


RESEARCH ARTICLE

Open Access



Comparative transcriptome analysis of *Gastrodia elata* (Orchidaceae) in response to fungus symbiosis to identify gastrodin biosynthesis-related genes

Chi-Chu Tsai^{1,2}, Keh-Ming Wu^{3†}, Tzen-Yuh Chiang^{4†}, Chun-Yen Huang^{1†}, Chang-Hung Chou^{5†}, Shu-Ju Li¹ and Yu-Chung Chiang^{6,7*} 

Abstract

Background: *Gastrodia elata* Blume (Orchidaceae) is an important Chinese medicine with several functional components. In the life cycle of *G. elata*, the orchid develops a symbiotic relationship with two compatible mycorrhizal fungi *Mycena* spp. and *Armillaria mellea* during seed germination to form vegetative propagation corm and vegetative growth to develop tubers, respectively. Gastrodin (p-hydroxymethylphenol-beta-D-glucoside) is the most important functional component in *G. elata*, and gastrodin significantly increases from vegetative propagation corms to tubers. To address the gene regulation mechanism in gastrodin biosynthesis in *G. elata*, a comparative analysis of *de novo* transcriptome sequencing among the vegetative propagation corms and tubers of *G. elata* and *A. mellea* was conducted using deep sequencing.

Results: Transcriptome comparison between the vegetative propagation corms and juvenile tubers of *G. elata* revealed 703 differentially expressed unigenes, of which 298 and 405 unigenes were, respectively up-regulated (fold-change ≥ 2 , q -value < 0.05 , the trimmed mean of M-values (TMM)-normalized fragments per kilobase of transcript per Million mapped reads (FPKM) > 10) and down-regulated (fold-change ≤ 0.5 , q -value < 0.05 , TMM-normalized FPKM > 10) in juvenile tubers. After Gene Ontology (GO) annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis, 112 up-regulated unigenes with KEGG Ortholog identifiers (KOids) or enzyme commission (EC) numbers were assigned to 159 isogroups involved in seventy-eight different pathways, and 132 down-regulated unigenes with KOids or EC numbers were assigned to 168 isogroups, involved in eighty different pathways. The analysis of the isogroup genes from all pathways revealed that the two unigenes TRINITY_DN54282_c0_g1 (putative monooxygenases) and TRINITY_DN50323_c0_g1 (putative glycosyltransferases) might participate in hydroxylation and glucosylation in the gastrodin biosynthetic pathway.

Conclusions: The gene expression of the two unique unigenes encoding monooxygenase and glycosyltransferase significantly increases from vegetative propagation corms to tubers, and the molecular basis of gastrodin biosynthesis in the tubers of *G. elata* is proposed.

Keywords: Gastrodin, Deep sequencing, Gene regulation, Monooxygenase, Glycosyltransferase

* Correspondence: yuchung@mail.nsysu.edu.tw

[†]Equal contributors

⁶Department of Biological Sciences, National Sun Yat-sen University, Kaohsiung 804, Taiwan

⁷Department of Biomedical Science and Environment Biology, Kaohsiung Medical University, Kaohsiung 807, Taiwan

Full list of author information is available at the end of the article



Background

Gastrodia elata is a rootless and leafless achlorophyllous orchid that grows in a symbiotic relationship with two compatible mycorrhizal fungi, *Mycena* spp. and *Armillaria mellea*, during seed germination and vegetative growth, respectively [1–3]. The seeds of *G. elata* are tiny and do not possess an endosperm, and these seeds germinate only when adequate nutrition is obtained through the digestion of the specific fungi, *Mycena* spp., which invades the embryonic cells of these seeds [4–9]. Currently, four fungi species, including *Mycena anoectochila*, *M. dendrobii* (Fig. 1a), *M. orchidicola*, and *M. osmundicola*, isolated from different species of orchids [10], promote the germination of *G. elata* seeds to form protocorms and further develop into vegetative propagation corms (Fig. 1b) [11–16]. Once vegetative propagation corms have been established from seed germination, *G. elata* undergoes vegetative growth through an established symbiotic association with the compatible mycorrhizal fungi, *A. mellea* (Fig. 1c), to yield juvenile tubers (Fig. 1d) [1, 6, 17]. The vegetative propagation corms of *G. elata* obtain nutrition and energy from *A. mellea* to develop into tubers, and the growth conditions of tubers are positively and closely associated with the hyphal development of this fungi [1, 9, 18, 19]. The hyphae of *A. mellea* develop well in the cortical layers of *G. elata* tubers [1, 20–22]; however, the cells in the pith of tubers digest the invaded hyphae to obtain nutrition and energy [1].

Since ancient times, *Gastrodia elata* has been used as a Chinese medicine for the cure of various conditions, including for its analgesic, antiepileptic, neuroprotective, anticonvulsant, and sedative effects against vertigo, general paralysis, and tetanus [23–26]. Numerous functional components isolated from *G. elata* have been characterized, such as gastrodin (4-hydroxymethylphenyl- β -D-

glucopyranoside) and the aglycone gastrodigenin (4-hydroxybenzyl alcohol) as a primary active ingredient [27–29], and the other related components, including 4-hydroxybenzaldehyde, vanillyl alcohol, and vanillin, also show potential anticonvulsant activity [30–36]. Moreover, other constituents in *G. elata*, including gastrodioside, 4-hydroxybenzyl methyl ether, 4-hydroxybenzoic acid, parishin, β -Sitosterol, bis (4-hydroxybenzyl) sulfide, N6-(4-hydroxybenzyl) adenine riboside, dauricine, citric acid, palmitate, and succinic acid, have been reported [37–39].

