

Article

# Genome-Wide Identification and Comparative Analysis for OPT Family Genes in *Panax ginseng* and Eleven Flowering Plants

He Su <sup>1,†</sup>, Yang Chu <sup>2,†</sup>, Junqi Bai <sup>1</sup>, Lu Gong <sup>1</sup>, Juan Huang <sup>1</sup>, Wen Xu <sup>1</sup>, Jing Zhang <sup>1</sup>, Xiaohui Qiu <sup>1</sup>, Jiang Xu <sup>2,\*</sup> and Zhihai Huang <sup>1,\*</sup>

<sup>1</sup> The Second Clinical College of Guangzhou University of Chinese Medicine, Guangzhou 510006, China; suhe@cau.edu.cn (H.S.); baijunqi@126.com (J.B.); gonglu0904@126.com (L.G.); juanhuangzi@126.com (J.H.); freexuwen@163.com (W.X.); ginniezj@163.com (J.Z.); qiuxiaohui@gzucm.edu.cn (X.Q.)

<sup>2</sup> Institute of Chinese Materia Medica, China Academy of Chinese Medicinal Sciences, Beijing 100700, China; ychu@icmm.ac.cn

\* Correspondence: jxu@icmm.ac.cn (J.X.); zhhuang7308@163.com (Z.H.); Tel.: +86-20-39318571 (Z.H.)

† These authors contributed equally to this work.

Received: 9 November 2018; Accepted: 17 December 2018; Published: 20 December 2018



**Abstract:** Herb genomics and comparative genomics provide a global platform to explore the genetics and biology of herbs at the genome level. *Panax ginseng* C.A. Meyer is an important medicinal plant for a variety of bioactive chemical compounds of which the biosynthesis may involve transport of a wide range of substrates mediated by oligopeptide transporters (OPT). However, information about the OPT family in the plant kingdom is still limited. Only 17 and 18 OPT genes have been characterized for *Oryza sativa* and *Arabidopsis thaliana*, respectively. Additionally, few comprehensive studies incorporating the phylogeny, gene structure, paralogs evolution, expression profiling, and co-expression network between transcription factors and OPT genes have been reported for ginseng and other species. In the present study, we performed those analyses comprehensively with both online tools and standalone tools. As a result, we identified a total of 268 non-redundant OPT genes from 12 flowering plants of which 37 were from ginseng. These OPT genes were clustered into two distinct clades in which clade-specific motif compositions were considerably conservative. The distribution of OPT paralogs was indicative of segmental duplication and subsequent structural variation. Expression patterns based on two sources of RNA-Sequence datasets suggested that some OPT genes were expressed in both an organ-specific and tissue-specific manner and might be involved in the functional development of plants. Further co-expression analysis of OPT genes and transcription factors indicated 141 positive and 11 negative links, which shows potent regulators for OPT genes. Overall, the data obtained from our study contribute to a better understanding of the complexity of the OPT gene family in ginseng and other flowering plants. This genetic resource will help improve the interpretation on mechanisms of metabolism transportation and signal transduction during plant development for *Panax ginseng*.

**Keywords:** *Panax ginseng*; oligopeptide transporter; flowering plant; phylogeny; transcription factor

## 1. Introduction

Peptide transportation is a widely observed phenomenon of translocating small peptides across a membrane in a carrier-mediated, energy-dependent manner [1]. Transported peptides are often hydrolyzed and the resulting amino acids are used as substrates for protein synthesis, sources of nitrogen and carbon [2,3], and signals for biological processes such as quorum sensing [4], yeast mating [5], and metal homeostasis regulation [6,7]. There are three distinct protein families related

to peptide transportation. The ATP binding cassette (ABC, TC 3.A.1) transporter superfamily is the largest transporter gene family. The members are able to translocate a wide variety of substrates including amino acids, sugars, peptides, proteins, and a large number of hydrophobic compounds and metabolites across extra-cellular and intracellular membranes [8,9]. In contrast to the ABC family, the proton-dependent oligopeptide transporter (PTR, TC 2.A.17) family utilizes a proton gradient other than ATP hydrolysis for dipeptide and tripeptide translocation [10]. The members of PTR proteins have been found in all kingdoms of life except the Archaea [1,11]. PTR also participates in amino acid and nitrate transportation [12]. In addition to dipeptides and tripeptides that are translocated by PTR proteins, tetra-peptides, penta-peptides, and some longer oligopeptides are translocated by a novel protein family known as the oligopeptide transporter (OPT, TC 2.A.67) family [10].

The OPT family is a group of electrochemical potential-driven transporters that catalyze their solutes in an energy-dependent symport manner. CaOPT1 was first cloned from *Candida albicans* (Robin) Berkhout and functional verified in *Schizosaccharomyces pombe* (Lindner) and subsequently defined as OPT but not an ABC or PTR protein by Jeff Becker's laboratory [13–15]. OPTs are suggested to play diverse roles in long-distance sulfur distribution, metal homeostasis, nitrogen mobilization, heavy metal sequestration by transporting glutathione, peptides, and meta-chelates [16]. Phylogenetically, the OPTs can be divided into Oligopeptide Transporter (PT) and Stripe-like (YSL) clades [16,17]. Genes in the YSL clade have been found in Archaea, eubacteria, fungi, and plants but not in animals, which function as metal chelate transporters [6,18,19] consisting of mugenic acids (MA) or nicotianamine (NA) while genes in the PT clade have only been identified in plants and fungi mediating long-distance metal distribution, nitrogen mobilization, glutathione translocation, and heavy metal sequestration [16,20–25].

In plants, the OPTs may play important roles in plant growth and abiotic and biotic stress responses [13,26]. The OPT member *ZmYS1*, which was first cloned by Curie et al. [6] but re-defined by Yen et al. [27], was proven to mediate the import of Fe-phytosiderophore complexes from soils and long-distance transport of iron-NA complexes [13]. Studies of two *AtOPT3* T-DNA mutants indicated that *AtOPT3* is of importance in both embryo development and iron deficiency signal transduction [1,7]. In addition, *AtOPT3* is found to be expressed in the phloem and functions in long-distance shoot-to-root signaling for Fe/Zn/Mn status. A lack of *AtOPT3* in *Arabidopsis thaliana* (*Arabidopsis*) led to the over-accumulation of cadmium in seeds [23]. Glutathione (GSH) is an essential sulfur-containing tripeptide that performs various important roles in plant processes, including detoxification of xenobiotics, heavy metal transport and resistance, controlling redox status, and long-distance transport of organic sulfur [22]. GSH is a precursor for plants to use to produce phytochelatin (PCs), which is the polymerized form of GSH, by which heavy metals can be transported to a central vacuole for detoxification [28]. *AtOPT4* and *AtOPT6* from *Arabidopsis* [21,22] and *BjGT1* [25] from *Brassica juncea* (*B. juncea*) are all capable of translocating Cd-GSH conjuncts. Moreover, GSH also plays important roles in plant growth and development in response to abiotic and biotic stresses.

The majority of members of OPT proteins seem to contain 16 TMSs including a few of which appear to have 17 TMSs. A homology-based analysis for each TMSs in the OPT family indicated that the 16-TMS proteins might have been generated by three sequential duplications from 2-TMS protein precursors. Additionally, gene fusion might be responsible for the 17-TMS proteins [17]. However, although the OPT proteins have been studied for more than two decades, the majority of studies still focus on model plants such as yeast, *Arabidopsis*, and *Oryza sativa* (rice) [1,21,29,30]. Thanks to the rapid development of whole-genome sequencing techniques, an exponential increase in genome information has provided us with great opportunities to identify more OPT genes in non-model plants and make comparisons among multiple species simultaneously. However, to the best of our knowledge, genome-wide identification of OPT proteins has only been conducted in *Ganoderma lucidum* [31], *Populus trichocarpa*, and *Vitis vinifera* [32]. *Panax ginseng*, which is a Traditional Chinese Medicine, has been used for several millennia and has become more and more popular around the world. It is the most commonly used medicinally species in the *Panax* genera in contrast to the other

four species: *Panax quinquefolius*, *Panax vietnamensis*, *Panax japonicus*, and *Panax notoginseng* [33]. Since we finished the genome assembly for *Panax ginseng* (*P. ginseng*) in our previous report [34], the interest in characterizing OPT genes in *P. ginseng* and comparing it with other genome-assembly-available species has increased.

Although more plant genomes have been mapped in the last decade, studies on the genome-wide identification and comparison of OPT genes among species are still limited. Information on the phylogeny, gene structure, expression patterns, and regulatory networks of OPT genes remains to be discovered. In the present study, we identified OPT genes from *P. ginseng* and 11 flowering plants with the purpose of uncovering the phylogenetic relationships and gene structures of OPT genes in flowering plants as well as investigating the expression profiles and regulators of OPT genes in *P. ginseng*. Our analysis, which combines these types of information, provides new insights into both the structural and functional roles of OPT genes in ginseng and serves as a valuable resource for further study of the roles OPT genes play in plant development and transport of secondary metabolism.

## 2. Results and Discussion

### 2.1. Identification of OPT Genes in *P. ginseng* and 11 Other Flowering Plants

We identified the OPT genes for *P. ginseng* and other species with TransportTP by setting as reference organisms *Oryza sativa* and *Arabidopsis thaliana* [35]. As a result, a total of 364 OPT candidates were identified in our study (Table 1). Seventeen of the 18 identified genes from *Arabidopsis* were in accordance with the reviewed records deposited in Swiss-Prot (release 2018\_10), and, although the other gene *At5g45450*, was not recorded in Swiss-Prot, it was regarded as an OPT gene recorded in GenBank. However, only 12 OPT genes were identified from rice, of which the accession numbers were not in accordance with those records deposited in Swiss-Prot. This might be due to a different version of the rice genome being used. Considering our genome assembly did not scale to a chromosome level, we conducted a manual curation of the 39 OPTs identified from *P. ginseng*. Thereafter, 37 OPT genes were kept for further analysis. In addition, we identified 54 and 26 OPT genes from poplar and grape in our study, respectively, while only 20 and 18 genes were identified by Cao et al. [32]. These results suggested that our identification of OPT genes was accurate and comprehensive.

