

## Article

# Evolutionary history of the GH3 family based on 48 sequenced plants and identification of jasmonic acid-relate GH3 genes in *Solanum tuberosum*

Chao Zhang<sup>1†</sup>, Leilei Zhang<sup>2†</sup>, Dongdong Wang<sup>1</sup>, Haoli Ma<sup>1</sup>, Bailin Liu<sup>1</sup>, Zheng Shi<sup>1</sup>, Xiaohui Ma<sup>1</sup>, Yue Chen<sup>1\*</sup>, Qin Chen<sup>1\*</sup>

<sup>1</sup> State Key Laboratory of Crop Stress Biology for Arid Areas, College of Agronomy, Northwest A&F University, Yangling, Shaanxi, People's Republic of China, 712100; zhangchao520@nwfau.edu.cn (C. Zhang); dongdong-1025@hotmail.com (D. Wang); mahaoli@nwsuaf.edu.cn (H. Ma); liubl@nwfau.edu.cn (B. Liu); summer\_shi@nwfau.edu.cn (Z. Shi); maxiaohui1216@nwfau.edu.cn (X. Ma)

<sup>2</sup> College of Agronomy, Liaocheng University, Liaocheng, Shandong People's Republic of China, 252059; zhangleilei@lcu.edu.cn (L. Zhang)

\* Correspondence: xnchenyue@nwfau.edu.cn (Y. Chen); chenpeter2289@nwfau.edu.cn (Q. Chen)

**Abstract:** Glycoside Hydrolase 3 (GH3) is a phytohormone-responsive family of genes that has been found in many plant species. It is implicated in the biological activity of indolacetic (IAA) and jasmonic acids (JA), and also affects plant growth and developmental processes and some stresses. In this study, GH3 genes were identified in 48 plants, which belong to algae, moss, fern, gymnosperm and angiosperm. No GH3 representative gene has been found in algae, and our research identified 4 genes in mosses, 19 in ferns, 7 in gymnosperms, and numerous in Angiosperms. The results showed that GH3 genes mainly occur in seed plants. Phylogenetic analysis of all GH3 genes showed three separate clades. Group I was related to JA adenylation, group II was related to IAA adenylation, and group III was separated from group II but the function was not clear. The structure of GH3 protein indicated highly conserved sequence in the plant kingdom. The analysis of JA-adenylation related to gene expression of GH3 in potato (*Solanum tuberosum*) showed that *StGH3.12* highly responded to Methyl Jasmonate (MeJA) treatment. Expression levels of *StGH3.1*, *StGH3.11*, and *StGH3.12* were high in flower and *StGH3.11* expression was also high in stolon. Our research revealed the evolution of the GH3 family, which is useful for studying the precise function about JA-adenylation GH3 genes in *S. tuberosum* under development and biotic stresses.

**Keywords:** GH3 family; sequencing plants; potato; jasmonic acids; tissues; biotic

## 1. Introduction

GH3 proteins are widespread in plant species. Differential hybridization screening was used to first isolate GH3 as an auxin-induced cDNA clone from etiolated soybean hypocotyls [1]. Subsequently, GH3 genes were identified from tobacco [2], *Arabidopsis* [3], rice [4] and other plants [5-9]. These proteins adjust the phytohormone signaling pathway. A two-step mechanism showed that the GH3 family conjugates amino acids to diverse acyl acid substrates. The first step indicates adenylation of an acyl acid hormone and release of pyrophosphate. In the second transfer step, the amine group of an amino acid nucleophilically displaces AMP, yielding the conjugated acyl acid [10]. GH3 proteins regulate the levels of phytohormones, including JA and IAA [11]. Group I GH3 genes catalyze additional amino acids to phytohormone JA-Ile and Group II GH3 genes catalyze additional amino acids to phytohormone IAA.

In published phylogenetic trees of *Arabidopsis*, GH3 proteins can be divided into three major groups on account of sequence similarity and acyl substrate specificity [12]. Group I proteins use JA as a substrate, which promotes formation of JA-Ile and boosts JA-mediated responses [9]; *AtGH3.11* (*JAR1*) was identified as a pre-receptor in jasmonate-responses [11]. Group II proteins can conjugate IAA to the free IAA and store phytohormone with an inactive form [13]. *AtGH3.12/PBS3*, belonging to Group III, uses the benzoates, 4-hydroxybenzoate (4HBA), and is better than Salicylic acid (SA) as the substrate [14]. Until now, the functions of GH3 genes have been reported primarily in *Arabidopsis*. *AtGH3.2/Ydk1-D* can promote short primary roots and reduce lateral root number [15]; *AtGH3.6/df1* restrains shoot elongation and lateral root formation [16]; *AtGH3.9* seems to be specifically active in roots [17].

IAA as the first identified phytohormone which is responsible for growth and development in flowering plants by influencing cell division, expansion, and differentiation [18, 19]. And group II genes from GH3 family are related in IAA. Similarly, JA as one of the major plant hormones involved in regulating the balance between growth and defense [20]. A huge number of research come to a decision is JA participates in biotic stresses in plants indeed. Thus, group I genes is not only take part in growth but also have key role in biotic stress based on JA biological function in plants.

With the continuous improvement of gene sequencing technology, genetic data have been generated for more and more species. The GH3 gene family could be identified in several plants, but no one studied a detailedly and completely evolutionary history. Even though Terol (2006) conducted a genome-wide analysis of GH3 in rice and its evolutionary history from 20 plant species were identified based on EST analysis [21]. But sequencing complete genomes are available for at least 48 plants until now, and covered stages from lower to higher plants. Thus, the evolution of GH3 genes may be studied more accurately based on these sequenced plant species. Here, we analyzed the GH3 gene family in 48 species which included algae, moss, fern, gymnosperm, and angiosperms using bioinformatics methods. On the other hand, we analyzed the gene characteristics, structures, phylogenetic relationships, gene ontology (GO) annotations, and expression patterns of GH3 genes in *Solanum tuberosum*. Thus, the results of this study not only reveal the evolution of the GH3 family but also provide some useful information for further research in potato, especially the biotic stress based on JA pathway.

## 2. Results and discussion

### 2.1 The GH3 family was identified in 48 plant species

To study the evolution of the GH3 family, we surveyed the sequences of GH3 family members in sequencing plant species. We used the GH3 full-length protein sequence of *Arabidopsis thaliana* [12] as a query to search the available genome sequences from 48 plant species including algae, mosses, ferns, gymnosperms, and angiosperms. We found 579 proteins in 48 plants. The number of GH3 proteins ranged from 0 to 38 (Table 1). According to the evolutionary relationships in plant, no GH3 proteins were acquired in algae. Only two proteins were identified in each moss, *Marchantia polymorpha* L. and *Physcomitrella patens*. Fern species *Selaginella moellendorffii* contained 19 GH3 proteins which was a large group compared with some other plants. We could not find genome information for the gymnosperm *Picea abies* in NCBI-Genome but it was reported by [22]; seven GH3 proteins were identified in *P. abies*. Numerous angiosperms have been sequenced, making it easier to determine how many GH3 genes had been acquired. The most GH3 proteins appeared from *Brassica rapa*, but it did not have the largest genome or total protein counts. Different plants had different GH3 proteins, but these numbers were not correlated with genome length or protein counts. According to the general trend, the GH3 protein appeared earliest in moss, and more proteins were identified with the evolution of plants. We found that the number of GH3 family genes apparently increased with the emergence of the seed plants (ferns, gymnosperms, and angiosperms).