Gastrodin, a simple glycoside comprising glucose and 4-hydroxybenzyl alcohol (the precursor of gastrodin), is the major phenolic compound of *G. elata*, and pharmacological tests have shown that this compound exhibits tranquilization, anti-inflammation, analgesia, cortical neuron protection, memory improvement, sedative, anticonvulsant, free radical scavenging, neuroprotective effect, anesthetic, and antioxidant effects [40–42]. Gastrodin was identified, characterized, and artificially synthesized at the end of 1970s [43]. Furthermore, gastrodin biosynthesis markedly increases from the growth stage of vegetative propagation corms to that of juvenile tubers, which have no flower buds [44]. In general, gastrodin production is derived from 4-hydroxybenzyl alcohol through a one-step glycosylation with different glucose donors. Therefore, one of key enzymes of gastrodin biosynthesis is glucosyl-transferase, a large family identified in various plants [45]. Glycosylation is typically the final step in the biosynthesis of secondary plant compounds, resulting in the formation of a large number of glucosides [45–48]. The glycosylation might increase solubility or decrease volatility compared with non-glycosylated molecules [47]. Toluenes are general components in plants that serve as the precursors of plant secondary compounds [49–52].

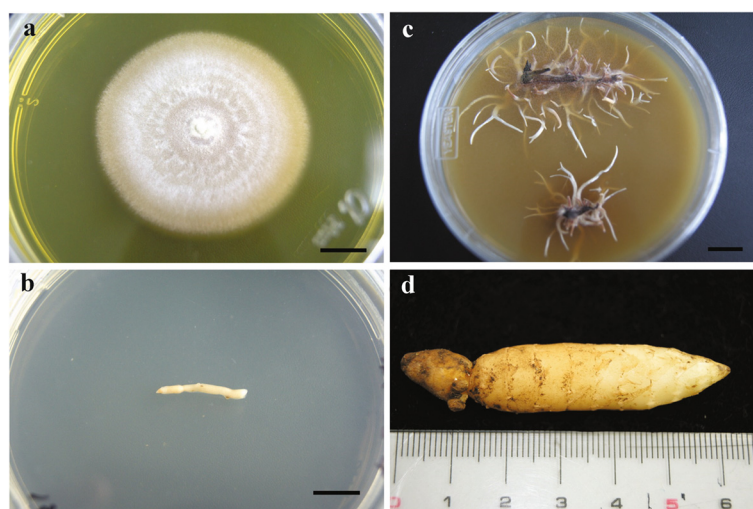


Fig. 1 Materials in the study. Fungus of *Mycena dendrobii* (a), vegetative propagation corm (b), *Armillaria mellea* (c), and juvenile tuber (d) of *Gastrodia elata*. Scale bars = 1 cm

Toluene is considered as a precursor of gastrodin [53]. The derivation of the metabolic pathway of 4-hydroxybenzyl alcohol derived from toluene is largely unknown in *G. elata* and other plants. However, the catalytic pathway from toluene to 4-hydroxybenzyl alcohol has been reported to involve hydroxylation through monooxygenase of cytochrome P450 (CYP450) [54, 55], a member of a large enzyme family in plants that catalyzes most of the oxidation steps in plant secondary metabolism [56–58]. The molecular basis of gastrodin biosynthesis remains largely unknown. In the present study, the comparative transcriptome analysis among *A. mellea*, the vegetative propagation corms and juvenile tubers of gastrodia was conducted using deep sequencing to reveal the gene regulation of gastrodin biosynthesis in *G. elata*.

Results and discussion

The transcriptome sequencing of NGS

De novo transcriptome sequencing has been used in various functional genomics studies, and is particularly suitable for gene expression profiling in non-model organisms without genomic sequences. The next generation sequencing (NGS) technology not only provides a comparative expressed sequence tag (EST) analysis for gene discovery on a genome-wide scale in non-model plants but also an efficient process for transcriptome sequencing and characterization. NGS platforms, such as the Illumina/Solexa Genome Analyzer and the Roche 454 GS FLX, have been widely used in recent years for the high-throughput sequencing of various organisms [59, 60]. Using these techniques for *de novo* transcriptome sequencing, EST databases have been successfully obtained for several medicinal herbs, including American ginseng [48], *Salvia miltiorrhiza* [61], sweet wormwood [62], *Euphorbia fischeriana* [63], *Taxus* [64], and other crops, such as chili pepper [65], maize [66], *Curcuma longa* [67], chestnut [68], *Eucalyptus* tree [69], olive [70], *Camellia sinensis* [71], sweet potato [72], *Arabidopsis* [73], and *Phalaenopsis* [74]. The Illumina platform is beneficial and useful for gene discovery because this technique can obtain deeper coverage and higher accuracy than Roche 454 sequencing technology [74]. Hence, the Illumina system was used in the present study to clarify the differential gene expression of different life stages of *G. elata*.

