree with the repetition value of 1000.

3.3 Sequence analysis of expansin genes

The sequence identity between genes was conducted by the bioinformatics software DNAMAN v7.0. The box-plot map of the expansin genes in nucleotide and in amino acid identity was drawn by using Origin9. SignalP v4.1 was used to predict the signal peptide. ProtParam proteomics online tool was used to obtain the molecular weight, protein isoelectric point, aliphatic index and other related physicochemical parameters of expansins. The intron-exon structure map of the gene was drawn by online tool GSDS.

3.4 Codon usage bias analysis

The codon usage bias of expansin genes in *Salix purpurea* was analyzed by CodonW v1.4.4 and CHIPS and

CUSP programs of EMBOSS. The neutrality plot analysis was carried out with GC3 and GC12 as horizontal and longitudinal coordinates, two-dimensional scatter plot of ENC was drawn with GC3S as abscissa and ENC as ordinate. PR2-plot analysis was carried out with G3/(G3+C3) as abscissa and A3/(A3+T3) as ordinate. All the genes were plotted in a 59-dimensional vector space based on RSCU values. The correspondence analysis was conducted between one-axis and two-axis, and the correlation analysis was conducted by SPSS 20.0 software.

3.5 The synteny analysis of expansin genes

All mapped expansin genes of *Salix purpurea* were conducted for synteny analysis by using Circos. Two or more genes in a <200 kb fragment were defined as tandem duplications (Holub, 2001). The Ka/Ks (non-synonymous substitution/synonymous substitution of each site) ratio was calculated for each pair of the duplicated genes by using ClustalW alignment tool and KaKs_Calculator2.0 software. The divergence time in evolution was estimated using the formula $T=Ks/2r$ ($r=1.5 \times 10^{-8}$, the divergence rate of plant nuclear gene) (Koch et al., 2000).

Authors' contributions

Yang Ruixia and Liu Xinru designed and carried out the study and contributed equally to this work. Lan Baoliang, Wang Hao, and Liu Xiao participated the data analysis and manuscript writing. Xu Jichen is the corresponding author for the project and gave the guidance on experiment data analysis and manuscript writing.

Acknowledgments

The research was granted by College Student Research and Career-creation Program of Beijing (S201910022050) and Beijing Forestry University Undergraduate Training Program for Innovation and Entrepreneurship (X201910022080)

References

- Brummell D.A., Harpster M.H., and Dunsmuir P., 1999, Differential expression of expansin gene family members during growth and ripening of tomato fruit, *Plant Molecular Biology*, 39(1): 161-169
<https://doi.org/10.1023/A:1006130018931>
PMid:10080718
- Castillo F.M., Canales J., Claude A., and Calderini D.F., 2018, Expansin genes expression in growing ovaries and grains of sunflower are tissue-specific and associate with final grain weight, *BMC Plant Biology*, 18(4): 327
<https://doi.org/10.1186/s12870-018-1535-7>
PMid:30514222 PMCid:PMC6280438
- Chen Y., Han Y., Zhang M., Zhou S., Kong X., and Wang W., 2016, Overexpression of the wheat expansin gene TaEXPA2 improved seed production and drought tolerance in transgenic tobacco plants, *PLoS One*, 11(4): e0153494
<https://doi.org/10.1371/journal.pone.0153494>
PMid:27073898 PMCid:PMC4830583
- Chen Y., Zhang B., Li C., Lei C., Kong C., Yang Y., and Gong M., 2019, A comprehensive expression analysis of the expansin gene family in potato (*Solanum tuberosum*) discloses stress-responsive expansin-like B genes for drought and heat tolerances, *PLoS One*, 14(7): e0219837
<https://doi.org/10.1371/journal.pone.0219837>
PMid:31318935 PMCid:PMC6638956
- Cosgrove D.J., 2000, New genes and new biological roles for expansins, *Current Opinion in Plant Biology*, 3(1): 73-78
[https://doi.org/10.1016/S1369-5266\(99\)00039-4](https://doi.org/10.1016/S1369-5266(99)00039-4)
- Cosgrove D.J., 2015, Plant expansins: diversity and interactions with plant cell walls, *Current Opinion in Plant Biology*, 25: 162-172
<https://doi.org/10.1016/j.pbi.2015.05.014>
PMid:26057089 PMCid:PMC4532548
- Cosgrove D.J., Bedinger P., and Durachko D.M., 1997, Group I allergens of grass pollen as cell wall-loosening agents, *Proc. Natl. Acad. Sci. USA*, 94(12): 6559-6564
<https://doi.org/10.1073/pnas.94.12.6559>
PMid:9177257 PMCid:PMC21089
- Dal Santo S., Vannozzi A., Tornielli G.B., Fasoli M., Venturini L., Pezzotti M., and Zenoni S., 2013, Genome-Wide analysis of the expansin gene superfamily reveals grapevine-specific structural and functional characteristics, *PLoS One*, 8(4): e62206
<https://doi.org/10.1371/journal.pone.0062206>
PMid:23614035 PMCid:PMC3628503
- Hao X., Li X.Y., La G.X., Dai D.D., and Yang T.G., 2015, Identification and bioinformatics analysis of expansin gene family in cucumber, *Fenzi Zhiwu Yuzhong (Molecular Plant Breeding)*, 13(10): 2280-2289