Table 1 GH3 gene identified from 48 sequenced plants.

Lineage	Organism	Genome total length (Mb) <sup>1</sup>	Protein counts <sup>1</sup>	Numbers of <i>GH3</i> protein	Tandem duplication (pairs) <sup>2</sup>	Segmental duplication (pairs) <sup>3</sup>	References
Algae							
	<i>Chlamydomonas reinhardtii</i>	120.405	14488	0	0	0	This study
	<i>Volvox carteri</i>	137.684	14436	0	0	0	This study
Mosses							
	<i>Marchantia polymorpha</i> L.	205.718	17956	2	0	0	This study
	<i>Physcomitrella patens</i>	477.948	35934	2	0	0	This study
Ferns							
	<i>Selaginella moellendorffii</i>	212.502	34746	19	2	1	This study
Gymnosperms							
	<i>Picea abies</i>	28354	about 20000	7	0	0	This study
Angiosperms							
	<i>Amborellaceae</i> <i>Amborella trichopoda</i> <sup>#</sup>	706.495	31494	6	0	0	This study
Eudicots							
	<i>Chenopodiaceae</i> <i>Beta vulgaris</i>	482.053	32874	5	0	1	This study
	<i>Brassicaceae</i> <i>Arabidopsis lyrata</i>	204.898	39161	17	10	0	This study
	<i>Arabidopsis thaliana</i>	116.846	39593	19	11	2	Jain[4]
	<i>Brassica oleracea</i>	501.692	56687	35	10	13	This study
	<i>Brassica rapa</i>	284.129	52553	38	8	13	This study
	<i>Capsella rubella</i>	133.064	28713	20	12	3	This study

<i>Solanaceae</i>	<i>Capsicum annuum</i>	2935.88	40627	7	0	1	This study
	<i>Solanum lycopersicum</i>	760.067	36010	17	6	4	This study
	<i>Solanum tuberosum</i>	705.934	37966	13	3	4	This study
<i>Leguminosae sp.</i>	<i>Cajanus cajan</i>	620.626	38965	12	0	0	This study
	<i>Cicer arietinum</i>	520.885	33107	7	0	1	This study
	<i>Glycine max</i>	953.339	71525	24	0	36	This study
	<i>Lotus japonicus</i>	394.455	N/A	7	0	0	This study
	<i>Medicago truncatula</i>	412.924	57661	17	7	2	Yang[6]
<i>Rosids</i>	<i>Phaseolus vulgaris</i>	535.413	32720	12	0	7	This study
	<i>Fragaria vesca</i>	214.373	31387	8	0	0	This study
	<i>Malus domestica</i>	1288.87	60549	15	1	4	Yuan[5]
	<i>Prunus mume</i>	234.03	29705	7	0	1	This study
	<i>Prunus persica</i>	212.767	32595	7	0	1	This study
<i>Cucurbitaceae</i>	<i>Pyrus bretschneideri</i>	508.551	47086	15	0	4	This study
	<i>Citrullus lanatus</i>	321.047	N/A	8	1	0	This study
	<i>Cucumis sativus</i>	323.986	25668	10	1	0	This study
<i>Actinidiaceae</i>	<i>Actinidia chinensis</i>	604.217	N/A	13	0	3	This study
<i>Caricaceae</i>	<i>Carica papaya</i>	370.419	26103	6	0	0	This study
<i>Euphorbiaceae</i>	<i>Ricinus communis</i>	350.622	27998	7	0	0	This study
<i>Lentibulariaceae</i>	<i>Utricularia gibba</i>	100.689	N/A	6	0	1	This study
<i>Malvaceae</i>	<i>Theobroma cacao</i>	335.437	37520	19	8	3	This study
	<i>Gossypium raimondii</i>	767.667	59057	32	12	11	This study
	<i>Eucalyptus grandis</i>	691.43	52554	9	0	2	This study
<i>Nelumbonaceae</i>	<i>Nelumbo nucifera</i>	797.494	38191	11	0	4	This study
<i>Rutaceae</i>	<i>Citrus sinensis</i>	323.528	43683	9	1	0	This study
<i>Salicaceae</i>	<i>Populus trichocarpa</i>	417.287	45942	13	0	8	This study

Monocots	<i>Vitaceae</i>	<i>Vitis vinifera</i>	486.197	41208	9	0	1	Bottcher[23]
	<i>Poaceae</i>	<i>Brachypodium distachyon</i>	218.345	33944	9	0	1	This study
		<i>Hordeum vulgare</i>	1825.17	N/A	5	0	0	This study
		<i>Oryza sativa</i>	374.423	36376	13	0	1	Jain[4]
		<i>Triticum urartu</i>	3747.05	24169	9	0	0	This study
		<i>Zea mays</i>	2145.45	58411	13	1	3	Feng[24]
	<i>Arecaceae</i>	<i>Elaeis guineensis</i>	1017.1	41887	9	0	2	This study
	<i>Musaceae</i>	<i>Musa acuminata</i>	472.231	47707	18	1	18	This study
	<i>Orchidaceae</i>	<i>Phalaenopsis equestris</i>	1064.2	29894	6	0	0	This study
Total					572			

N/A indicated the data cannot found.

<sup>1</sup>The data was came from NCBI ([www.ncbi.nlm.nih.gov/genome/](http://www.ncbi.nlm.nih.gov/genome/)).

<sup>2</sup>JGI (<https://phytozome.jgi.doe.gov/pz/portal.html>) contained hole information.

<sup>3</sup>The related information obtained from PGDD (<http://chibba.agtec.uga.edu/duplication/>).

<sup>#</sup>*Amborella trichopoda* was known as the sister group of the remaining flowering plant in molecular phylogenetic analyses [25].