Sequencing and *de novo* assembly

A total of 21,045,338 (2x75 bases, 51 % GC), 18,436,794 (2x75 bases, 44 % GC) and 18,253,900 (2x75 bases, 45 % GC) high-quality paired-end (PE) reads were generated from Illumina HiSeq2000 platform, and approximately 3 giga bases of sequence data were obtained for each of *A. mellea*, vegetative propagation corm, and juvenile tuber of *G. elata* (Table 1). These short sequence reads have

been deposited in NCBI under GEO accession number GSE73633. High-quality bases (above Q20) were more than 97 % for all three samples, indicating an excellent quality (Q20 means 1 error per 100 sequenced bases), while another high quality indicator was the abundance peak of average sequence quality per read located around Q38 (less than 0.0158 % of error rate) for all three samples. The high-quality PE reads were used for *de novo* transcriptome shotgun assembly (TSA) to build transcript isoforms based on paired-end information. From a total of 161,517 assembled transcript isoforms (≥ 200 bases), 134,441 transcripts were selected as the representative unigenes with longest length for all loci (i.e., genes). The mapping rates of these high quality reads from all three samples against the total transcripts were all above 86 % (Table 2). The final N50 lengths of 1592 and 1184 bases, and the total lengths of 137,618,051 and 92,177,843 bases were calculated for the transcripts and the unigenes, respectively (Table 2) (Fig. 2). N50 statistics are widely used to assess the quality of the assembly, and the higher the N50 value representing the assembly the better the quality [67]. Compared with other plant transcriptome sequencing of *de novo* TSAs [61, 67, 69, 71, 72, 75–78], the N50 value obtained in the present study was above the average and adequate for further analysis (Table 3).

Functional annotation and Gene Ontology classification

All unigenes were annotated according to the sequence similarity search against NCBI non-redundant protein sequence (nr) database using BLASTX algorithm. A total of 59,932 unique sequences were annotated, accounting for 44.58 % of the total unigenes (Table 4). Gene Ontology (GO) assignment were performed for the functional categorization of the annotated unigenes. A total of 11,645 unigenes were mapped to GO terms, accounting for 8.66 % of the unigenes (Table 4). Because multiple GO terms can be assigned to the same unigene [79], totally 58,488 GO terms were assigned in the present study. The GO annotation showed that these unigenes represent diverse functionalities and are involved in various metabolic pathways. In *A. mellea*, 9101, 7484, and 5694 GO terms, respectively, represent molecular function, biological process and cellular component categories [See Additional file 1: Figure S1]. In the molecular function category, the terms integral to “binding” (GO:0005488) and “catalytic activity” (GO:0003824) were shown as the most frequently occurring, constituting 19.23 % (4285) and 18.56 % (4136) of the level 2 GO terms, respectively. “Metabolic process” (GO:0008152) and “biological regulation” (GO:0065007) were the most frequently occurring under the biological process category, representing 9.50 % (2116) and 6.97 % (1552) of the level 2 GO terms, respectively. In the cellular

Table 1 Basic Statistics of RNA-Seq generated from *Armillariella mellea*, vegetative propagation corm and juvenile tuber of *Gastrodia elata* encoding by Illumina HiSeq2000 platform

| | Sample | | |
|----------------------------|-----------------------------|-----------------------------|--------------------------|
| | Vegetative propagation corm | Juvenile tuber ^a | <i>Armillaria mellea</i> |
| Total number of read pairs | 18,436,794 | 18,253,900 | 21,045,338 |
| Total nucleotides (nt) | 2,765,519,100 | 2,738,085,000 | 3,156,800,700 |
| GC percentage | 44 % | 45 % | 51 % |
| ≥Q20 percentage | 97 % | 97 % | 97 % |
| Read length | 75 | 75 | 75 |

^aThe tubers of *Gastrodia elata* have established symbiotic associations with *Armillaria mellea*

component category, “cell part” (GO:0044464) was the most frequently occurring, representing 19.20 % (4277) of the total level 2 GO terms.

In vegetative propagation corm of *G. elata*, 17,371, 9875, and 6915 GO terms, respectively, represent molecular function, biological process, and cellular component categories [See Additional file 1: Figure S1]. In the molecular function category, the terms integral to “binding” (GO:0005488) and “catalytic activity” (GO:0003824) occurred most frequently, representing 24.84 % (8484) and 20.71 % (7076) of the total level 2 GO terms,

respectively. In the biological process category, “metabolic process” (GO:0008152) was the most frequently occurring, representing 11.95 % (4083) of the total level 2 GO terms. In the cellular component category, “cell part” (GO:0044464) was the most frequently observed, representing 15.69 % (5359) of the total level 2 GO terms.

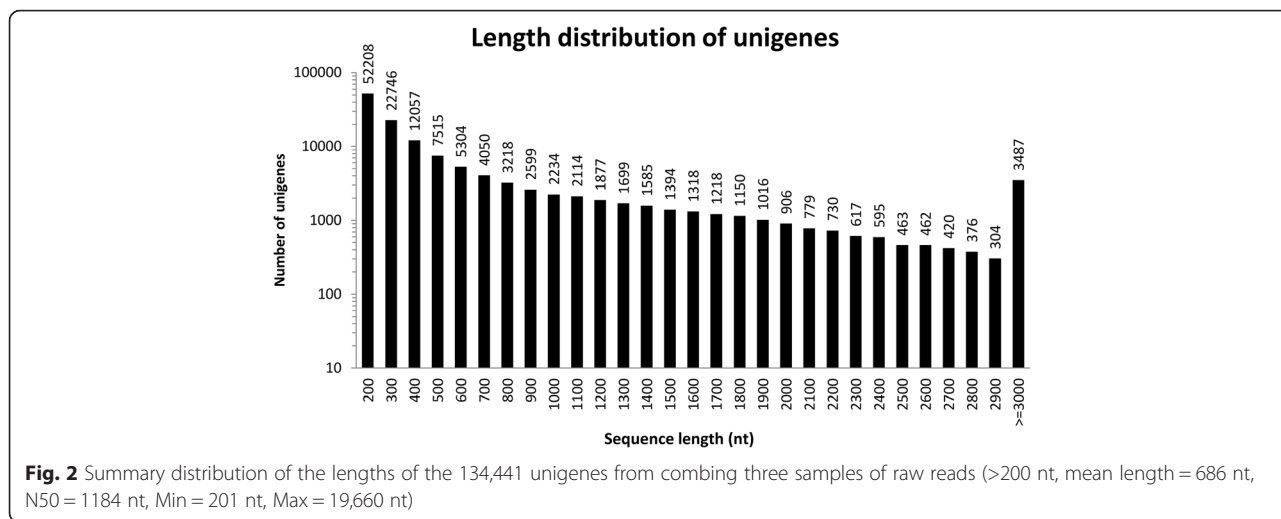
In the juvenile tuber of *G. elata*, 10,259, 5927, and 3753 GO terms were shown for molecular function, biological process and cellular component categories, respectively [See Additional file 1: Figure S1]. In the