- Harikrishnan S.L., Pucholt P., and Berlin S., 2015, Sequence and gene expression evolution of paralogous genes in willows, *Scientific Reports*, 5: 18662
<https://doi.org/10.1038/srep18662>
PMid:26689951 PMCID:PMC4687058
- Holub E.B., 2001, The arms race is ancient history in Arabidopsis, the wildflower, *Nature Reviews Genetics*, 2(7): 516
<https://doi.org/10.1038/35080508>
PMid:11433358
- Hou L., Zhang Z., Dou S., Zhang Y., Pang X., and Li Y., 2019, Genome-wide identification, characterization, and expression analysis of the expansin gene family in Chinese jujube (*Ziziphus jujuba* Mill.), *Planta*, 249(3): 815
<https://doi.org/10.1007/s00425-018-3020-9>
PMid:30411169
- Jiang F., Lopez A., Jeon S., De Freitas S.T., Yu Q., Wu Z., Labavitch J.M., Tian S., Powell A., and Mitcham E., 2019, Disassembly of the fruit cell wall by the ripening-associated polygalacturonase and expansin influences tomato cracking, *Horticulture Research*, 6: 17
<https://doi.org/10.1038/s41438-018-0105-3>
PMid:30729007 PMCID:PMC6355925
- Koch M.A., Haubold B., and Mitchell-Olds T., 2000, Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in Arabidopsis, Arabis, and related genera (Brassicaceae), *Molecular Biology and Evolution*, 17(10): 1483-1498
<https://doi.org/10.1093/oxfordjournals.molbev.a026248>
PMid:11018155
- Kong H., Landherr L.L., Frohlich M.W., Leebens-Mack J., Ma H., and dePamphilis C.W., 2007, Patterns of gene duplication in the plant SKP1 gene family in angiosperms: Evidence for multiple mechanisms of rapid gene birth, *Plant J.*, 50(5): 873-885
<https://doi.org/10.1111/j.1365-313X.2007.03097.x>
PMid:17470057
- Li H., Shi Y., Ding Y., and Xu J., 2014, Bioinformatics analysis of expansin gene family in poplar genome, *Beijing Linze Daxue Xuebao (Journal of Beijing Forestry University)*, 36(2): 59-67
- Li N., Pu, Y., Gong Y., Yu Y., and Ding H., 2016, Genomic location and expression analysis of expansin gene family reveals the evolutionary and functional significance in *Triticum aestivum*, *Genes & Genomics*, 38(5): 1021-1030
<https://doi.org/10.1007/s13258-016-0446-y>
- Lin T., Ni Z., Shen M., and Chen L., 2002, High-frequency codon analysis and its application in codon analysis of tobacco, *Xiamen Daxue Xuebao (Journal of Xiamen University (Natural Science))*, 41(5): 551-554
- Liu Q., Feng Y., and Dong H., 2004, Comparative studies on synonymous codon usage bias in twenty species, *Xibei Nonglin Keji Daxue Xuebao (Journal of Northwest A & F University (Natural Science Edition))*, 32(7): 67-71
- Lu P., Kang M., Jiang X., Dai F., Gao J., and Zhang C., 2013, RhEXPA4, a rose expansin gene, modulates leaf growth and confers drought and salt tolerance to Arabidopsis, *Planta*, 237(6): 1547-1559
<https://doi.org/10.1007/s00425-013-1867-3>
PMid:23503758
- Noh S.A., Lee H.S., Kim Y.S., Paek K.H., Shin J.S., and Bae J.M., 2013, Down-regulation of the IbEXP1 gene enhanced storage root development in sweetpotato, *J. Exp. Bot.*, 64(1): 129-142
<https://doi.org/10.1093/jxb/ers236>
PMid:22945944 PMCID:PMC3528024
- Sampedro J., and Cosgrove D.J., 2005, The expansin superfamily, *Genome Biol.*, 6(12): 242
<https://doi.org/10.1186/gb-2005-6-12-242>
PMid:16356276 PMCID:PMC1414085
- Sampedro J., Carey R.E., and Cosgrove D.J., 2006, Genome histories clarify evolution of the expansin superfamily: new insights from the poplar genome and pine ESTs, *Journal of Plant Research*, 119(8): 11-21
<https://doi.org/10.1007/s10265-005-0253-z>
PMid:16411016
- Shi Y., Xu X., Li H., Xu Q., and Xu J., 2014, Bioinformatics analysis of the expansin gene family in rice, *Yichuan (Hereditas)*, 36(8): 809-820
- Yan A., Wu M., Yan L., Hu R., Ali I., and Gan Y., 2014, AtEXP2 is involved in seed germination and abiotic stress response in Arabidopsis, *PLoS One*, 9(1): e85208
<https://doi.org/10.1371/journal.pone.0085208>
PMid:24404203 PMCID:PMC3880340
- Zenoni S., Fasoli M., Tornielli G.B., Dal Santo S., Sanson A., de Groot P., Sordo S., Citterio S., Monti F., and Pezzotti M., 2011, Overexpression of PhEXPA1 increases cell size, modifies cell wall polymer composition and affects the timing of axillary meristem development in *Petunia hybrida*, *New Phytol.*, 191(3): 662-77
<https://doi.org/10.1111/j.1469-8137.2011.03726.x>
PMid:21534969