**Table 1.** Statistics of OPT genes predicted from *P. ginseng* and 11 flowering plants.

Species	Predicted	De-Redundant	Final
Ginseng	39	37	37
<i>Arabidopsis</i>	18	16	17
Rice	12	10	18
Sorghum	38	26	26
Carrot	25	16	16
Potato	42	29	29
Tomato	23	17	17
Cassava	29	21	21
Clover	31	25	25
Cacao	27	19	19
Poplar	54	28	28
Grape	26	24	23
Total	364	268	276

37 OPT genes were left after manual curation for OPT genes from Ginseng. OPT genes predicted for *Arabidopsis* and rice were replaced with 17 and 18 reviewed OPT genes retrieved from Swiss-Prot. De-redundant indicated only one gene could be kept if there were genes that had 100% similarity to it.

Since we found some OPTs were highly redundant (similar to each other with 100% similarity) within species, we removed the redundant OPTs in order to reduce the subsequent calculation consumption using CD-HIT software [36] by setting the sequence identity threshold to 100%. Lastly, a total of 268 OPT genes were kept. Because a candidate OPT gene named “GSVIVT01007176001”

identified from grape contained too many “X”s, we excluded it from further analyses. In order to generate robust results from subsequent studies, we replaced those predicted OPT genes from *Arabidopsis* and rice with reviewed OPT genes retrieved from Swiss-Prot. Furthermore, we introduced two other experimentally verified OPT genes from *B. juncea* and *Zea mays* (*BjGT1* and *Maize\_YS\_1* respectively [6,25]) into our study. Lastly, 278 OPT genes were used for further analysis (Supplementary File 1).

## 2.2. Protein Properties of OPT Genes for OPT Genes Identified in *P. ginseng* and 11 Other Flowering Plants

By examining the properties of OPT genes for each plant species, we found that the number of amino acid residues varied among species. Generally, the number of amino acid residues for OPT genes in *Arabidopsis*, rice, sorghum, and cassava ranged from 552 to 766, which is higher than the rest of those studied species (ranged from 184 to 941, as for *P. ginseng*). The number of residues ranged from 348 to 919 (Table S1, Figure 1D). The distribution of molecular weight for OPT genes was similar to the distribution pattern of residue numbers (Figure 1B). The grand average of hydropathicity (GRAVY) value is a measure of protein hydrophobicity [37]. Our results suggested that GRAVY for those OPT genes mainly ranged from 0.30 to 0.60 (Figure 1A). As OPT genes with the lowest and the highest GRAVY values (0.029 for Potri.017G150620.2.p and 0.87 for PGSC0003DMP400037534) were filtered out for lacking OPT-specific information for further phylogenetic analysis (Figure 2), we expanded the confident range of GRAVY values from 0.329–0.628 to 0.114–0.659 compared with the previous study [32]. In addition, the isoelectric point (pI) of the majority of OPT genes was around 9.0, which suggests that the electrochemical properties of OPT genes might be less varied in the plant kingdom (Figure 1C).

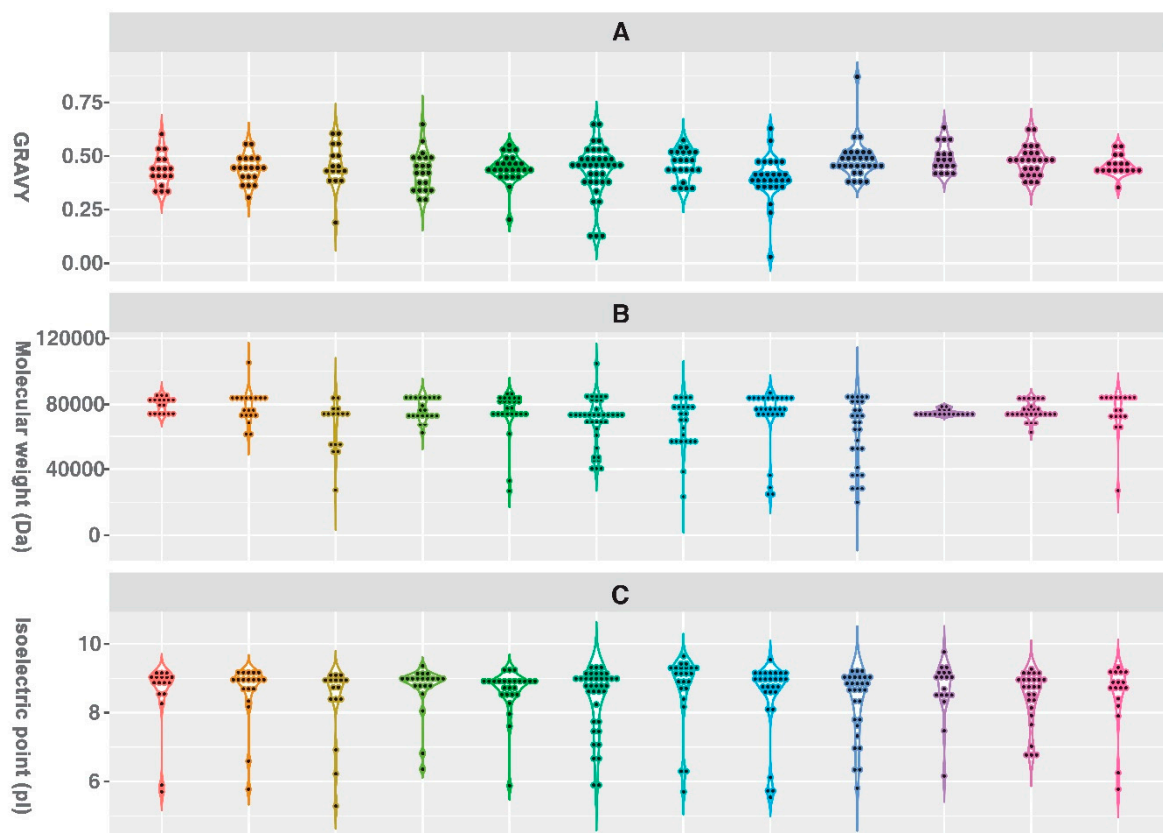
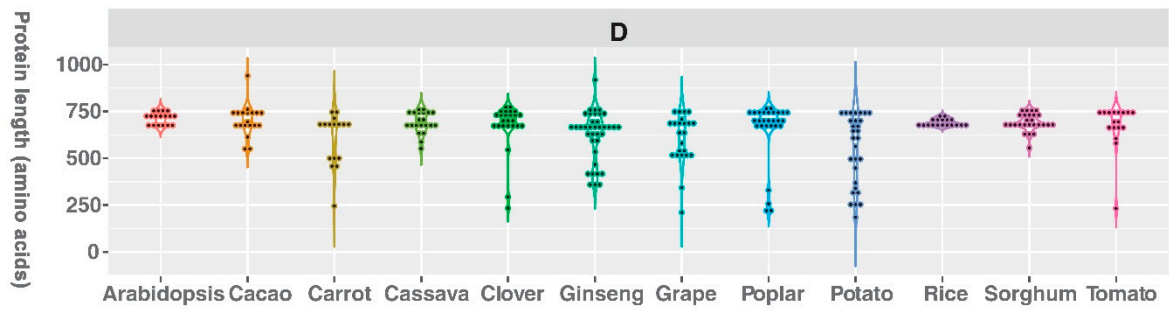
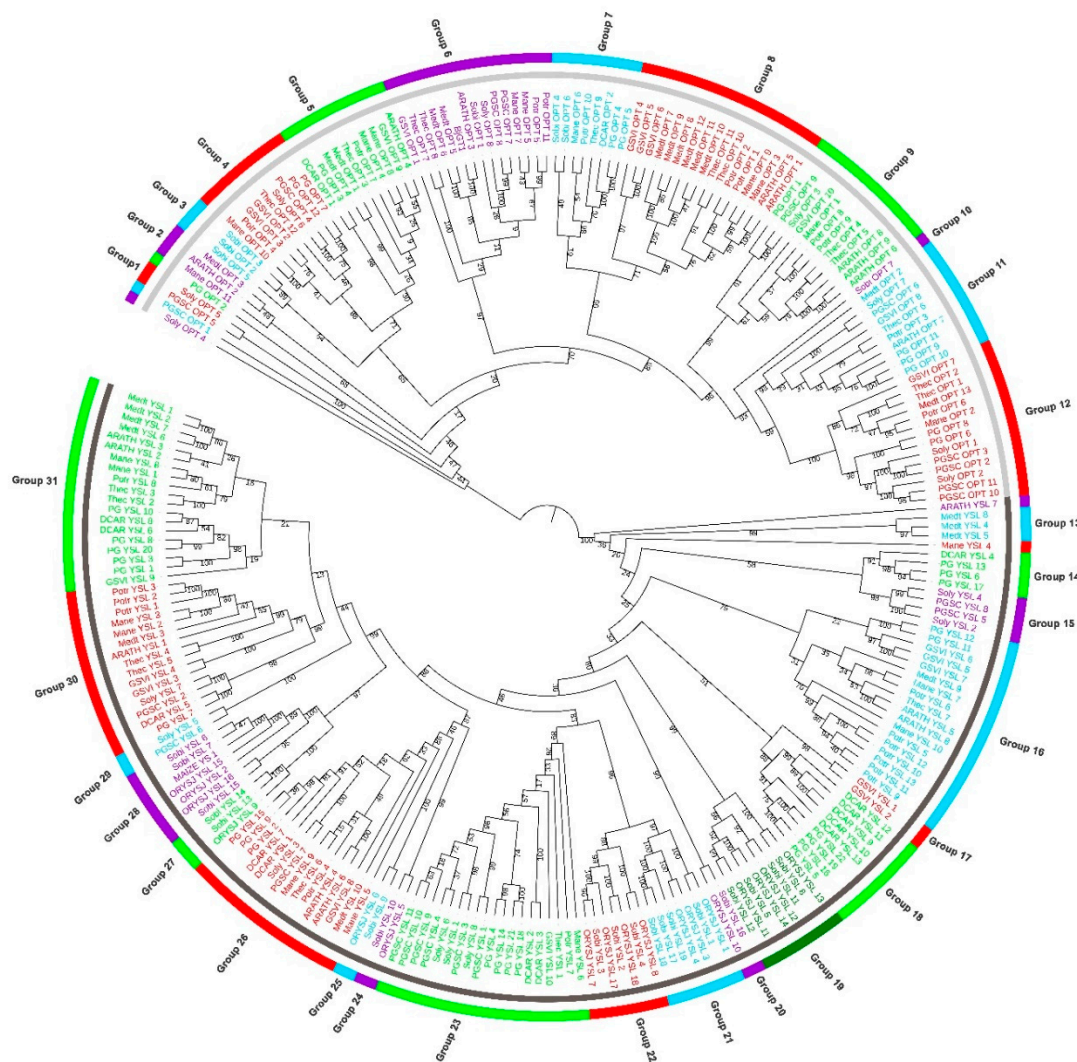


Figure 1. Cont.