## 2.2 Analysis of the phylogenetic tree and duplication patterns

Phylogenetic tree of the GH3 members in 48 plants was constructed based on the similarity of protein sequences (Figure 1). This tree contained three clades and consistent with previous studies [4, 5, 26]. The functional groups, group I and group II, which are related to JA and IAA, respectively, and these parts of genes were present in all major plant lineages, including the mosses. Three genes in mosses pertained to group I which contained the most ancient genes. GH3 genes might have played important roles in environmental adaptation in mosses. This gene group was present in all seed plants. It could be inferred that the function of group I was indispensable in all plants. One GH3 protein belonged to group II in mosses, but more group II genes appeared in seed plants; suggesting that group II had a key role in plants evolution. Group III genes were only present in eudicots. From the phylogenetic tree analysis, group III was closer to group II and belonged to one of the branches. Moreover, group II and III had the same plant lineages, and evolved from the moss *Marchantia polymorpha* L. (*Mapoly0053s0073.1.p*). Group III members were in eudicots including all brassicaceae (*Arabidopsis thaliana*, *Brassica oleracea*, *Brassica rapa*, *Capsella rubella*, and *Arabidopsis lyrata*), *Gossypium raimondii*, and *Theobroma cacao*; this result was consistent with Singh [27]. Therefore, we speculated that group III was the latest evolution, and Group III proteins were suggested to play a role in response to (a)biotic stress by Okrent [11].

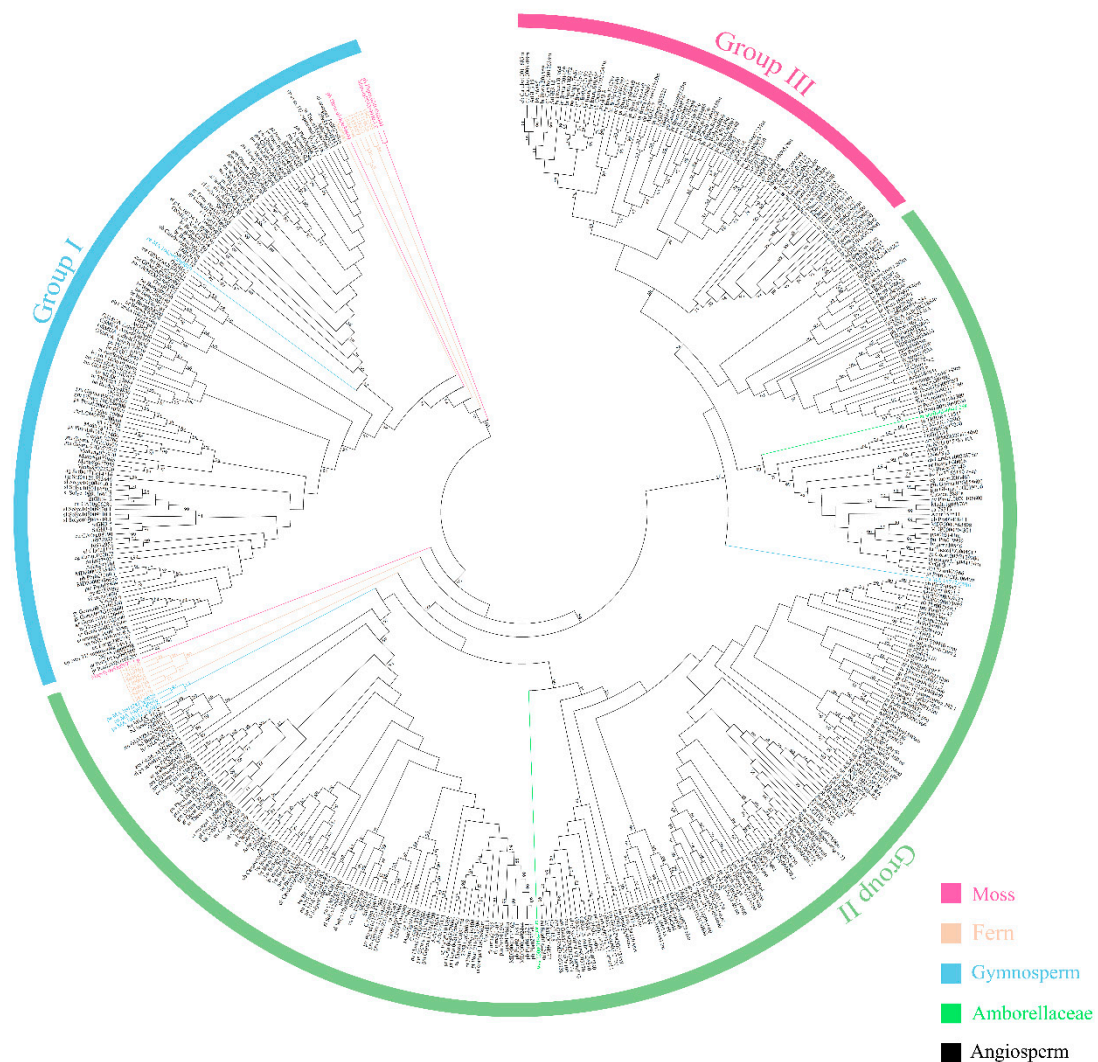




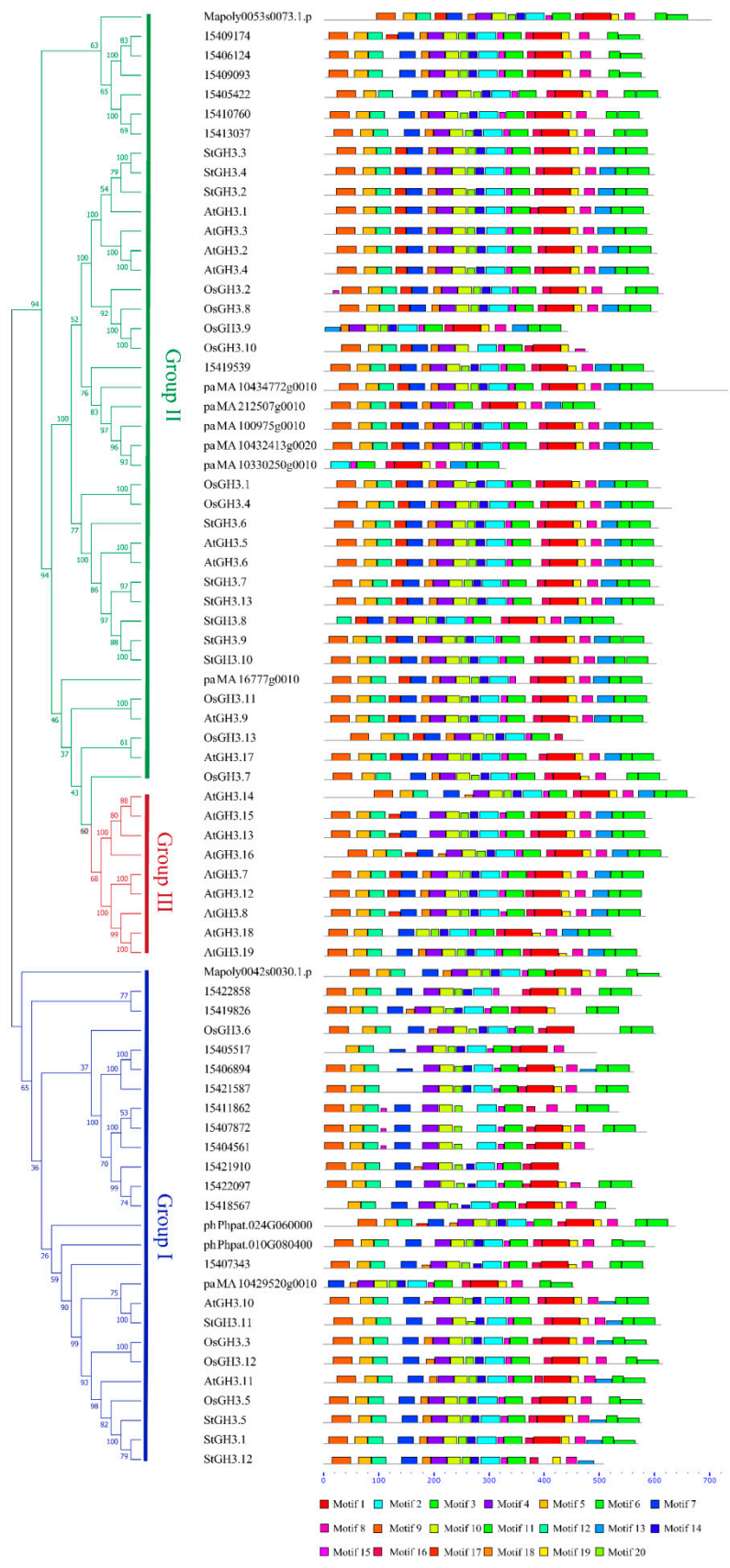
Figure 1. Unrooted Neighbor-Joining tree constructed with 48 plants GH3 proteins. Different colors in phylogenetic tree represent the proteins belong to different lineages. The distinction of subfamily is represented by 3 different color curves.