- Zhang H., Ding Y., Zhi J., Li X., Liu H., and Xu J., 2018a, Over-expression of the poplar expansin gene, PtoEXPA12, in tobacco plants enhanced cadmium accumulation, *Int. J. Biol. Macromol.*, 116: 676-682
<https://doi.org/10.1016/j.ijbiomac.2018.05.053>
PMid:29758311
- Zhang H., Li J., Wang R., Zhi J., and Yin P., 2018b, Comparative analysis of expansin gene codon usage patterns among eight plant species, *J. Biomol. Struct. Dyn.*, 37(4): 1-19
<https://doi.org/10.1080/07391102.2018.1442746>
PMid:29480091
- Zhang S., Xu R., Gao Z., Chen C., Jiang Z., and Shu H., 2014a, A genome-wide analysis of the expansin genes in *Malus domestica*, *Molecular Genetics and Genomics*, 289(2): 225-236
<https://doi.org/10.1007/s00438-013-0796-y>
PMid:24378555
- Zhang W., Yan H., Chen W., Liu J., Jiang C., Jiang H., Zhu S., and Cheng B., 2014b, Genome-wide identification and characterization of maize expansin genes expressed in endosperm, *Molecular Genetics and Genomics*, 289(6): 1061-1074
<https://doi.org/10.1007/s00438-014-0867-8>
PMid:25213600
- Zhu Y., Wu N., Song W., Yin G., Qin Y., Yan Y., and Hu Y., 2014, Soybean (*Glycine max*) expansin gene superfamily origins: segmental and tandem duplication events followed by divergent selection among subfamilies, *BMC Plant Biology*, 14(1): 93
<https://doi.org/10.1186/1471-2229-14-93>
PMid:24720629 PMCid:PMC4021193