**Figure 1.** Protein properties for OPT genes identified from *P. ginseng* and 11 flowering plants. (A) Grand average of hydropathicity, GRAVY. (B) Molecular weight. (C), Isoelectric point, pI. (D) Number of amino acid residues.



**Figure 2.** Phylogenetic relationships of OPT genes in *P. ginseng* and 11 other species. Tomato (Soly), Potato (PGSC), Cassava (Mane), *Arabidopsis* (ARATH), Clover (Medt), Poplar (Potr), Grape (GSVI), Cacao (Thec), Sorghum (Sobi), Carrot (DCAR), and Rice (ORYSJ). BjGT1 from *Brassica juncea* and Maize YS1 from *Zea Mays* (maize) are experimentally validated OPT proteins that were retrieved from GenBank database. Light gray in the inner circle indicates the PT clade. Dark gray refers to the YSL clade.

Further analysis conducted with WOLF PSORT (<http://wolftpsort.org>) enabled us to predict the probable protein localization for each candidate OPT identified in our study. It was found that all candidate OPTs were most likely to be located in the plasma and vacuolar membranes. The results were in accordance with a previous study [32]. Furthermore, 17 OPT genes were predicted to only be located in the plasma membrane. The remaining OPT genes were predicted to be not only in the plasma but also in at least one of the following: vacuole, chloroplast, cytoplasm, nucleus, mitochondria, Golgi apparatus, or endoplasmic reticulum (Table S1).

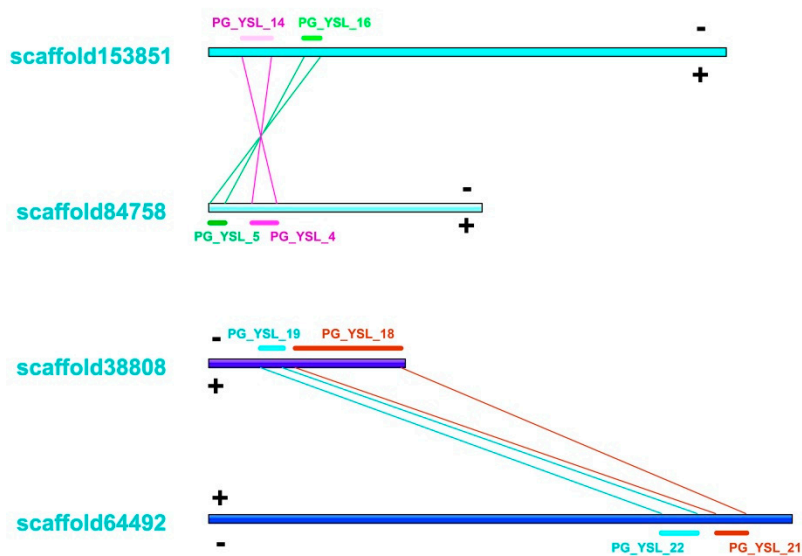
### 2.3. Phylogenetic Analyses, Classification, and Functional Relatedness of the OPT Genes Identified in *P. ginseng* and 11 Other Flowering Plants

To unravel the phylogenetic relationships of OPT genes in flowering plants, we conducted a phylogenetic analysis for those genes from 12 flowering plants. All OPT genes were clustered into two major distinct clades known as PT and YSL clade for which the results were consistent with those of previous reports [13,16,32]. However, what was different from previous studies was that the rice OPT genes in the PT clade were not included because no rice OPT genes in this clade were available in the Swiss-Prot database (release 2018\_10). Therefore, only OPTs from the YSL clade were used in this phylogenetic analysis. Based on the bootstrap permutation test and the relationships of each OPT gene, we further classified the PT clade into 12 subgroups (Groups 1-12) and the YSL clade into 19 subgroups (Groups 13-31). Groups 23 and 31 included the largest number of members in the YSL clade (each with 19 members). Groups 9-12 formed a highly confident larger group with a bootstrap value of 96% in the OPT clade and Groups 23 and 27-30 formed another group in the YSL clade with a supporting value of 99%, which suggested that those members were likely to have evolved by recent gene duplication from a common ancestor. However, Soly\_OPT\_4 (Tomato), PGSC\_OPT\_1 (Potato), and PG\_OPT\_2 (Ginseng) failed to be grouped with any other PT genes due to a lack of supporting information by maximum likelihood analyses. In addition, ARTH\_YSL\_7 (*Arabidopsis*) and Mane\_YSL\_4 (Cassava) also failed to be grouped with any other genes in the YSL clade. Although Sobi\_OPT\_7 (Sorghum) seemed likely to stand alone, it was in fact grouped with Groups 11 and 12 with a bootstrap value of 93% (Figure 2). Furthermore, the motif structures of the genes described below also supported the group classifications (Figure S1). Moreover, ARATH\_OPT\_3 and BjGT3 (*Brassica juncea*), ARATH\_OPT\_1 and ARATH\_OPT\_5, ARATH\_OPT\_6 and ARATH\_OPT\_8 and ARATH\_OPT\_9, ARATH\_YSL\_5 and ARATH\_YSL\_8, ARATH\_YSL\_4 and ARATH\_YSL\_6, ORYSJ\_YSL\_7 (rice) and ORYSJ\_YSL\_17 were grouped together, with the phylogenetic relationships in accordance with previous study reports [16,30,32]. The consistency of our findings with previous findings indicated that our phylogenetic study was properly conducted and the results were reliable. However, it was interesting to find out that ARATH\_YSL\_7, which has been reported to be sub-grouped with ARATH\_YSL\_5 and ARATH\_YSL\_8 [30], failed to be grouped with any other OPT members in the YSL clade in our study.

Genes with the same functions were often closely related, as found in both a previous study [32] and our study. *BjGT1*, which is the first cloned and characterized OPT gene from *Brassica juncea*, was experimentally validated to be a glutathione transporter mediating cadmium absorption [25]. ARATH\_OPT\_3, which is another OPT gene that was cloned and characterized in *Arabidopsis*, was reported to be involved in the sensing and translocation of Cd (as well as Fe and Zn) [1,7,23]. These two functionally similar genes were clustered together in our study. In addition, Mazie\_YSL\_1, the first experimentally validated OPT gene responsible for transport of Fe(III)-phytosiderophore chelates, was clustered together with ORYSJ\_YSL\_15 and ORYSJ\_YSL\_2 (Figure 2). ORYSJ\_YSL\_15 has been suggested to be responsible for iron uptake from rhizosphere and for phloem transport of iron by transporting Fe(III)-phytosiderophore chelates while ORYSJ\_YSL\_2 has been suggested to be responsible for phloem transport of iron by transporting Fe(III)-nicotianamine chelates [38,39]. Furthermore, ARATH\_YSL\_2 and ARATH\_YSL\_3 clustered together in Group 31 were both reported to be involved in transport of nicotianamine-chelated metals in the vasculature [40,41]. These results supported the idea that genes with the same functions were closely related. Based on the hypothesis,

it would be interesting to test if PG\_YSL\_1 is involved in iron-transportation since it was clustered together with ARATH\_YSL\_1 that was found to be involved in transport of iron-nicotianamine chelates [41,42]. Similarly, it would be interesting to test if PG\_OPT\_1 is akin to ARATH\_OPT\_6, which was reported to be involved in the transport of glutathione derivatives and metal complexes [21,43,44] and to test whether PG\_OPT\_9,10 and PG\_OPT\_11 are involved in increasing plant sensitivity to Cd like ARATH\_OPT\_7 functions [43].

We identified a total of 45 pairs of paralogs from the phylogenetic analyses (Table S2), which accounted for 11.1% to 70.6% of all OPT candidates in each studied species and shared similar structures within each group (Figure S1). We found that some OPT genes in ginseng were tandemly clustered on the same scaffold (Table S3) and those genes were location-related. For example, PG\_OPT\_10 and PG\_OPT\_11 were neighbor paralogs with 1122 bp in between. These genes might be formed by tandemly segmental duplication. PG\_YSL\_8 and PG\_YSL\_10 constitute a special tandemly clustered paralogs with a 3214 bp-long shared region. This paralogs pair might be generated by a crossover of chromosome after whole-genome duplication or by gene fusion. It would be interesting to test whether this OPT cluster was functional in further studies. In addition, PG\_YSL\_4-PG\_YSL\_5 and PG\_YSL\_14-PG\_YSL\_16 formed a special type of gene cluster block in which PG\_YSL\_4-PG\_YSL\_14 and PG\_YSL\_5-PG\_YSL\_16 were identified as paralogs oriented in the same direction. PG\_YSL\_18-PG\_YSL\_19 and PG\_YSL\_21-PG\_YSL\_22 constituted another special type of block, in which PG\_YSL\_18-PG\_YSL\_21 and PG\_YSL\_19-PG\_YSL\_22 were paralogs oriented in opposite directions (Figure 3). From this section of the study, we speculated that both types of cluster blocks were generated from segmental duplication or whole-genome duplication. Since *P. ginseng* is a tetraploid plant, we prefer to believe that genes from those blocks were more likely to be generated by whole-genome duplication. The paralogs blocks arranged in the opposite direction were likely to be generated by subsequent segmental inversion of the chromosome after segmental duplication.