Group I and II GH3 genes have the same ancient origin, and the evolutionary explosions of these groups were caused by many duplication events. Early duplication events were identified in group I and II of fern (Figure 1).

For further GH3 evolutionary relationships, prediction of the molecular evolution rates could clarify the gene evolution process. Selective pressure was estimated by calculating the ratio of the nonsynonymous substitution rate to the synonymous substitution rate (Ka/Ks value), allowing us to analyze the molecular evolutionary rates (Table S1). Ka/Ks values was calculated for each duplication event, and almost every gene pair evolved at a Ka/Ks value lower than one except *ta\_Thecc1EG031553/ta\_Thecc1EG031554* gene pair in *Theobroma cacao* (the Ka/Ks value is 1.638). This gene pair experienced positive selection pressure in evolution. The other gene pairs indicated that the gene was subject to purification selection. Compared with each stages duplications, the angiosperms had many duplication events. We found out the duplication event in fern firstly. That Ka/Ks Values were all less than 1, and two kinds of duplication had similar value (0.318 to 0.388) in *Selaginella moellendorffii*. The numbers of GH3 gene duplications in dicots were more than in monocots. Interestingly, a plenty of tandem and segmental duplications occurred in *Brassicaceae*. The most segmental duplications were identified in *Glycine max* and there were no tandem duplications. The number of GH3 proteins and duplication events suggested that the reason for the increase in GH3 proteins was due to numerous duplication events in angiosperms. Plants evolve with changes in the environment and positive selection promoted the exchange of gene function to survive. Therefore, the selective patterns can partly explain the evolutionary patterns of the genes.

### 2.3 Structural analysis of GH3 Proteins

Seven species from different stages from moss to angiosperm were selected as the candidates to analyze their structures. The prediction of motifs was an essential method of protein analysis, and the motifs of GH3 proteins were identified using the MEME program. Twenty motifs were searched in these proteins (Figure 2). The GH3 protein family members have highly conserved motifs. The vast majority of proteins contained about 20 motifs, and they were in the same order. Compared with these proteins showed that moss and fern totally had 17-19 motifs and it was not existed the 17th motif in it. *PaGH3* protein (*pa\_MA\_10330250g0010*) only had 10 motifs, which was the lowest motifs in all proteins. Correspondingly, it also had the shortest length of protein. Be different with other gene family, such as WRKY [28], the motifs in subgroup proteins had no significantly differences.



**Figure 2.** The constitution of gene motifs with 7 plants. The 7 plants are from different lineage, respectively. The name of GH3 proteins on the left and construct their phylogenetic tree. The different-colors boxes on the right are indicated the different motif.



Because the GH3 proteins combined amino acids to different acyl acid substrates, they must have had the corresponding active site. Predicted protein structure could show the related structures. We took 10 proteins as candidates to analyze the information from research by Westfall [9]. These 10 proteins belonged to four different evolution stages and contained  $\alpha 5$ ,  $\alpha 6$ ,  $\beta 8-9$  and P-loop, but the sequences had some differences in diverse group (Figure 3). Alignment of proteins from each group showed they had same sequences. Also, the highly conserved sequences, Lys<sup>428</sup>/Lys<sup>435</sup> were in each protein, which are proposed to interact with amino acid substrates in group I and Lys<sup>146</sup> accepted acidic amino acids, whereas Ser<sup>151</sup> was specifically conjugated to isoleucine in group II.

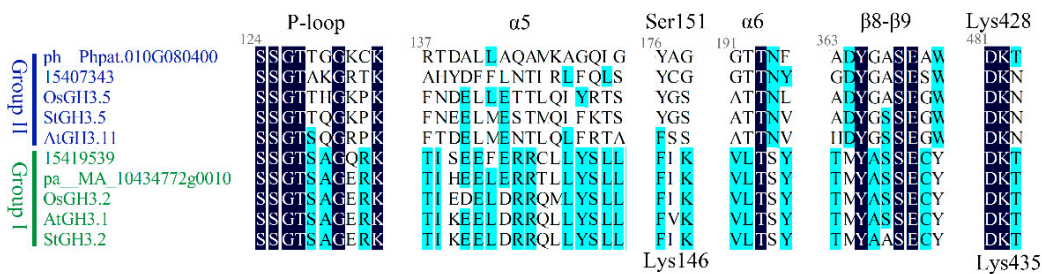


Figure 3. Predictions of protein structure with 10 GH3 proteins. The selected proteins are belonged to group I and II, which had definitely function. The grey numbers stand for the location of amino acids. The black annotates symbolize the secondary structure of protein.

2.4 Gene ontology annotation and RNA-seq data analysis of *StGH3* proteins

To better understand the biological processes of GH3 proteins, we selected potato GH3 proteins for GO analysis using the NCBI database. Three *StGH3* proteins (*StGH3.1*, *StGH3.5* and *StGH3.12*) participated in multiple signal transductions (signal transduction, response to stress, small molecule metabolic process, and immune system process). Moreover, these three proteins had predicted ligase activity. *StGH3.5* participated in enzyme binding and nucleotidyl transferase activity. The prediction of genes cellular component showed that *StGH3-5* participated in the component of vacuole (Figure 4A). Compared with the databases in the KEGG website, *StGH3* group I genes (except *StGH3.11*) participated in the JA pathway (Figure 4B). Thus, these 3 proteins had important roles in JA adenylation.