Table 2 Summary of the *de novo* transcriptome shotgun assembly from all Illumina sequences

| | |
|---|----------------|
| Assembled transcripts | |
| Total number of transcripts | 161,517 |
| Total bases of transcripts | 137,618,051 nt |
| Longest transcript | 19,660 nt |
| Average transcript length | 852 nt |
| Median transcript length | 420 nt |
| N50 transcript length | 1592 nt |
| Read mapping rate of sample <i>Armillaria mellea</i> | 88.21 % |
| Read mapping rate of sample juvenile tuber ^a | 88.73 % |
| Read mapping rate of sample vegetative propagation corm | 86.75 % |
| Assembled unigenes | |
| Total number of transcripts | 134,441 |
| Total bases of unigenes | 92,177,843 nt |
| Longest unigene | 19,660 nt |
| Average unigene length | 686 nt |
| Median unigene length | 358 nt |
| N50 unigene length | 1184 nt |
| Predicted peptides | |
| Total number of peptides | 50,084 |
| Total amino acids of peptides | 14,597,170 aa |
| Longest peptide | 5370 aa |
| Average unigene length | 291 aa |
| Median peptide length | 199 aa |

Abbreviations: aa amino acids, nt nucleotides

^aThe tubers of *Gastrodia elata* have established symbiotic associations with *Armillaria mellea*



molecular function category, the terms integral to “binding” (GO:0005488) and “catalytic activity” (GO:0003824) were most frequently observed, representing 27.18 % (5420) and 19.53 % (3895) of the total level 2 GO terms, respectively. Metabolic process (GO:0008152) was the most frequently observed under the biological process category, representing 12.37 % (2466) of the total level 2 GO terms. In the cellular component category, “cell part” (GO:0044464) was the most frequently observed, representing 14.8 % (2951) of the total level 2 GO terms.

Differential expression analysis between *A. mellea* and juvenile tuber of *Gastrodia elata*

The comparative analysis of the transcriptomes of the *A. mellea* and *G. elata* (symbiosis with *A. mellea*) juvenile tubers was conducted based on the combined transcriptome assembly of all three samples. Among the total 134,441 unigenes, 49,890 and 71,722 unigenes were aligned with reads from *A. mellea* and *G. elata* juvenile

Table 3 Comparative analysis of plant transcriptome N50 values

| Organisms | N50 (in bases) |
|--|----------------|
| <i>Ipomoea batatas</i> (Wang et al. [72]) | 765 |
| <i>Eucalyptus grandis</i> (Mizrachi et al. [69]) | 1640 |
| <i>Salvia miltiorrhiza</i> (Hua et al. [61]) | 533 |
| <i>Camellia sinensis</i> (Shi et al. [71]) | 506 |
| <i>Cajanus cajan</i> (Kudapa et al. [76]) | 1510 |
| <i>Myrica rubra</i> (Feng et al. [75]) | 708 |
| <i>Capsicum frutescens</i> (Liu et al. [65]) | 1108 |
| <i>Curcuma longa</i> (Annadurai et al. [67]) | 1448 |
| <i>Cuscuta pentagona</i> (Ranjan et al. [77]) | 1550 |
| <i>Hevea brasiliensis</i> (Salgado et al. [78]) | 837 |
| Average of above studies | 1061 |
| This study | 1592 |

tubers, respectively [See Additional file 2: Figure S2]. Among these unigenes, only 5547 unigenes were identified in both samples of deep sequencing data. To evaluate differential gene expression, the absolute value of the log2-FC (fold changes) ≥ 1, the *q*-values < 0.05 and the TMM-normalized FPKM > 0.3 were used as the criteria to determine the significance of gene expression differences [80]. A total of 292 differentially expressed unigenes were revealed in the transcriptome comparison, of which sixty-nine unigenes were significantly up-regulated (log2-FC ≥ 1, FPKM > 0.3, *q*-values < 0.05) in *A. mellea* [See Additional file 3: Table S1], and 223

Table 4 Summary statistics of unigenes with functional annotations for all combined assembly and for each sample

| Functional annotation | No. of unigene hits | Percentage (%) |
|---|---------------------|----------------|
| All (134,441 unigenes) | | |
| NR | 59,932 | 44.58 |
| GO | 11,645 | 8.66 |
| KEGG | 27,008 | 20.09 |
| Vegetative propagation corm (82,712 unigenes) | | |
| NR | 33,638 | 40.67 |
| GO | 7338 | 8.87 |
| KEGG | 15,073 | 18.22 |
| Juvenile tuber ^a (71,722 unigenes) | | |
| NR | 27,296 | 38.06 |
| GO | 4396 | 6.13 |
| KEGG | 11,301 | 15.76 |
| <i>Armillaria mellea</i> (49,890 unigenes) | | |
| NR | 29,401 | 58.93 |
| GO | 3900 | 7.82 |
| KEGG | 13,243 | 26.54 |

^aThe tubers of *Gastrodia elata* have established symbiotic associations with *Armillaria mellea*

unigenes in *G. elata* juvenile tubers were expressed at significantly higher levels ($\log_2\text{-FC} \leq -1$, $\text{FPKM} > 0.3$, $q\text{-values} < 0.05$) (Fig. 3a) [See Additional file 4: Table S2]. Of 292 the differentially expressed unigenes, only 106 unigenes can be assigned with KEGG Ortholog identifiers (KOids) or enzyme commission (EC) numbers corresponding to biological pathways for cellular functions and molecular interactions after KEGG analysis. Among these, twenty-five up-regulated unigenes from *A. mellea* and *G. elata* juvenile tuber were assigned KOids or EC numbers corresponding to twenty-eight isogroups involved in sixteen different pathways [See Additional file 5:

Table S3]; and eighty-one down-regulated unigenes were corresponding to 134 isogroups involved in sixty different pathways [See Additional file 6: Table S4].