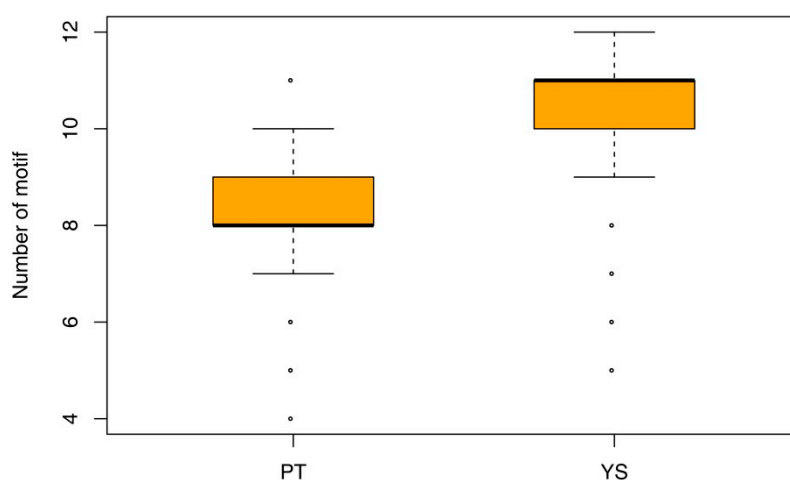
**Figure 3.** Chromosome locations for two special types of clusters of paralogs blocks. We used the ginseng genome version1 finished by Xu et al. [34] in this study. The scaffold refers to the DNA sequences in the ginseng genome that were generated by bridging non-gapped contigs (assembled with short gun sequencing reads) with mate-pair sequencing reads. A scaffold is equivalent to a chromosome segment.

$K_s$  (synonymous substitution rate) is a widely accepted concept for gene duplication time estimation. In general, the lower  $K_s$  is, the more recently gene duplication occurred [32]. Since a codon-based alignment of PG\_YSL\_6 and PG\_YSL\_17 failed to generate, calculation of  $K_s$  was excluded from this study (Table S2). Aligned sequences were nearly identical after removing gaps from

Potr\_YSL\_2/Potr\_YSL\_3 (poplar), Thec\_OPT\_7/Thec\_OPT\_8 (cacao), and Thec\_YSL\_4/Thec\_YSL\_5. The estimation of  $K_s$  for these paralogs also failed. Additionally,  $K_s$  values for PG\_OPT\_9/PG\_OPT\_10 and Sobi\_YSL\_13/Sobi\_YSL\_14 were estimated as 0, suggested that they were generated by a very recent duplication event. It was interesting to find that gene duplication of OPT paralogs occurred more recently in the YSL clade than in the PT clade in *P. ginseng*. The phenomenon was similar to grape and clover but contrary to cacao, cassava, and *Arabidopsis*. The duplication event for paralogs occurred more recently in ginseng, potato, poplar, cacao, grape, clover, and sorghum (about 0 to 5 MYA) than in carrot, cassava, *Arabidopsis*, and rice, which indicates that ginseng and other species or their common ancestor might have suffered a high level of gene loss during evolution because of the lack of an older duplication event such as 94.2 MYA for ARATH\_OPT\_6/ARATH\_OPT\_9 [45].

#### 2.4. Conserved Domains and Motif Analysis for OPT Genes Identified in *P. ginseng* and 11 Other Flowering Plants

By searching against the Conserved Domain Database (CDD) [46] with 278 OPT genes, all genes were annotated as OPT genes. However, only 267 were predicted to have specific domains, wherein all the 258 OPT genes used in the phylogeny analysis were covered. Because domain analysis could not provide information about smaller individual motifs and more divergent patterns, we conducted a study of motif analysis with MEME software (Supplementary File 2). As a result, 30 distinct motifs were identified in these genes. Detailed information of those motifs is presented in Supplementary File 3. It is interesting that the motif composition of OPT members in the PT clade is distinct from that in the YSL clade (Figure S1), which was in accordance with the conclusions generated by phylogenetic analysis. In addition, the number of motifs of OPT genes from the PT clade (ranging from 4 to 11, with a median value of 8) was distinct from that of the YSL clade (ranging from 5 to 12, with a median value of 11), which suggests the clade-specific structure of each OPT gene (Figure 4). Furthermore, we found nine motifs (Motif\_1,3,6,13,14,19,23,15,29) unique to the PT clade and 10 motifs (Motif\_7,8,12,16,17,21,22,24,26,30) unique to the YSL clade, respectively. Six motifs (Motif\_10,14,19,23,102,106) were frequently shared by PT clade members (94.4%, 94.4%, 86.9%, 91.6%, 95.3%, and 99.1%, respectively) and 10 motifs (Motif\_2,8,9,15,16,18,21,22,26,28) were frequently shared by the YSL clade members (92.0%, 88.7%, 71.3%, 38.7%, 92.0%, 92.0%, 92.7%, 81.3%, 84.7%, 97.3%, and 94.0%, respectively) (Table S4). Those findings might give us new insights into how OPT genes evolved since being separated from their common ancestor and how they functionally diverged during the subsequent evolution process.

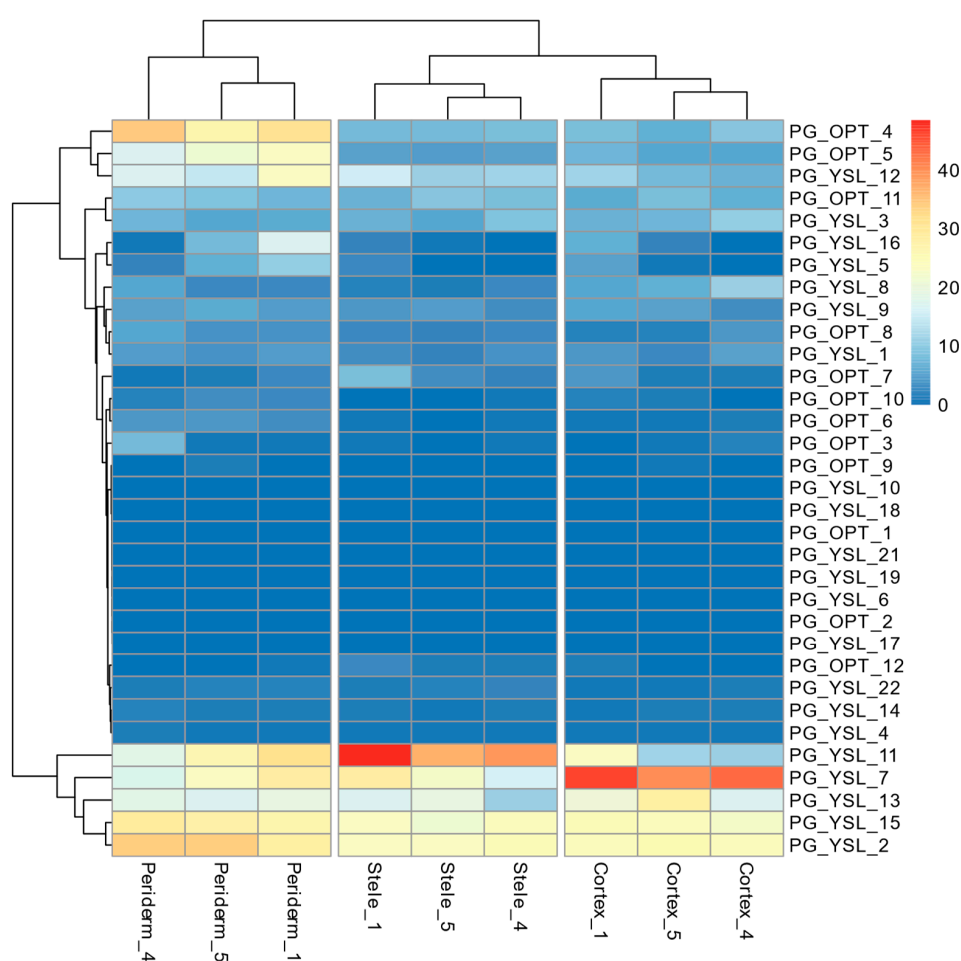


**Figure 4.** Number of motifs in OPT genes from PT and YS clades. These static results were calculated with xml output of MEME analysis (Supplementary File 5) by our custom R scripts. The boxplot was generated by the built-in function “boxplot” in R.