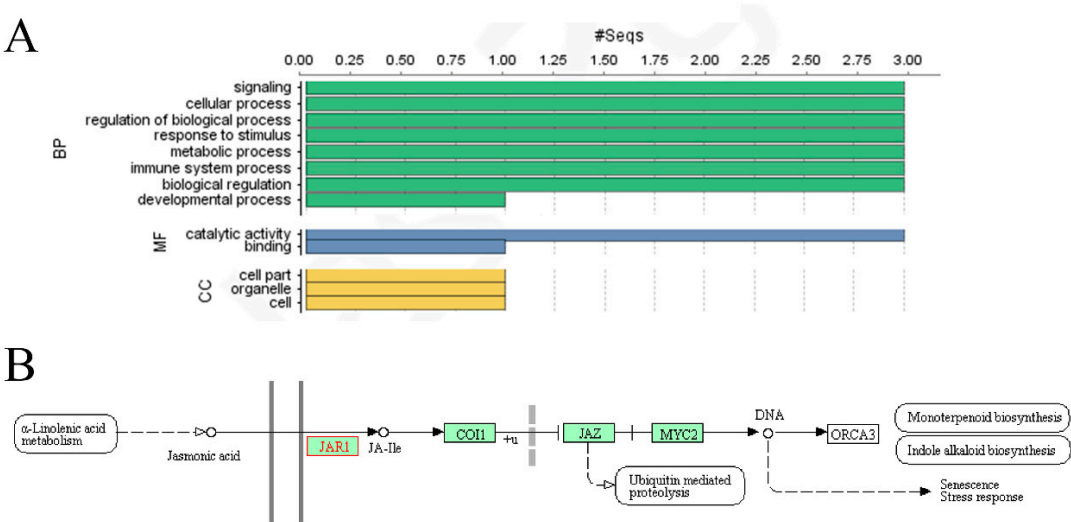


Figure 4. The information of gene ontology annotation and KEEG pathway. (A) *StGH3* proteins GO annotation. BP, MF and CC indicate Biological Process, Molecular Function and Cellular Component, respectively. (B)The

GH3 genes related KEEG pathway. *StGH3.1*, *StGH3.5* and *StGH3.12* are predicted as a part of JA pathway. The red word in red box is where the genes work.

To analyze the expression patterns of *StGH3* genes in various treatments, (Table S2) was downloaded from PGSC. We processed the RNA-seq database and generated a heatmap (Figure 5). Ten different treatments were part of this analysis. All *StGH3* genes were down regulated under BAP (6-Benzylaminopurine) treatment and slightly enhanced or reduced under *Phytophthora infestans* treatment. *StGH3.2* was up regulated under seven stressors and three treatments resulted in higher expression. *StGH3.10* was strongly up regulated and *StGH3.13* was down regulated to the point of no signal under heat treatment. Combined GO annotation and RNA-seq data analysis showed *StGH3.1*, *StGH3.5*, and *StGH3.12* participated in the stress regulation. These genes were all up regulated under heat and biotic stresses. The change of *StGH3* genes expression regulated by JA has not been reported, and these results should be analyzed following treatment of JA or MeJA.

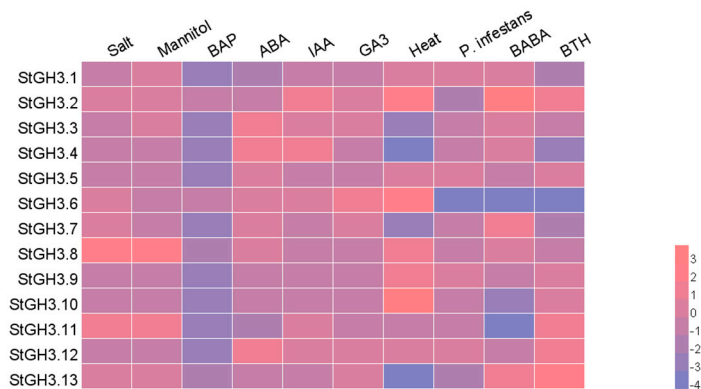


Figure 5. The heatmap with RNA-seq database under different biotic, abiotic and hormones treatments. The RNA-seq database is processed by log2. And each stress or hormonal data were compared with control.

2.5 Expression analysis of *StGH3* group I genes

To demonstrate the function of GH3 proteins in plant tissues and under JA or MeJA treatment, qRT-PCR analysis was used to reveal the expression. We mainly focused on the group III genes which were related to the JA pathway through prediction. Some researchers used MeJA to treat the potato tuber slice and tissue culture seedling, and results showed that JA and MeJA have an induced dilation effect on potato tuber cells and had a strong tuber inducing activity; also treating a single stalk of potato could promote tuber formation [29, 30] Thus, qRT-PCR was used to analyze the genes expression.

The tissue culture seedling was treated by 10<sup>-5</sup> M MeJA [31, 32] and collected materials treated in 4 different time periods. The *StGH3* group I genes were studied the expression changing in these treatments (Figure 6A, Table S3). These four genes were all up-regulated in different time period. Compared with untreated sample, *StGH3.1*, *StGH3.5*, and *StGH3.12* were up regulated 6 hours later; the expression level was gradually decreased from 6 hours to 24 hours. The highest change in expression occurred in *StGH3.12*, at 8-fold higher than the control after 6 hours of treatment. Overall, the response time of genes to MeJA treatment was about 6 hours. The most responsive gene was *StGH3.12* and other genes typically had no obvious response, especially *StGH3.11*, which was down regulated under MeJA treatment. But if a gene was down regulated in 1 hour, it might be affected by the MeJA concentration because the jasmonic acid compound was poorly soluble in MS liquid medium.

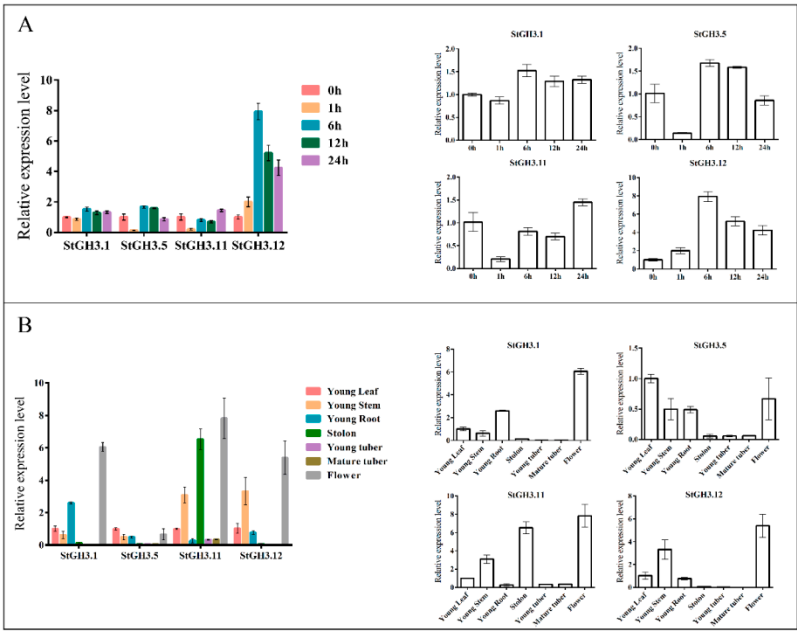


Figure 6. The JA-adenylation *StGH3* genes expression in different tissues and the expression change under MeJA treatment in different time. (A) The expression change of MeJA treatment in different times. (B) The expression in different tissues.