Differential expression analysis between the juvenile tuber and vegetative propagation corm of *Gastrodia elata*

The comparative analysis of transcriptomes of the juvenile tubers (symbiosis with *A. mellea*) and vegetative propagation corms (asymbiosis with *A. mellea*) was conducted also based on the combined transcriptome assembly of all three samples. Among the total 134,441 unigenes, 71,722 and 82,712 unigenes were aligned with

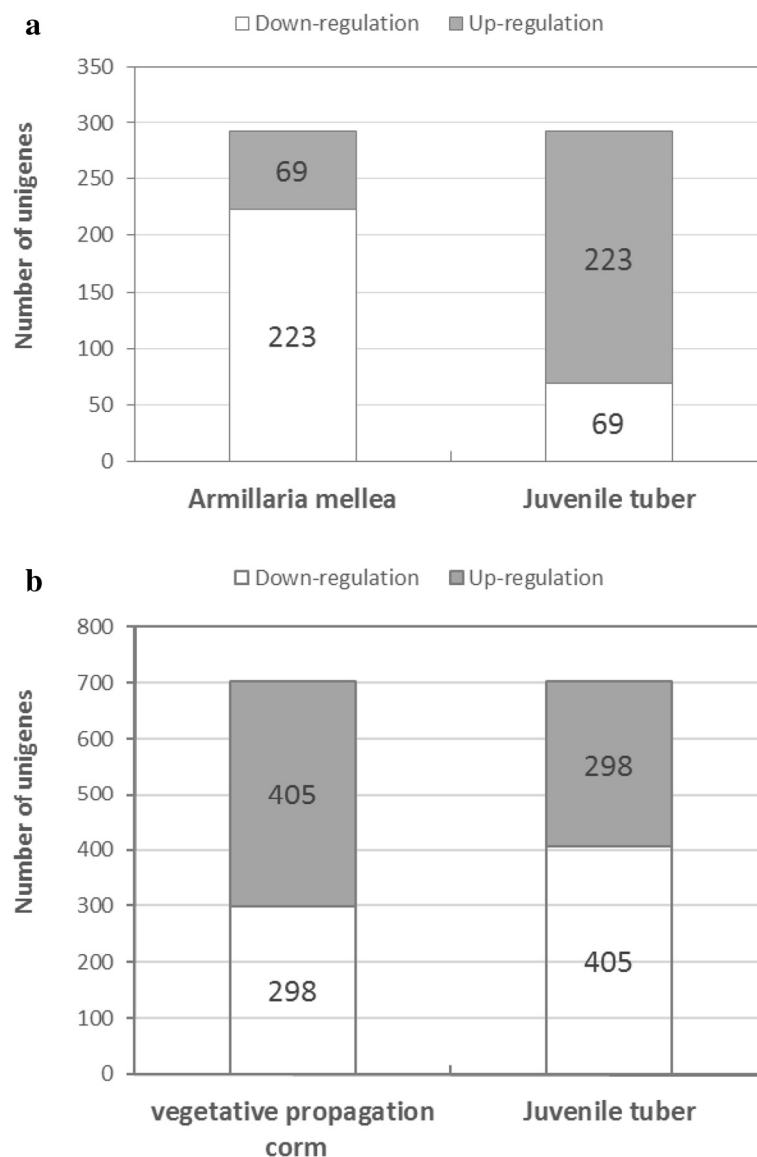


Fig. 3 Differentially expressed unigenes between **a** *Armillaria mellea* and juvenile tuber of *Gastrodia elata* (under the criteria: the absolute value of the $\log_2\text{-FC} \geq 1$, the $q\text{-values} < 0.05$ and the TMM-normalized FPKM > 0.3), and **b** vegetative propagation corm and juvenile tuber (under the criteria: the absolute value of the $\log_2\text{-FC} \geq 1$, the $q\text{-values} < 0.05$ and the TMM-normalized FPKM > 10). Numbers of up- and down-regulated unigenes were shown in boxes