### 2.5. Profiling of Expression Patterns for OPT Genes Identified in *P. ginseng*

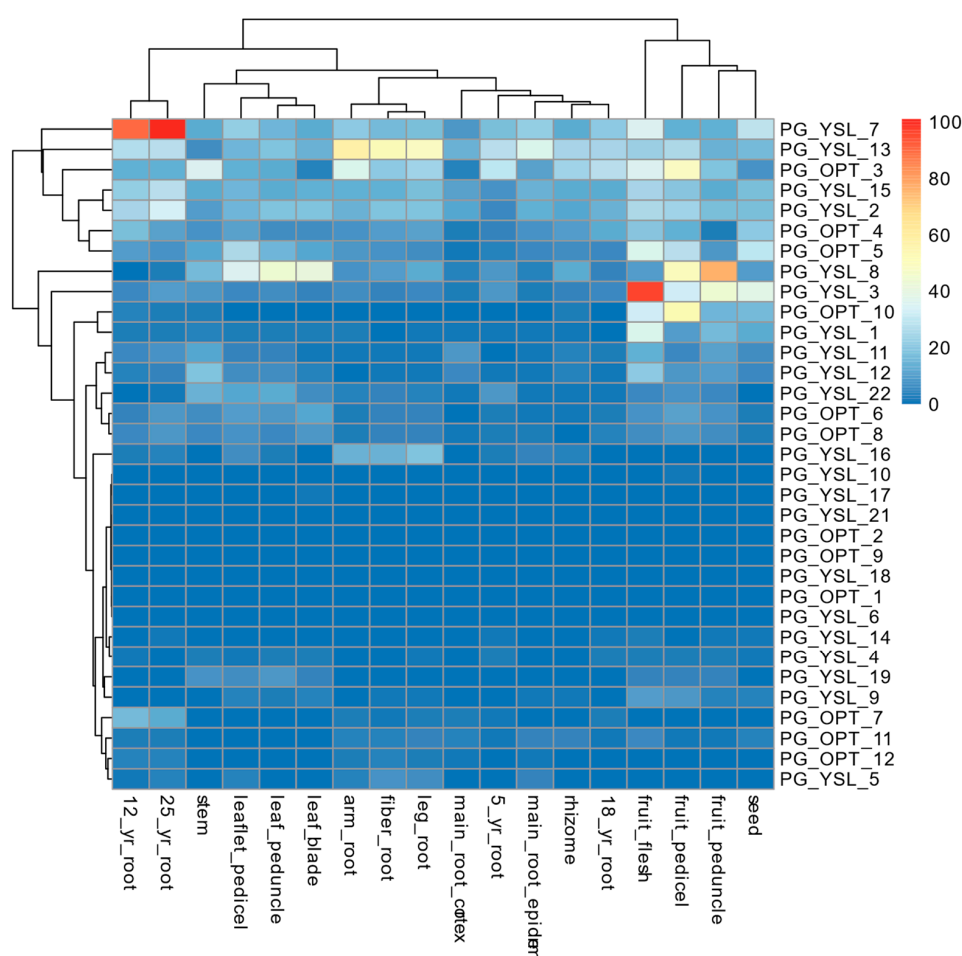
In order to examine the expression patterns of the OPT genes in *P. ginseng*, we performed a comprehensive expression analysis by using two sets of RNA-Seq datasets: one from our previous study about *P. ginseng* root [47] and one from a public study about 18 kinds of tissues. In general, genes in the YSL clade were more highly expressed than genes in the PT clade except in the periderm (Figure 5). PG\_YSL\_2,13 and PG\_YSL\_15 were expressed evenly in the root with little difference among tissues, which suggests that they might be constitutive OPTs. OPT genes exhibited distinct tissue-specific expression manners. For example, PG\_OPT\_4,5 and PG\_YSL\_12 were more likely to be expressed highly in periderm than in the stele or cortex. PG\_YSL\_11 and PG\_YSL\_7 had the highest expression in the stele and cortex, respectively, while they were still expressed at a considerably high level in other tissues. The different expression patterns for those OPT genes indicated that a wide range of substrates might be transported in different parts of the plant root.



**Figure 5.** Expression of OPT genes of different tissues in ginseng root. Gene expression of OPT genes was calculated with RNA-Seq data generated by our previous study on *P. ginseng* root. The hierarchical clustered heat map was plotted with ‘pheatmap’ planted in R package named ‘pheatmap’.

Due to the nature of sink tissue of fruit and seeds in plants, the expression characteristics of OPT genes of these tissues are expected to share more common traits than those of other tissues. Based on the expression data, fruit flesh, fruit pedicel, fruit peduncle, and seeds were clustered together, which suggests that the similar expression pattern of those OPT genes might contribute to methods of metabolism relocation. The expression of PG\_YSL\_1,3 and PG\_YSL\_7 both peaked in fruit flesh compared with the fruit pedicel, fruit peduncle, and seed, which indicates that lateral transportation might be the most active transportation process during seed development. Additionally, the leaf

blade, leaf pedicel, leaf peduncle, and stem, which are physically connected organs forming a complex vascular transportation system in plants, were clustered together by their similar expression pattern, wherein PG\_YSL\_8 was expressed at the highest level (except for the stem). Moreover, the arm root, fiber root, and leg root were clustered together, and PG\_OPT\_13 and PG\_YSL\_16 were highly expressed. It was interesting that PG\_YSL\_7 was highly expressed in 12-year-old and 25-year-old roots but minimally expressed in five-year-old and 18-year-old roots that were clustered with the main root cortex, the main root epiderm, and the rhizome. Taken together, the expression patterns found in our study and Wang's [48] both suggested that OPT genes were expressed in tissue-specific and location-specific manners by which the transportation and distribution of oligopeptides and their conjugates with metals, signals, etc. were shaped in different ginseng tissues [41] (Figure 6).



**Figure 6.** Expression of the OPT genes in 18 organs of ginseng. Gene expression of OPT genes was re-calculated with RNA-seq data from public research by taking our published genome as a reference.

Based on the phylogenetic analysis described above, PG\_OPT\_4 and PG\_OPT\_5 were grouped with ARATH\_OPT\_1 and ARATH\_OPT\_5, which were proven to be OPT transporters for penta-peptide (KLLLG) in an energy-dependent manner by yeast complementation assay [20,49]. High expression of those genes exclusively in the root suggested that the penta-peptide-related metabolism (metabolism substrates, signal molecules, etc.) transportation might be activated. PG\_YSL\_12 was identified as another periderm-specific expressed gene found in this study. It was expressed more highly in the stem and fruit flesh than in other tissues (Figure 6). PG\_YSL\_12 was clustered with ARATH\_YSL\_5 and ARATH\_YSL\_8 into one group, which indicates that it might be involved in the transport of nicotianamine-chelated metals (metals-NA) just as ARATH\_YSL\_2 was in the transport of Fe-NA across the plasma membrane in leaf cells, involving lateral movement of iron away from

the xylem [40]. Furthermore, ARATH\_YSL\_8 might also be involved directly in iron uptake by leaf cells [13].

ARATH\_YSL\_1 and ARATH\_YSL\_3 were experimentally verified OPT proteins, which were found to be able to mediate Fe transportation to and from vascular tissues [41]. ARATH\_YSL\_3 was a sister branch to ARATH\_YSL\_2, which had been functionally confirmed to be expressed in both roots and shoots and to mediate transport of metal-NA complexes [40], which indicates their functional similarity. ARATH\_YSL\_1 was clustered into another sister group to a larger group including ARATH\_YSL\_2 and ARATH\_YSL\_3, which suggests that members of these two larger group might share some functional similarities. MAIZE\_YSL\_1 (*ZmYSL1*) known as a proton-coupled symporter transports iron complexed by plant-derived Fe(III) chelators (phytosiderophores, PS) by scavenging from soil, termed Strategy II [13]. It formed another cluster in the YSL clade with some YSL genes from sorghum and rice with a bootstrap value of 97% (Figure 2). Considering the functional similarity of ARATH\_YSL\_1,2,3 and *ZmYSL1*, OPT genes from groups 27 and 28 and groups 29–31 were suggested to form two sister groups that might be involved in the transport of Fe.

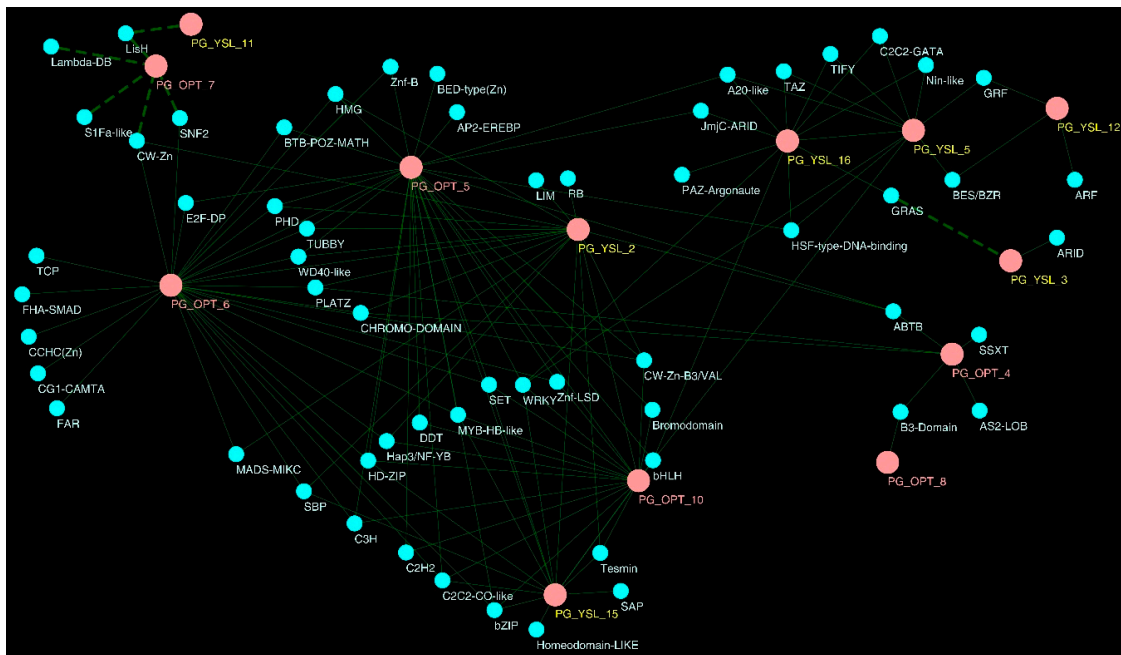
The OPT paralogs were more likely to be generated by segmental tandem duplication rather than transposition [32]. The expression pattern of those duplicated genes may differ if they suffered evolutionary divergence such as neofunctionalization. [50]. No similar expression patterns of duplicated paralogs were identified in the study about poplar and grape [32]. However, we detected five similarly expressed paralogs pairs in this study wherein PG\_OPT\_4-PG\_OPT\_5 had similar expression patterns both in our previous study and in Wang's study. PG\_YSL\_4-PG\_YSL\_14 and PG\_YSL\_18-PG\_YSL\_21 were expressed similarly but with a very low expression level. PG\_YSL\_6-PG\_YSL\_8 and PG\_YSL\_11-PG\_YSL\_12 were reported to be expressed similarly in Wang's report but not in ours. However, PG\_YSL\_5/PG\_YSL\_16 was similarly expressed in our study but not in Wang's study. The similar expression patterns found in our study might be because of the relatively short time has been experienced in *Ginseng* paralogs compared with those paralogs in poplar and grape. On the other hand, the phenomenon that a majority of the identified paralogs in *Ginseng* did not have similar expression patterns, which indicates functional diversification might be a result of long-term evolution—adapting to changing environmental conditions after gene duplication.