In addition, we studied the expression of these four genes in seven different tissues to verify whether there was a specific expression in a potato tissue (Figure 6B, Table S3). We selected the young leaf as the standard to process the qRT-PCR data by the  $2^{-\Delta\Delta C_t}$  method. Three genes had high expression in flower, only *StGH3.5* had lower expression in each tissue. The expression level of *StGH3.11* was also highly expressed in stolon and the other genes only had low expression. None of these genes had high expression in young tubers or mature tubers. Combined with results of research by Koda and others [29, 30], the related with *StGH3* genes and potato tuber was not due to these own high expression in tuber, but indirectly proved GH3 gene regulation tuber cell enlargement through the regulated JA biological processes.

3. Conclusion

We performed an evolutionary analysis of the GH3 protein family in the plant kingdom to reveal gene structure, phylogenetic relationships, and the evolution of GH3 genes in each group. Group I and group II genes were occurred in mosses. And group I contained a lot of ancient genes, thus we inferred that the function of group I was indispensable in all plants. Multiple group II genes appeared in seed plants; hence, we could infer that group II genes played crucial roles in plants. Group III members did not appear until the angiosperm period, so we speculated that group III was the latest group and was closely related with group II. This conclusion will provide a reference for the evolutionary relationship of the GH3 family in plants. And our analysis also revealed that group I genes were related to JA response and several genes were also involved in physiological processes of various tissues and responded some stresses in potato. The results of this study will increase our understanding of the evolutionary relationships in the GH3 family and also serve as the basis for the functional identification of potato GH3 genes. In subsequent studies, we will devote to reveal and prove the function of GH3 genes about JA biological process and improves resistance to some biotic stresses based on JA regulation in potato.

4. Materials and Methods

4.1 Mining GH3 genes from various species

Gene, protein, and transcript sequences from 48 species of algae, mosses, ferns, gymnosperms, and angiosperms were downloaded from the NCBI. Blast 2.6.0 was used to search for homologous

sequences based on *AtGH3* [12] as query with E-value  $\leq 10^{-10}$ . The acquired data was uploaded to NCBI-CDD website (<http://www.ncbi.nlm.nih.gov/cdd>) and Pfam website (<http://pfam.xfam.org/>) to search the domain. All identified GH3 genes were aligned using the multiple sequence alignment tool ClustalX2 (<http://www.clustal.org/clustal2/>). After excluding small portions of genes with divergent sequences, the others were considered as putative genes.

#### 4.2 Structure analysis of GH3 proteins

To better understand GH3 proteins, a portion of GH3 proteins to analyze their structures. These proteins were predicted from the motifs using MEME (<http://meme-suite.org/tools/meme>), with the parameters set to default, only select the number of motifs is 20. Predictions of protein secondary structure prediction came from PHDsec server, which could upload the data to PRABI Lyon Gerland website (<https://prabi.ibcp.fr/htm/site/web/home>), and DNAMAN software (<https://www.lynnon.com/pc/framepc.html>) was used to analyze multiple sequence alignment.

#### 4.3 Construction of phylogenetic tree

Phylogenetic analysis of the GH3 family was conducted using MEGA7 software (<https://www.megasoftware.net/>). Seventeen species of plants were included in the phylogenetic tree. Phylogenetic trees were produced using the Neighbor-Joining (NJ) method [27] with the following parameters: 1000 bootstrap replications, Poisson model, and pairwise deletions.

#### 4.4 Analysis of gene duplications

Tandemly duplicated gene pairs were identified by comparing their physical locations on chromosomes and their homology (more than 50%). We defined paralogous genes as those existing in the same chromosome within a 50-kb physical distance that were tandem duplicated pairs [28, 33]. The segmental duplication of each gene was ensured through Plant Genome Duplication Database (PGDD) website (<http://chibba.agtec.uga.edu/duplication/>). Ka/Ks values were calculated using DnaSP6 software [34].

#### 4.5 GO annotation and RNA-seq data analysis

Blast2GO software (<https://www.blast2go.com/>) was used to analyze the gene ontology (GO). The full-length amino acid sequences were uploaded to the program, and the NCBI database was chosen as the reference to analyze molecular function, cellular component, and biological process. The KEGG pathway of proteins could be searched on the KEGG website (<http://www.kegg.jp/kegg/kegg2.html>).

The RNA-seq data [35] were downloaded from PGSC website ([http://solanaceae.plantbiology.msu.edu/pgsc\\_download.shtml](http://solanaceae.plantbiology.msu.edu/pgsc_download.shtml)). The selected raw data were transformed by log2, and then HemI software [36] was used to visualize the expression.

#### 4.6 Plant growth conditions and treatments

The potato material (DM1-3-516-R44) was grown with 30 g/L sugars and solid Murashige & Skoog Basal Medium with Vitamins (MS, USA) as culture medium for four weeks in a plant incubator at 25±1°C under 10 000 Lx in light for 16 h and 20±1°C under 0 Lx for 8h. For MeJA treatment, 10<sup>-4</sup> M MeJA was added to liquid MS medium with 20 g/L sugars. After culturing for 1 h, 6 h, 12 h, and 24 h, samples were taken to analyze the expression.

B5141-6 (Lenape)×Wauseon was planted in the greenhouse for 3 months. Seven tissues (young leaf, young stem, young root, stolon, young tuber, mature tuber, and flower) were selected to test gene expression.

#### 4.7 Expression analysis of *StGH3* genes

Total RNA from four different types of tissue culture seedlings were treated with MeJA (Invitrogen, USA) and seven different tissues were used for reverse transcription into cDNA (Takara, Japan). Elongation factor 1- $\alpha$  (*ef1- $\alpha$* ) was used as a reference gene to quantify the expression of *StGH3* genes [37]. Bio-Rad Real-Quantitative real-time PCR analysis System (CFX96, USA) was used to analyze expression levels. The expression data of each treatment was based on three technical replicates. Relative expression levels were calculated using  $2^{-\Delta\Delta Ct}$  [38, 39].

**Supplementary Materials:** Supplementary materials can be found online.

**Funding:** This work was supported by the National Natural Science Foundation of China (No. 31601358), and the Project of Science and Technology from Shaanxi Province (No. 2017ZDXM-NY-004).