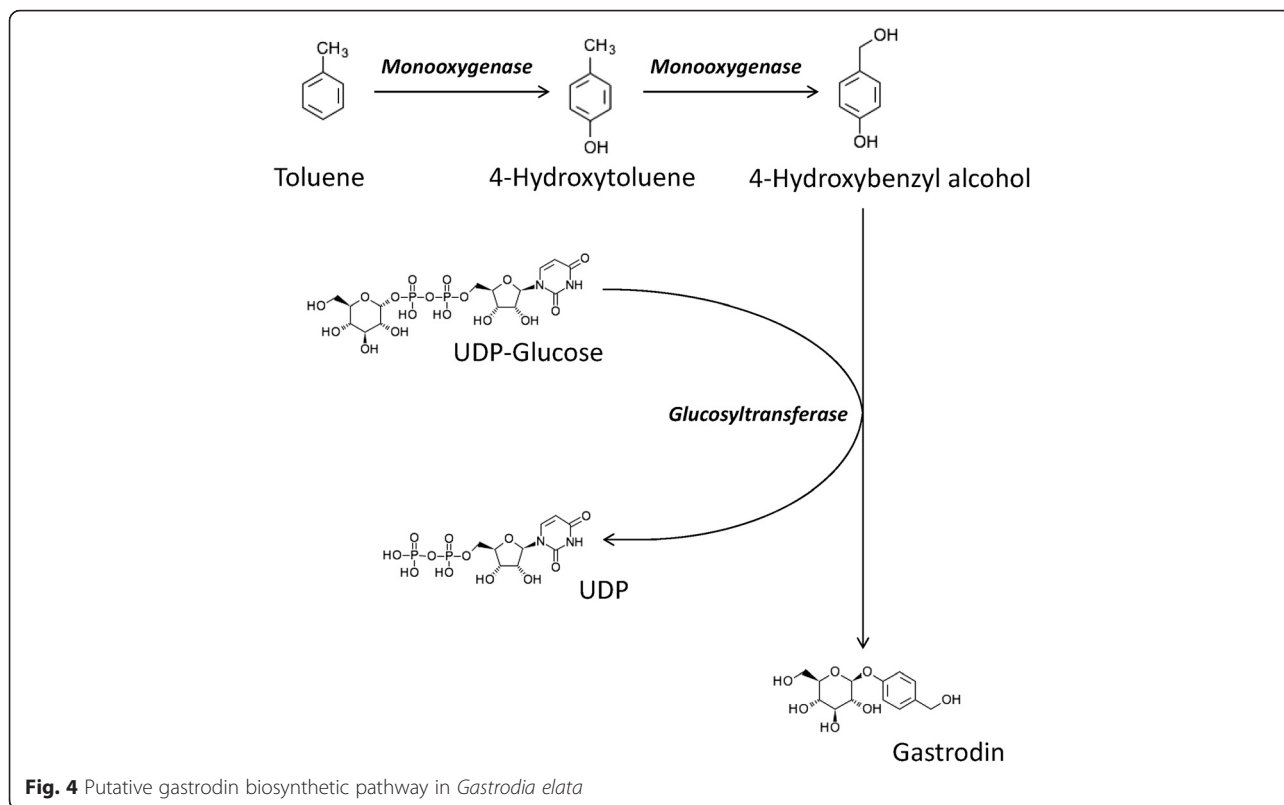
reads from juvenile tubers and vegetative propagation corms, respectively [See Additional file 7: Figure S3]. Among these, 63,317 unigenes were identified in both samples, and 22,942 differentially expressed unigenes were revealed in the transcriptome comparison using the same criteria (the absolute value of the \log_2 -FC ≥ 1 , the q -values < 0.05 and the TMM-normalized FPKM > 0.3). 7383 and 15,559 unigenes were expressed at higher levels in *G. elata* juvenile tubers and vegetative propagation corms, respectively (data not shown). To focus on highly and differentially expressed unigenes by modifying the threshold of TMM-normalized FPKM larger than ten, 703 highly and differentially expressed unigenes were revealed in the transcriptome comparison. Among which 298 unigenes were significantly up-regulated (\log_2 -FC ≥ 1 , q -values < 0.05 , TMM-normalized FPKM > 10) in juvenile tubers [See Additional file 8: Table S5], and 405 unigenes in vegetative propagation corms were expressed at significantly higher levels (\log_2 -FC ≤ -1 , q -values < 0.05 , TMM-normalized FPKM > 10) (Fig. 3b) [See Additional file 9: Table S6]. Of 703 the differentially expressed genes, only 244 unigenes can be assigned to KOids or EC numbers corresponding to biological pathways for cellular functions and molecular interactions after KEGG analysis. Among these, 112 up-regulated (\log_2 -FC ≥ 1 , q -value < 0.05 , TMM-normalized FPKM > 10) unigenes from juvenile tubers compared with vegetative propagation corms were assigned to KOids or EC numbers corresponding to 159 isogroups involved in seventy-eight different pathways [See Additional file 10: Table S7]; and 132 down-regulated (\log_2 -FC ≤ -1 , q -value < 0.05 , TMM-normalized FPKM > 10) unigenes were assigned to KOids or EC numbers corresponding to 168 isogroups involved in eighty different pathways [See Additional file 11: Table S8].

Kusano (1911) first reported that *G. elata* existed in a mycorrhizal relationship with the wood-rotting pathogen *A. mellea*; however, this relationship was uncharacterized [18]. Until 1980, Zhang and Li showed that *G. elata* digests the invasive hyphae of *A. mellea* as the source of nutrition [1]. Lan et al. (1986) also confirmed that *A. mellea* was used as the source of nutrition for *G. elata* through the observation of labeled materials from *A. mellea* in the transverse section of *G. elata*. The labeled materials appeared in mitochondria, the endoplasmic reticulum and vacuoles of *G. elata* cortical cells [20]. When the hyphae of *A. mellea* are disconnected between wood (source of nutrition for *A. mellea*) and *G. elata*, the growth of *G. elata* terminates and this organism dies; therefore, the role of *A. mellea* for *G. elata* was considered as the food for survival [21]. According to the differential gene expression in *G. elata* in response to *A. mellea* symbiosis, unigene TRINITY_DN70668_c0_g1 is significantly induced [See Additional file 8: Table S5], as high as ~ 7 folds, and this gene was annotated as a gastrodianin (i.e., gastrodia