## 2.6. Analysis of Co-Expression Network between OPT Genes and Potent Transcription factor for *P. ginseng*

The regulation of gene expression in all living cells is dominated by transcriptional initiation, which is regulated by transcription factors, ancillary transcription regulators, and chromatin regulators. Therefore, we conducted an analysis focusing on the co-expression between all transcriptionally modulated genes in the ginseng genome and all transcription factors in order to reveal regulators for OPT genes in *Ginseng*. PlantTFcat is a useful tool for identifying proteins with signature domains specific to 108 major transcription regulators families [51]. We assessed the *Ginseng* genome for identifying those proteins. A total of 5073 distinct genes in the *P. ginseng* genome have been predicted to be transcription factors wherein there are 5457 members (Supplementary File 4). Genes annotated as different transcription factors by PlantTFcat (such as PG39956, annotated as Znf-B, LisH, WD40-like, or PLATZ) were removed from further analysis. The expression values of transcription factors that were mapped by many genes (such as MYB-HB-like, mapped by 334 genes) were determined by the median values of those genes. Lastly, a total of 59 transcription factors and 13 OPT genes were used for network analysis. We used non-parametric Spearman's rank-order correlation for our co-expression analysis due to its robustness for generating biologically relevant gene networks [52].

The matrix of all correlation values for expression values between each pair of transcription factor and OPT gene from a set of nine biological samples is shown in Table S5. At a conservative threshold of  $\rho \geq |0.85|$ , 141 positive and 11 negative correlations involving 13 OPT genes were found (Table S6). The number of transcription factors correlated to an OPT gene ranged from 1 to 27, while it ranged only 1 to 5 for the number of OPT genes correlated with a transcription factor (Tables S7 and S8, Figure 7). For example, PG\_OPT\_5 and PG\_OPT\_6 positively correlated with 26 and 27 transcription factors,

respectively, while *bHLH* and *WRKY* only correlated with five OPT genes. Our findings suggested that the initiation of transcription of OPT genes might be dominated by a complicated synergetic regulation system consisting of a number of transcription factors. Additionally, transcription factors might act as pleiotropic regulators participating in a variety of transcription regulations for OPT genes. On the other hand, because 19 out of 59 transcription factors were linked to only one OPT gene and two out of 13 OPT genes were linked to only one transcription factor, the results suggested that the transcription of some OPT genes was regulated by specific transcription factors and some transcription factors had specific target genes to regulate.



**Figure 7.** Regulatory gene networks involving transcription master genes and OPT genes. A stringent threshold ( $\rho \geq |0.85|$ ) was used and the visualization was produced in Cytoscape-3.6.1. Nodes for OPT genes are represented by pink circles (yellow labels represent OPT genes in the YSL clade. Pink labels represent OPT genes in PT clade). Nodes for blue circles represent transcriptional regulators. Positive interactions are indicated by green lines and negative interactions are indicated by green dashed lines.

### 3. Materials and Methods

#### 3.1. Sequence Retrieval and Identification of OPT Genes

We identified OPT genes from *P. ginseng* and 11 other flowering plants by using TransportTP (<http://bioinfo3.noble.org/transporter/> [35]). Proteome sequences from 11 flowering plants (*Arabidopsis* TAIR10, rice v7, tomato iTAG2.4, potato v4.03, carrot v2.0, *Manihot esculenta* v6.1, *Medicago truncatula* Mt4.0v1, Poplar v3.1, Grape Genoscope 12X, Cacao v1.1, Sorghum v3.1.1) were retrieved from the Phytozome database (V12.1) [53], wherein the assembly version is followed by each species name. Ginseng proteome sequences were retrieved from <http://ginseng.vicp.io:23488/index.php/index/download.html>. The proteome sequences for each species were then used for the identification of OPT genes by searching against TransporterTP, setting the E-value threshold to 0.1, and setting *Arabidopsis thaliana* and *Oryza sativa* as the reference organisms.

Subcellular localization of those OPT proteins was predicted with WOLF-PSORT [54,55]. Isoelectric point (pI), molecular weight, and grand average hydropathicity (GRAVY) values were estimated with functions planted in the Peptides package (<https://github.com/dosorio/Peptides/>) for R.



### 3.2. Phylogenetic Analysis for OPT Genes

Phylogenetic analysis of these OPT genes was conducted on their conserved domains identified by CDD (Conserved Domain Database, [56,57]) and planted in NCBI with default parameters (50369 PSSMs, e-value of 0.01, maximum number of hits 500). Multiple sequence alignments of those conserved OPT protein were performed with MAFFT v7.158b [58] and followed by manual comparison and refinement. Aligned regions that contain over 50% gaps or ambiguous sites were removed and sequences that contained gaps for more than 50% of the remaining sequence were deleted. Lastly, 258 OPT genes were left for further phylogenetic analysis. In order to select the best evolutionary model for phylogeny reconstruction, we used the function 'modelTest' planted in the R package named 'phangorn' [59] with the parameter 'model' set to 'all' and found 'LG+G+F' was the best model. A maximum likelihood method of phylogenetic analysis based on RAxML v 8.2.9 [60] software was conducted, with the parameter "--bootstop-perms" set to 1000, "--m" set to PROTGAMMALG, and the "--e" set to 0.001. After finishing the reconstruction of the phylogeny of OPT genes for these species, the topology was plotted by the online tool iTOL [61].

### 3.3. Estimation of Duplication Time for OPT Paralogs

Pairwise alignment of protein sequences of the OPT paralogs was aligned with MAFFT software, and codon-based pairwise alignment of nucleotide sequences were generated by using PAL2NAL [62]. The  $K_a$  and  $K_s$  values for paralogous genes were estimated by the program yn00 planted in the PAML package with default parameters [63]. Assuming a molecular clock, the synonymous substitution rates ( $K_s$ ) of the paralogous genes could be regarded as a proxy for time estimation of the segmental duplication events. The approximation of date for duplication events was estimated with the following formula:  $T = K_s/2\lambda$ , where  $\lambda$  denotes clock-like rates of synonymous substitution. In this study,  $1.5 \times 10^{-8}$  substitutions/synonymous site/year was used for *Arabidopsis*,  $6.5 \times 10^{-9}$  for rice, sorghum, cassava, grape and cacao,  $9.1 \times 10^{-9}$  for poplar [32],  $1.08 \times 10^{-8}$  for clover,  $6.68 \times 10^{-9}$  for *P. ginseng*,  $2.69 \times 10^{-9}$  for potato, and  $2.91 \times 10^{-9}$  for carrot.  $\lambda$  for each species was deduced or collected from previous studies [64,65].

### 3.4. Analysis of Motif Composition for OPT Genes

Conserved motif analysis for OPT genes in the *P. ginseng* genome was conducted with MEME (<http://meme.sdsc.edu>). The OPT candidates were run locally with MEME with the following parameters: number of repetitions = any, maximum number of motifs = 30. The other parameters were kept as default values.

### 3.5. Profiling Expression of OPT Genes for *P. ginseng*

The gene expression of *P. ginseng* was profiled by RNA-Seq datasets from our previous study [47] and public research [48]. Those datasets could be retrieved from the SRA database by searching BioProject id PRJNA369187 and PRJNA302556. These raw datasets from the SRA database were first converted into FASTQ files by sratoolkit.2.8.0 [66] and then quality controlled by Trimmomatic-0.36 [67]. Lastly, reference-based gene expression of those biological samples was estimated with the HISAT2+StringTie pipeline [68]. FPKM values for each gene were used as gene expression levels. A hierarchical clustered heatmap for OPT genes was plotted with the pheatmap package [69], wherein "manhattan" distance was used for both row-based and column-based clustering.

### 3.6. Identification of Regulatory Network between OPT Genes and Transcription Factors for *P. ginseng*

The *P. ginseng* proteome sequence dataset was submitted to the PlantTFcat analysis tool (<http://plantgrn.noble.org/PlantTFcat/> [51]) for the identification and classification of transcription factors, chromatin modifiers, and other transcriptional regulators into protein families. Genes that mapped to more than one transcription factors were removed from further analysis. In addition, median

values of those genes referring to the same transcription factors were regarded as the transcription factors' expression value. In this study, we used our previous RNA-Seq dataset for construction of the co-expression network between OPT genes and transcription factors. FPKMs of all genes including OPT genes and transcription factors were combined and used for the calculation of Spearman's rank correlation coefficient to predict potential gene regulatory networks. The correlation coefficient ( $\rho$ ) for each gene pair was calculated by the built-in function "cor" in R, and a threshold of  $\rho \geq |0.85|$  was regarded as significant co-expression. Visualization of the network was created in Cytoscape 3.6.1 [70].

#### 4. Conclusions

This study is the first to investigate the chromosomal location, expression profiling, and transcriptional regulation networks of *P. ginseng* OPT genes and provide a comparative genome analysis addressing the phylogeny, gene structure, and paralogs duplication history of the OPT gene family in *P. ginseng* and 11 flowering plants. Chromosomal location analyses revealed that structural variation occurred after segmental duplication, expression profiling, and transcriptional co-expression networks analyses, which indicates that both specific and pleiotropic transcription regulators might be involved in the regulation of OPT genes' expression. Phylogenetic analyses suggested two well-supported clades in the OPT family, which can be further classified into 12 or 19 distinct groups. Motif compositions are conserved in each clade and clade-specific motifs were frequently occupied within each clade. Estimations for paralogs divergence history indicated that the majority of OPT paralogs in *P. ginseng* might have emerged from recent duplications, which was different from the history of *Arabidopsis* or cassava. The study of expression profiles in different organs and tissues of *P. ginseng* has provided insights into possible functional divergence among OPT members and important functional roles in the plant development of some OPT members. These data may provide valuable information for future functional investigations of this gene family.

**Supplementary Materials:** See the word file of "The list of supplementary materials." All supplementary materials are available online.

**Author Contributions:** Z.H. conceived and designed the research framework. H.C., J.X., and Y.C. prepared the sample and performed the experiments. J.X. and Y.C. provided many important suggestions for data analysis. H.S. analyzed the data. H.S. and J.X. wrote the manuscript. J.B., L.G., J.H., W.X., J.Z., X.Q., and Z.H. make revisions to the final manuscript. All authors have read and approved the final manuscript.