**Conflicts of Interest:** The authors have no conflicts of interest to declare.

#### Author Contributions:

C.Z., C.Y. and C.Q. designed experiments; C.Z., C.Q., L.Z., and D.W. carried out experiments; C.Z., Z.S. and X.M. analyzed experimental results. C.Z., H.M. and B.L. analyzed data and developed analysis tools. C.Z., L.Z. and B.L. wrote the manuscript.

#### References

1. Hagen, G.; Guilfoyle, T. J., Rapid induction of selective transcription by auxins. *Molecular & Cellular Biology* **1985**, 5, (6), 1197.
2. Roux, C.; Perrot-Rechenmann, C., Isolation by differential display and characterization of a tobacco auxin-responsive cDNA Nt-gh3, related to GH3. *FEBS Lett* **1997**, 419, (1), 131-6.
3. Tanaka, S.; Mochizuki, N.; Nagatani, A., Expression of the AtGH3a gene, an Arabidopsis homologue of the soybean GH3 gene, is regulated by phytochrome B. *Plant Cell Physiol* **2002**, 43, (3), 281-9.
4. Jain, M.; Kaur, N.; Tyagi, A. K.; Khurana, J. P., The auxin-responsive GH3 gene family in rice (*Oryza sativa*). *Funct Integr Genomics* **2006**, 6, (1), 36-46.
5. Yuan, H.; Zhao, K.; Lei, H.; Shen, X.; Liu, Y.; Liao, X.; Li, T., Genome-wide analysis of the GH3 family in apple (*Malus x domestica*). *BMC Genomics* **2013**, 14, 297.
6. Yang, Y.; Yue, R.; Sun, T.; Zhang, L.; Chen, W.; Zeng, H.; Wang, H.; Shen, C., Genome-wide identification, expression analysis of GH3 family genes in *Medicago truncatula* under stress-related hormones and *Sinorhizobium meliloti* infection. *Applied microbiology and biotechnology* **2015**, 99, (2), 841-54.
7. Mittag, J.; Gabrielyan, A.; Ludwig-Muller, J., Knockout of GH3 genes in the moss *Physcomitrella patens* leads to increased IAA levels at elevated temperature and in darkness. *Plant Physiol Biochem* **2015**, 97, 339-49.
8. Liao, D.; Chen, X.; Chen, A.; Wang, H.; Liu, J.; Liu, J.; Gu, M.; Sun, S.; Xu, G., The characterization of six auxin-induced tomato GH3 genes uncovers a member, SlGH3.4, strongly responsive to arbuscular mycorrhizal symbiosis. *Plant Cell Physiol* **2015**, 56, (4), 674-87.
9. Westfall, C. S.; Zubieta, C.; Herrmann, J.; Kapp, U.; Nanao, M. H.; Jez, J. M., Structural basis for prereceptor modulation of plant hormones by GH3 proteins. *Science* **2012**, 336, (6089), 1708-11.



10. Mellor, N.; Band, L. R.; Pencik, A.; Novak, O.; Rashed, A.; Holman, T.; Wilson, M. H.; Voss, U.; Bishopp, A.; King, J. R.; Ljung, K.; Bennett, M. J.; Owen, M. R., Dynamic regulation of auxin oxidase and conjugating enzymes AtDAO1 and GH3 modulates auxin homeostasis. *Proc Natl Acad Sci U S A* **2016**, 113, (39), 11022-7.
11. Okrent, R. A.; Wildermuth, M. C., Evolutionary history of the GH3 family of acyl adenylases in rosids. *Plant Mol Biol* **2011**, 76, (6), 489-505.
12. Westfall, C. S.; Herrmann, J.; Chen, Q.; Wang, S.; Jez, J. M., Modulating plant hormones by enzyme action: the GH3 family of acyl acid amido synthetases. *Plant Signal Behav* **2010**, 5, (12), 1607-12.
13. Westfall, C. S.; Sherp, A. M.; Zubieta, C.; Alvarez, S.; Schraft, E.; Marcellin, R.; Ramirez, L.; Jez, J. M., Arabidopsis thaliana GH3.5 acyl acid amido synthetase mediates metabolic crosstalk in auxin and salicylic acid homeostasis. *Proc Natl Acad Sci U S A* **2016**, 113, (48), 13917-13922.
14. Okrent, R. A.; Brooks, M. D.; Wildermuth, M. C., Arabidopsis GH3.12 (PBS3) conjugates amino acids to 4-substituted benzoates and is inhibited by salicylate. *J Biol Chem* **2009**, 284, (15), 9742-54.
15. Takase, T.; Nakazawa, M.; Ishikawa, A.; Kawashima, M.; Ichikawa, T.; Takahashi, N.; Shimada, H.; Manabe, K.; Matsui, M., ydk1-D, an auxin-responsive GH3 mutant that is involved in hypocotyl and root elongation. *Plant J* **2004**, 37, (4), 471-83.
16. Nakazawa, M.; Yabe, N.; Ichikawa, T.; Yamamoto, Y. Y.; Yoshizumi, T.; Hasunuma, K.; Matsui, M., DFL1, an auxin-responsive GH3 gene homologue, negatively regulates shoot cell elongation and lateral root formation, and positively regulates the light response of hypocotyl length. *Plant J* **2001**, 25, (2), 213-21.
17. Khan, S.; Stone, J. M., Arabidopsis thaliana GH3.9 influences primary root growth. *Planta* **2007**, 226, (1), 21-34.
18. Paponov, I. A.; Teale, W.; Lang, D.; Paponov, M.; Reski, R.; Rensing, S. A.; Palme, K., The evolution of nuclear auxin signalling. *BMC Evol Biol* **2009**, 9, 126.
19. Ren, H.; Gray, W. M., SAUR Proteins as Effectors of Hormonal and Environmental Signals in Plant Growth. *Mol Plant* **2015**, 8, (8), 1153-64.
20. Withers, J.; Yao, J.; Mecey, C.; Howe, G. A.; Melotto, M.; He, S. Y., Transcription factor-dependent nuclear localization of a transcriptional repressor in jasmonate hormone signaling. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, 109, (49), 20148-20153.
21. Terol, J.; Domingo, C.; Talon, M., The GH3 family in plants: genome wide analysis in rice and evolutionary history based on EST analysis. *Gene* **2006**, 371, (2), 279-90.
22. Nystedt, B.; Street, N. R.; Wetterbom, A.; Zuccolo, A.; Lin, Y. C.; Scofield, D. G.; Vezzi, F.; Delhomme, N.; Giacomello, S.; Alexeyenko, A., The Norway spruce genome sequence and conifer genome evolution. *Nature* **2013**, 497, (7451), 579.
23. Böttcher, C.; Boss, P. K.; Davies, C., Acyl substrate preferences of an IAA-amido synthetase account for variations in grape (*Vitis vinifera* L.) berry ripening caused by different auxinic compounds indicating the importance of auxin conjugation in