antifungal proteins, GAFFPs) gene, which digests the cell wall of *A. mellea* [81]. This result suggested that *G. elata* digests the invasive hyphae of *A. mellea* as a source of nutrition according to Zhang and Li (1980) [1]. In the present study, the low-level gene expression of the gastrodianin biosynthetic gene was detected in vegetative propagation corms (i.e., primary corms), which differentiated from protocorms, suggesting that the symbiotic relationship between *A. mellea* and *G. elata* only can be developed during the vegetative propagation corms of *G. elata* [82]. According to previous reports, there are two copies of gastrodianin biosynthetic genes in *G. elata* [81], and these two genes (unigenes TRINITY_DN70668_c0_g1 and TRINITY_DN48867_c0_g1) were also identified through deep sequencing data in the present study, only unigene TRINITY_DN70668_c0_g1 was significantly induced in the juvenile tubers of *G. elata* in response to *A. mellea* symbiosis [See Additional file 8: Table S5]. The result was consistent with the previous gene expression study of the gastrodianin biosynthetic gene in *G. elata* [81].

Identification and validation of candidate genes involved in gastrodin biosynthesis

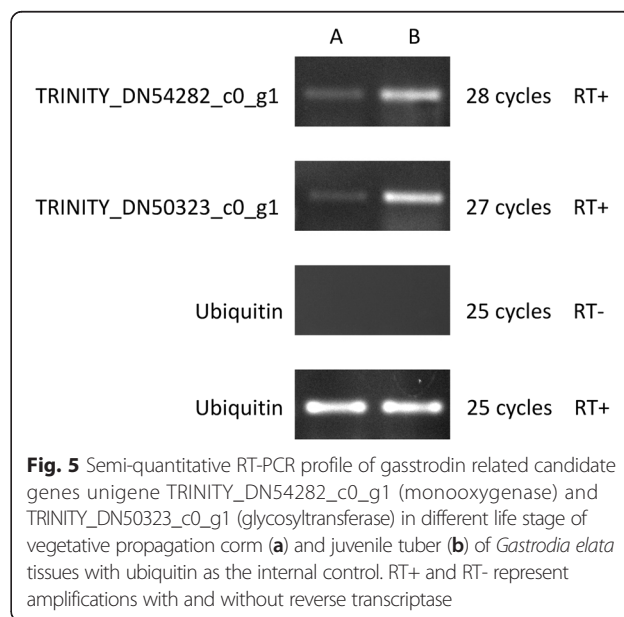
The mechanism and related genes in the gastrodin biosynthesis pathway are currently unknown. To the best of our knowledge, gastrodin (4-hydroxymethylphenyl- β -D-glucopyranoside) is a simple glycoside comprising glucose and 4-hydroxybenzyl alcohol [27]. The last biosynthesis enzyme of gastrodin is glucosyltransferase [45]. Gastrodins are synthesized from 4-hydroxybenzyl alcohol with UDP-glucose via glucosylation catalyzed through glucosyltransferase. The precursor of gastrodin, 4-hydroxybenzyl alcohol, is catalyzed through cresols degradation (toluene degradation) from toluene through two steps of hydroxylation via monooxygenase (CYP450) [54, 55]. Therefore, both monooxygenase and glucosyltransferase are considered two key enzymes for gastrodin biosynthesis [53]. According to the chemical structure of gastrodin and gastrodin precursors, analyzed in previous reports, the putative gastrodin biosynthetic pathway is shown in Fig. 4. However, both monooxygenase (CYP450) and glucosyltransferase belong to a large enzyme families involved in different biosynthesis pathways in various plants [45, 54, 55]. Moreover, 4-hydroxybenzyl alcohol and gastrodin were also detected in *Anoectochilus formosanus* [83]. To determine the candidate genes involved in gastrodin biosynthesis, the comparative analysis of the transcriptomes between vegetative propagation corm and juvenile tuber of *G. elata* was conducted. A total of thirty and forty-six unigenes, respectively, were annotated to monooxygenase and glucosyltransferase among the 63,317 unigenes expressed both in the juvenile tubers and the vegetative propagation corms of *G. elata*. Under the criteria



of $\log_2\text{-FC} \geq 1$, TMM-normalized FPKM > 3 and q -values < 0.05 , four putative monoxygenases were selected. Among them, unigene TRINITY_DN54282_c0_g1 was the most abundance in FPKMs and differentially expressed one (~2.4 times higher in the juvenile tubers). Under the same criteria, three putative glucosyltransferases were selected and the unigene TRINITY_DN50323_c0_g1 was the highest differentially expressed one (~3.2 times higher in the juvenile tubers).

To validate the differential expression of gastrodin biosynthesis-related unigenes TRINITY_DN54282_c0_g1 and TRINITY_DN50323_c0_g1, we investigated the expression of these genes in different life stages between the vegetative propagation corms and juvenile tubers of *G. elata* using semi-quantitative RT-PCR and real-time PCR (quantitative RT-PCR, qRT-PCR), and compared the results with fold-changes of FPKM (RNA-Seq). The specific primers for semi-quantitative RT-PCR and qRT-PCR were designed. In semi-quantitative RT-PCR, expression profiling revealed the differential expression of both monoxygenases and glucosyltransferases between the vegetative propagation corms and juvenile tubers of *Gastrodia elata* (Fig. 5). Both monoxygenases and glucosyltransferase genes were up-regulated in the juvenile tubers and were considered as gastrodin biosynthetic-related genes, as gastrodin production markedly increases from the growth stage of vegetative

propagation corms to that of juvenile tubers [44]. In addition, the differential expression of the two genes was also validated through qRT-PCR analysis as shown in Fig. 6. The expression levels of the unigene TRINITY_DN54282_c0_g1 were up-regulated up to 6.5 times in juvenile tubers and 2.4 times for qRT-PCR and RNA-