**Funding:** This work was supported by grants from several funds supported by Guangdong Forestry Department, Guangdong Food and Drug Administration and Guangdong Provincial Bureau of traditional Chinese Medicine (2017KT1835, 2018KT1050, 2018TDZ16, 2018KT1138, 2018KT1228, and 2018KT1230), National Nature Science Foundation of China (81803672), standardized research and application of precise powder decoction pieces in traditional Chinese Medicine, and Construction Project of TCM Hospital Preparation by Special Fund of Strong Province Construction in TCM, Guangdong, China (No. 6).

**Conflicts of Interest:** The authors declare no conflict of interest.

#### References

1. Stacey, M.G.; Koh, S.; Becker, J.; Stacey, G. Atopt3, a member of the oligopeptide transporter family, is essential for embryo development in *Arabidopsis*. *Plant Cell* **2002**, *14*, 2799–2811. [[CrossRef](#)] [[PubMed](#)]
2. Perry, J.R.; Basrai, M.A.; Steiner, H.Y.; Naider, F.; Becker, J.M. Isolation and characterization of a *Saccharomyces cerevisiae* peptide transport gene. *Mol. Cell. Biol.* **1994**, *14*, 104–115. [[CrossRef](#)] [[PubMed](#)]
3. Steiner, H.Y.; Naider, F.; Becker, J.M. The ptr family: A new group of peptide transporters. *Mol. Microbiol.* **1995**, *16*, 825–834. [[CrossRef](#)] [[PubMed](#)]
4. Swift, S.; Throup, J.P.; Williams, P.; Salmond, G.P.; Stewart, G.S. Quorum sensing: A population-density component in the determination of bacterial phenotype. *Trends Biochem. Sci.* **1996**, *21*, 214–219. [[CrossRef](#)]
5. Kuchler, K.; Sterne, R.E.; Thorner, J. *Saccharomyces cerevisiae* ste6 gene product: A novel pathway for protein export in eukaryotic cells. *Embo J.* **1989**, *8*, 3973–3984. [[CrossRef](#)] [[PubMed](#)]
6. Curie, C.; Panaviene, Z.; Loulergue, C.; Dellaporta, S.L.; Briat, J.-F.; Walker, E.L. Maize yellow stripe1 encodes a membrane protein directly involved in Fe(III) uptake. *Nature* **2001**, *409*, 346–349. [[CrossRef](#)] [[PubMed](#)]

7. Stacey, M.G.; Patel, A.; McClain, W.E.; Mathieu, M.; Remley, M.; Rogers, E.E.; Gassmann, W.; Blevins, D.G.; Stacey, G. The Arabidopsis atopt3 protein functions in metal homeostasis and movement of iron to developing seeds. *Plant Physiol.* **2008**, *146*, 589–601. [[CrossRef](#)] [[PubMed](#)]
8. Dean, M.; Hamon, Y.; Chimini, G. The human atp-binding cassette (abc) transporter superfamily. *J. Lipid Res.* **2001**, *42*, 1007–1017. [[CrossRef](#)] [[PubMed](#)]
9. Higgins, C.F. Abc transporters: From microorganisms to man. *Ann. Rev. Cell Biol.* **1992**, *8*, 67–113. [[CrossRef](#)]
10. Hauser, M.; Narita, V.; Donhardt, A.M.; Naider, F.; Becker, J.M. Multiplicity and regulation of genes encoding peptide transporters in *saccharomyces cerevisiae*. *Mol. Membr. Biol.* **2001**, *18*, 105–112. [[CrossRef](#)]
11. Newstead, S. Recent advances in understanding proton coupled peptide transport via the pot family. *Curr. Opin. Struct. Biol.* **2017**, *45*, 17–24. [[CrossRef](#)] [[PubMed](#)]
12. Williams, L.; Miller, A. Transporters responsible for the uptake and partitioning of nitrogenous solutes. *Ann. Rev. Plant Biol.* **2001**, *52*, 659–688. [[CrossRef](#)] [[PubMed](#)]
13. Lubkowitz, M. The opt family functions in long-distance peptide and metal transport in plants. In *Genetic Engineering: Principles and Methods*; Setlow, J.K., Ed.; Springer: Boston, MA, USA, 2006; pp. 35–55.
14. Lubkowitz, M.A.; Hauser, L.; Breslav, M.; Naider, F.; Becker, J.M. An oligopeptide transport gene from *candida albicans*. *Microbiology* **1997**, *143*, 387–396. [[CrossRef](#)] [[PubMed](#)]
15. Lubkowitz, M.A.; Barnes, D.; Breslav, M.; Burchfield, A.; Naider, F.; Becker, J.M. *Schizosaccharomyces pombe* isp4 encodes a transporter representing a novel family of oligopeptide transporters. *Mol. Microbiol.* **1998**, *28*, 729–741. [[CrossRef](#)] [[PubMed](#)]
16. Lubkowitz, M. The oligopeptide transporters: A small gene family with a diverse group of substrates and functions? *Mol. Plant* **2011**, *4*, 407–415. [[CrossRef](#)] [[PubMed](#)]
17. Gomolplitinant, K.M.; Saier, M., Jr. Evolution of the oligopeptide transporter family. *J. Membr. Biol.* **2011**, *240*, 89. [[CrossRef](#)] [[PubMed](#)]
18. Feng, S.; Tan, J.; Zhang, Y.; Liang, S.; Xiang, S.; Wang, H.; Chai, T. Isolation and characterization of a novel cadmium-regulated yellow stripe-like transporter (snysl3) in *solanum nigrum*. *Plant Cell Rep.* **2017**, *36*, 281–296. [[CrossRef](#)] [[PubMed](#)]
19. Murata, Y.; Ma, J.F.; Yamaji, N.; Ueno, D.; Nomoto, K.; Iwashita, T. A specific transporter for iron(III)-phytosiderophore in barley roots. *Plant J.* **2006**, *46*, 563–572. [[CrossRef](#)] [[PubMed](#)]
20. Koh, S.; Wiles, A.M.; Sharp, J.S.; Naider, F.R.; Becker, J.M.; Stacey, G. An oligopeptide transporter gene family in *arabidopsis*. *Plant Physiol.* **2002**, *128*, 21–29. [[CrossRef](#)]
21. Wongkaew, A.; Asayama, K.; Kitaiwa, T.; Nakamura, S.-I.; Kojima, K.; Stacey, G.; Sekimoto, H.; Yokoyama, T.; Ohkama-Ohtsu, N. Atopt6 protein functions in long-distance transport of glutathione in *arabidopsis thaliana*. *Plant Cell Physiol.* **2018**. [[CrossRef](#)]
22. Zhang, Z.; Xie, Q.; Jobe, T.O.; Kau, A.R.; Wang, C.; Li, Y.; Qiu, B.; Wang, Q.; Mendoza-Cózatl, D.G.; Schroeder, J.I. Identification of atopt4 as a plant glutathione transporter. *Mol. Plant* **2016**, *9*, 481–484. [[CrossRef](#)] [[PubMed](#)]
23. Mendoza-Cózatl, D.G.; Xie, Q.; Akmakjian, G.Z.; Jobe, T.O.; Patel, A.; Stacey, M.G.; Song, L.; Demoin, D.W.; Jurisson, S.S.; Stacey, G. Opt3 is a component of the iron-signaling network between leaves and roots and misregulation of opt3 leads to an over-accumulation of cadmium in seeds. *Mol. Plant* **2014**, *7*, 1455–1469. [[CrossRef](#)] [[PubMed](#)]
24. Vasconcelos, M.W.; Li, G.W.; Lubkowitz, M.A.; Grusak, M.A. Characterization of the pt clade of oligopeptide transporters in rice. *Plant Genome* **2008**, *1*, 77–88. [[CrossRef](#)]
25. Bogs, J.; Bourbonloux, A.; Cagnac, O.; Wachter, A.; Rausch, T.; Delrot, S. Functional characterization and expression analysis of a glutathione transporter, bjgt1, from *brassica juncea*: Evidence for regulation by heavy metal exposure. *Plant Cell Environ.* **2003**, *26*, 1703–1711. [[CrossRef](#)]
26. Carole, D.M.; Beno, T.P. Role of glutathione in plant signaling under biotic stress. *Plant Signal. Behav.* **2012**, *7*, 210–212.
27. Yen, M.-R.; Tseng, Y.-H.; Saie, M., Jr. Maize yellow stripe1, an iron-phytosiderophore uptake transporter, is a member of the oligopeptide transporter (opt) family. *Microbiology* **2001**, *147*, 2881–2883. [[CrossRef](#)] [[PubMed](#)]
28. Cobbett, C.; Goldsbrough, P. Phytochelatins and metallothioneins: Roles in heavy metal detoxification and homeostasis. *Annu. Rev. Plant Biol.* **2003**, *53*, 159–182. [[CrossRef](#)]
29. Bourbonloux, A.; Shahi, P.; Chakladar, A.; Delrot, S.; Bachhawat, A.K. Hgt1p, a high affinity glutathione transporter from the yeast *saccharomyces cerevisiae*. *J. Biol. Chem.* **2000**, *275*, 13259–13265. [[CrossRef](#)]