- plant development. *Journal of Experimental Botany* **2011**, 62, (12), 4267-4280.
24. Feng, S.; Yue, R.; Tao, S.; Yang, Y.; Zhang, L.; Xu, M.; Wang, H.; Shen, C., Genome-wide identification, expression analysis of auxin-responsive GH3 family genes in maize (*Zea mays* L.) under abiotic stresses. *J Integr Plant Biol* **2015**, 57, (9), 783-95.
  25. GROßE-VELDMANN; KOROTKOVA; REINKEN; LOBIN; Barthlott, W., *Amborella trichopoda* – Cultivation of the most ancestral angiosperm in Botanic Gardens. *Veldmann* **2011**, 19, 143-155.
  26. Kuang, J. F.; Zhang, Y.; Chen, J. Y.; Chen, Q. J.; Jiang, Y. M.; Lin, H. T.; Xu, S. J.; Lu, W. J., Two GH3 genes from longan are differentially regulated during fruit growth and development. *Gene* **2011**, 485, (1), 1-6.
  27. Singh, V. K.; Jain, M.; Garg, R., Genome-wide analysis and expression profiling suggest diverse roles of GH3 genes during development and abiotic stress responses in legumes. *Front Plant Sci* **2014**, 5, 789.
  28. Zhang, C.; Wang, D.; Yang, C.; Kong, N.; Shi, Z.; Zhao, P.; Nan, Y.; Nie, T.; Wang, R.; Ma, H.; Chen, Q., Genome-wide identification of the potato WRKY transcription factor family. *PLoS One* **2017**, 12, (7), e0181573.
  29. Takahashi, K.; Fujino, K.; Kikuta, Y.; Koda, Y., Expansion of potato cells in response to jasmonic acid. *Plant Science* **1994**, 100, (1), 3-8.
  30. Koda, Y.; Kikuta, Y.; Tazaki, H.; Tsujino, Y.; Sakamura, S.; Yoshihara, T., Potato tuber-inducing activities of jasmonic acid and related compounds. *Phytochemistry* **1991**, 30, (5), 1435-1438.
  31. Ulloa, R. M.; Raíces, M.; Macintosh, G. C.; Maldonado, S.; Téllezziñón, M. T., Jasmonic acid affects plant morphology and calcium-dependent protein kinase expression and activity in *Solanum tuberosum*. *Physiol Plant* **2002**, 115, (3), 417.
  32. Sarkar, D.; Pandey, S. K.; Sharma, S., Cytokinins antagonize the jasmonates action on the regulation of potato ( *Solanum tuberosum* ) tuber formation in vitro. *Plant Cell Tissue & Organ Culture* **2006**, 87, (3), 285.
  33. Kong, H.; Landherr, L. L.; Frohlich, M. W.; Leebens - Mack, J.; Ma, H.; DePamphilis, C. W., Patterns of gene duplication in the plant SKP1 gene family in angiosperms: evidence for multiple mechanisms of rapid gene birth. *The Plant journal : for cell and molecular biology* **2007**, 50, (5), 873.
  34. Rozas, J.; Ferrermata, A.; Sánchezdelbarrio, J. C.; Guiraorico, S.; Librado, P.; Ramosonsins, S. E.; Sánchezgracia, A., DnaSP 6: DNA Sequence Polymorphism Analysis of Large Datasets. *Molecular Biology & Evolution* **2017**.
  35. Potato Genome Sequencing, C.; Xu, X.; Pan, S.; Cheng, S.; Zhang, B.; Mu, D.; Ni, P.; Zhang, G.; Yang, S.; Li, R.; Wang, J.; Orjeda, G.; Guzman, F.; Torres, M.; Lozano, R.; Ponce, O.; Martinez, D.; De la Cruz, G.; Chakrabarti, S. K.; Patil, V. U.; Skryabin, K. G.; Kuznetsov, B. B.; Ravin, N. V.; Kolganova, T. V.; Beletsky, A. V.; Mardanov, A. V.; Di Genova, A.; Bolser, D. M.; Martin, D. M.; Li, G.; Yang, Y.; Kuang, H.; Hu, Q.; Xiong, X.; Bishop, G. J.; Sagredo, B.; Mejia, N.; Zagorski, W.; Gromadka, R.; Gawor, J.; Szczesny, P.; Huang, S.; Zhang, Z.; Liang, C.; He, J.; Li, Y.; He, Y.; Xu, J.; Zhang, Y.; Xie, B.; Du, Y.; Qu, D.; Bonierbale, M.; Ghislain, M.;

- Herrera Mdel, R.; Giuliano, G.; Pietrella, M.; Perrotta, G.; Facella, P.; O'Brien, K.; Feingold, S. E.; Barreiro, L. E.; Massa, G. A.; Diambra, L.; Whitty, B. R.; Vaillancourt, B.; Lin, H.; Massa, A. N.; Geoffroy, M.; Lundback, S.; DellaPenna, D.; Buell, C. R.; Sharma, S. K.; Marshall, D. F.; Waugh, R.; Bryan, G. J.; Destefanis, M.; Nagy, I.; Milbourne, D.; Thomson, S. J.; Fiers, M.; Jacobs, J. M.; Nielsen, K. L.; Sonderkaer, M.; Iovene, M.; Torres, G. A.; Jiang, J.; Veilleux, R. E.; Bachem, C. W.; de Boer, J.; Borm, T.; Kloosterman, B.; van Eck, H.; Datema, E.; Hekkert, B.; Goverse, A.; van Ham, R. C.; Visser, R. G., Genome sequence and analysis of the tuber crop potato. *Nature* **2011**, 475, (7355), 189-95.
36. Liu, C.; Wang, X.; Xu, Y.; Deng, X.; Xu, Q., Genome-wide analysis of the R2R3-MYB transcription factor gene family in sweet orange (*Citrus sinensis*). *Mol Biol Rep* **2014**, 41, (10), 6769-85.
  37. Nicot, N.; Hausman, J. F.; Hoffmann, L.; Evers, D., Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *J Exp Bot* **2005**, 56, (421), 2907-14.
  38. Liu, B.; Zhang, G.; Murphy, A.; De, K. D.; Tai, H.; Bizimungu, B.; Si, H.; Li, X. Q., Differences between the Bud End and Stem End of Potatoes in Dry Matter Content, Starch Granule Size, and Carbohydrate Metabolic Gene Expression at the Growing and Sprouting Stages. *Journal of Agricultural & Food Chemistry* **2016**, 64, (5), 1176.
  39. Zhang, C.; Wang, D.; Yang, C.; Kong, N.; Shi, Z.; Zhao, P.; Nan, Y.; Nie, T.; Wang, R.; Ma, H., Genome-wide identification of the potato WRKY transcription factor family. *Plos One* **2017**, 12, (7), e0181573.