30. Liu, T.; Zeng, J.; Xia, K.; Fan, T.; Li, Y.; Wang, Y.; Xu, X.; Zhang, M. Evolutionary expansion and functional diversification of oligopeptide transporter gene family in rice. *Rice* **2012**, *5*, 1–14. [[CrossRef](#)]
31. Xiang, Q.; Shen, K.; Yu, X.; Zhao, K.; Gu, Y.; Zhang, X.; Chen, X.; Chen, Q. Analysis of the oligopeptide transporter gene family in ganoderma lucidum: Structure, phylogeny, and expression patterns. *Genome* **2017**, *60*, 293–302. [[CrossRef](#)]
32. Cao, J.; Huang, J.; Yang, Y.; Hu, X. Analyses of the oligopeptide transporter gene family in poplar and grape. *BMC Genom.* **2011**, *12*, 465. [[CrossRef](#)] [[PubMed](#)]
33. Yun, T.K. Brief introduction of panax ginseng c.A. Meyer. *J. Korean Med. Sci.* **2001**, *16* (Suppl.), S3–S5.
34. Jiang, X.; Yang, C.; Baosheng, L.; Shuiming, X.; Qinggang, Y.; Rui, B.; He, S.; Linlin, D.; Xiwen, L.; Jun, Q. *Panax ginseng* genome examination for ginsenoside biosynthesis. *GigaScience* **2017**, *6*, 1–15.
35. Li, H.; Benedito, V.A.; Udvardi, M.K.; Zhao, P.X. Transporttp: A two-phase classification approach for membrane transporter prediction and characterization. *BMC Bioinform.* **2009**, *10*, 418. [[CrossRef](#)] [[PubMed](#)]
36. Li, W.; Godzik, A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **2006**, *22*, 1658–1659. [[CrossRef](#)] [[PubMed](#)]
37. Kyte, J.; Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, *157*, 105–132. [[CrossRef](#)]
38. Shintaro, K.; Haruhiko, I.; Daichi, M.; Michiko, T.; Hiromi, N.; Satoshi, M.; Nishizawa, N.K. Osysl2 is a rice metal-nicotianamine transporter that is regulated by iron and expressed in the phloem. *Plant J.* **2010**, *39*, 415–424.
39. Haruhiko, I.; Takanori, K.; Tomoko, N.; Michiko, T.; Yusuke, K.; Kazumasa, S.; Mikio, N.; Hiromi, N.; Satoshi, M.; Nishizawa, N.K. Rice osysl15 is an iron-regulated iron(III)-deoxymugineic acid transporter expressed in the roots and is essential for iron uptake in early growth of the seedlings. *J. Biol. Chem.* **2009**, *284*, 3470–3479.
40. DiDonido, D., Jr.; Roberts, L.A.; Sanderson, T.; Eisley, R.B.; Walker, E.L. Arabidopsis yellow stripe-like2 (ysl2): A metal-regulated gene encoding a plasma membrane transporter of nicotianamine–metal complexes. *Plant J.* **2004**, *39*, 403–414. [[CrossRef](#)]
41. Waters, B.M.; Chu, H.H.; Didonato, R.J.; Roberts, L.A.; Eisley, R.B.; Lahner, B.; Salt, D.E.; Walker, E.L. Mutations in arabidopsis yellow stripe-like1 and yellow stripe-like3 reveal their roles in metal ion homeostasis and loading of metal ions in seeds. *Plant Physiol.* **2006**, *141*, 1446–1458. [[CrossRef](#)]
42. Marie, L.J.; Adam, S.; Stéphane, M.; Jean-François, B.; Catherine, C. A loss-of-function mutation in atysl1 reveals its role in iron and nicotianamine seed loading. *Plant J.* **2010**, *44*, 769–782.
43. Cagnac, O.; Bourbouloux, A.; Chakrabarty, D.; Zhang, M.-Y.; Delrot, S. Atopt6 transports glutathione derivatives and is induced by primisulfuron. *Plant Physiol.* **2004**, *135*, 1378–1387. [[CrossRef](#)] [[PubMed](#)]
44. Pike, S.; Patel, A.; Stacey, G.; Gassmann, W. Arabidopsis opt6 is an oligopeptide transporter with exceptionally broad substrate specificity. *Plant Cell Physiol.* **2009**, *50*, 1923–1932. [[CrossRef](#)] [[PubMed](#)]
45. Jaillon, O.; Aury, J.M.; Noel, B.; Policriti, A.; Clepet, C.; Casagrande, A.; Choisne, N.; Aubourg, S.; Vitulo, N.; Jubin, C. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **2007**, *449*, 463–467. [[PubMed](#)]
46. Marchlerbauer, A.; Bo, Y.; Han, L.; He, J.; Lanczycki, C.J.; Lu, S.; Chitsaz, F.; Derbyshire, M.K.; Geer, R.C.; Gonzales, N.R. Cdd/sparcle: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* **2017**, *45*, D200–D203. [[CrossRef](#)] [[PubMed](#)]
47. Zhang, J.J.; Su, H.; Zhang, L.; Liao, B.S.; Xiao, S.M.; Dong, L.L.; Hu, Z.G.; Wang, P.; Li, X.W.; Huang, Z.H. Comprehensive characterization for ginsenosides biosynthesis in ginseng root by integration analysis of chemical and transcriptome. *Molecules* **2017**, *22*, 889. [[CrossRef](#)]
48. Wang, K.; Jiang, S.; Sun, C.; Lin, Y.; Rui, Y.; Yi, W.; Zhang, M. The spatial and temporal transcriptomic landscapes of ginseng, panax ginseng c. A. Meyer. *Sci. Rep.* **2015**, *5*, 18283. [[CrossRef](#)]
49. Osawa, H.; Stacey, G.; Gassmann, W. Scept1 and atopt4 function as proton-coupled oligopeptide transporters with broad but distinct substrate specificities. *Biochem. J.* **2006**, *393*, 267–275. [[CrossRef](#)]
50. Prince, V.E.; Pickett, F.B. Splitting pairs: The diverging fates of duplicated genes. *Nat. Rev. Genet.* **2002**, *3*, 827–837. [[CrossRef](#)]
51. Dai, X.; Sinharoy, S.; Udvardi, M.; Zhao, P.X. Planttfcat: An online plant transcription factor and transcriptional regulator categorization and analysis tool. *BMC Bioinform.* **2013**, *14*, 321. [[CrossRef](#)]



52. Sapna, K.; Nie, J.; Chen, H.S.; Hao, M.; Ron, S.; Xiang, L.; Lu, M.Z.; Taylor, W.M.; Wei, H. Evaluation of gene association methods for coexpression network construction and biological knowledge discovery. *PLoS ONE* **2012**, *7*, e50411.
53. Goodstein, D.M.; Shu, S.; Russell, H.; Rochak, N.; Hayes, R.D.; Joni, F.; Therese, M.; William, D.; Uffe, H.; Nicholas, P. Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res.* **2012**, *40*, D1178–D1186. [[CrossRef](#)] [[PubMed](#)]
54. Yu, N.Y.; Wagner, J.R.; Laird, M.R.; Melli, G.; Lo, R.; Dao, P.; Sahinalp, S.C.; Ester, M.; Foster, L.J.; Brinkman, F.S.L. Psorb 3.0. *Bioinformatics* **2010**, *26*, 1608–1615. [[CrossRef](#)] [[PubMed](#)]
55. Horton, P.; Park, K.J.; Obayashi, T.; Fujita, N.; Harada, H.; Adamscollier, C.J.; Nakai, K. Wolf psort: Protein localization predictor. *Nucleic Acids Res.* **2007**, *35*, 585–587. [[CrossRef](#)] [[PubMed](#)]
56. Marchlerbauer, A.; Anderson, J.B.; Chitsaz, F.; Derbyshire, M.K.; Deweesescott, C.; Fong, J.H.; Geer, L.Y.; Geer, R.C.; Gonzales, N.R.; Gwadz, M. Cdd: Specific functional annotation with the conserved domain database. *Nucleic Acids Res.* **2009**, *37*, D205–D210. [[CrossRef](#)] [[PubMed](#)]
57. Marchler-Bauer, A.; Lu, S.; Anderson, J.B.; Chitsaz, F.; Derbyshire, M.K.; DeWeese-Scott, C.; Fong, J.H.; Geer, L.Y.; Geer, R.C.; Gonzales, N.R.; et al. Cdd: A conserved domain database for the functional annotation of proteins. *Nucleic Acids Res.* **2011**, *39*, D225–D229. [[CrossRef](#)]
58. Katoh, K.; Kuma, K.; Toh, H.; Miyata, T. Mafft version 5: Improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **2005**, *33*, 511–518. [[CrossRef](#)]
59. Schliep, K.P. Phangorn: Phylogenetic analysis in r. *Bioinformatics* **2011**, *27*, 592–593. [[CrossRef](#)]
60. Stamatakis, A. Raxml version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [[CrossRef](#)]
61. Letunic, I.; Bork, P. *Interactive Tree of Life (ItoI): An Online Tool for Phylogenetic Tree Display and Annotation*; Oxford University Press: Oxford, UK, 2007; pp. 78–82.
62. Suyama, M.; Torrents, D.; Bork, P. Pal2nal: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **2006**, *34*, W609–W612. [[CrossRef](#)]
63. Yang, Z. Paml: A program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **1997**, *13*, 555–556. [[CrossRef](#)] [[PubMed](#)]
64. The Potato Genome Sequencing Consortium; Xu, X.; Pan, S.; Cheng, S.; Zhang, B.; Mu, D.; Ni, P.; Zhang, G.; Yang, S.; Li, R.; et al. Genome sequence and analysis of the tuber crop potato. *Nature* **2011**, *475*, 189–195. [[CrossRef](#)] [[PubMed](#)]
65. Young, N.D.; Debelle, F.; Oldroyd, G.E.D.; Geurts, R.; Cannon, S.B.; Udvardi, M.K.; Benedito, V.A.; Mayer, K.F.X.; Gouzy, J.; Schoof, H.; et al. The medicago genome provides insight into the evolution of rhizobial symbioses. *Nature* **2011**, *480*, 520–524. [[CrossRef](#)] [[PubMed](#)]
66. Sherry, S. Ncbi sra toolkit technology for next generation sequence data. *Pump Ind. Anal.* **2000**, *3*, 2230–2234.
67. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)] [[PubMed](#)]
68. Pertea, M.; Kim, D.; Pertea, G.M.; Leek, J.T.; Salzberg, S.L. Transcript-level expression analysis of rna-seq experiments with hisat, stringtie and ballgown. *Nat. Protoc.* **2016**, *11*, 1650–1667. [[CrossRef](#)] [[PubMed](#)]
69. Kolde, R. Pheatmap: Pretty Heatmaps. R Package Version 1.0.8. Available online: <https://CRAN.R-project.org/package=pheatmap> (accessed on 18 December 2018).
70. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [[CrossRef](#)] [[PubMed](#)]

**Sample Availability:** Root samples of the *Panax ginseng* are available from the authors.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).