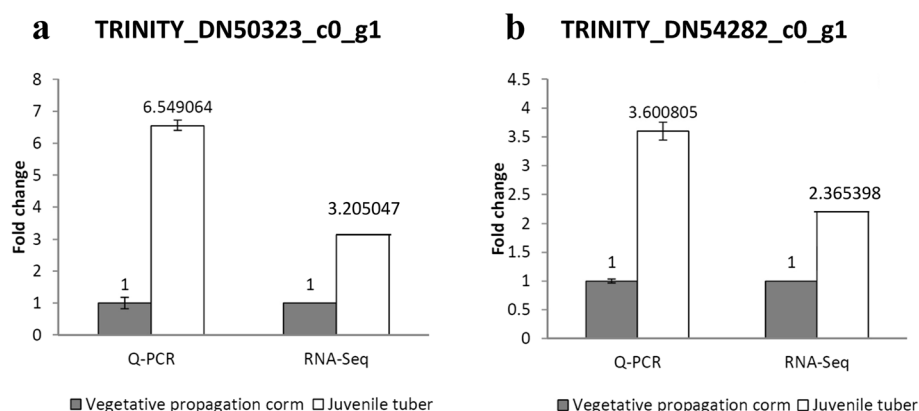


Fig. 6 Quantitative real-time PCR (Q-PCR) validations of (a) unigene TRINITY_DN50323_c0_g1 and (b) unigene TRINITY_DN54282_c0_g1 of RNAseq results (TMM-normalized FPKM fold changes). Comparison of differential expression values between the juvenile tuber (white bar) and vegetative propagation corm (grey bar) of *Gastrodia elata* determined by qRT-PCR and RNAseq

Seq, respectively. Similarly, qRT-PCR and RNA-Seq showed that the expression levels for unigene TRINITY_DN50323_c0_g1 were up-regulated approximately 3.6 and 3.2 times, respectively, in juvenile tubers. The results of both semi-quantitative RT-PCR (Fig. 5) and qRT-PCR analysis (Fig. 6) were consistent with the RNA-Seq data.

Monoxygenases belong to cytochrome P450 proteins, the largest family of plant proteins, which catalyze most of the oxidation steps in plant secondary metabolism [57, 58]. The comparison of the chemical structures of 4-hydroxybenzyl alcohol and the precursor toluene (Fig. 4), revealed two steps catalyzed through monoxygenase, the conversion of toluene to 4-hydroxytoluene and the conversion of 4-hydroxytoluene to 4-hydroxybenzyl alcohol. Therefore, unigene TRINITY_DN54282_c0_g1 (EC: 1.14.13.-) was considered as a toluene monoxygenase gene, consistent with the KEGG pathway annotation (Fig. 7). Generally, glycosylation is the last step in the biological biosynthesis of secondary metabolism because sugar conjugation results in both the increased water solubility and stability of the compounds [84–86]. In *G. elata*, glycosyltransferase catalyzes the last step of the gastrodin biosynthesis pathway, which converts 4-hydroxybenzyl alcohol to gastrodin [45] (Fig. 4). Therefore, unigene TRINITY_DN50323_c0_g1 (EC: 2.4.1.12) was considered as a glycosyltransferase gene, consistent with the KEGG pathway annotation (Fig. 8). In short, both unigenes TRINITY_DN54282_c0_g1 and TRINITY_DN50323_c0_g1 might be key enzyme genes that, respectively, participate in the hydroxylation (Fig. 7) and glycosylation (Fig. 8) of gastrodin (Fig. 4).

Cloning of gastrodin biosynthesis related genes

The full-length cDNA sequences of monoxygenases (unigene TRINITY_DN54282_c0_g1) and glycosyltransferases

(unigene TRINITY_DN50323_c0_g1) from *G. elata* were further isolated through RACE analysis. The nucleotide sequence of the full-length monoxygenase cDNA has an open reading frame (ORF) of 1476 nucleotides spanning from the first initiation codon (ATG) to the termination codon (TGA), an in-frame stop codon located 12 nt upstream from the initiation codon and an out-of-frame ATG located upstream of the main ORF. The complete ORF encodes a protein of 491 amino acids with a predicted molecular mass of 55.8 kDa (Fig. 9a). In addition, the nucleotide sequence of the full-length glycosyltransferase cDNA has an ORF of 1635 nucleotides spanning from the first initiation codon ATG to the termination codon TGA, an in-frame stop codon located 12 nt upstream from the initiation codon and an out-of-frame ATG located upstream of the main ORF. The complete ORF encodes a protein of 544 amino acids with a predicted protein of 63.1 kDa (Fig. 9b).

Notably, both monoxygenases (unigene TRINITY_DN54282_c0_g1) and glycosyltransferases (unigene TRINITY_DN50323_c0_g1) induced in response to fungi symbiosis possess an out-of-frame upstream ATG and an in-frame stop codon in the main ORF within the 5'UTR. In mammals, upstream ATGs/upstream ORFs significantly reduce protein expression levels through a reduction of the translation efficiency [87, 88] or mRNA decay [89–91]. Upstream AUGs/upstream ORFs in the 5' UTR efficiently disrupt the translation of the downstream coding sequence, thereby reducing the translation efficiency of the main coding region [87, 88]. According to deep sequencing, semi-quantitative RT-PCR, and real-time RT-PCR, both monoxygenases (unigene TRINITY_DN54282_c0_g1) and glycosyltransferase (unigene TRINITY_DN50323_c0_g1) genes were co-expressed in vegetative propagation corms. Therefore, the upstream ATGs of these two genes might result in

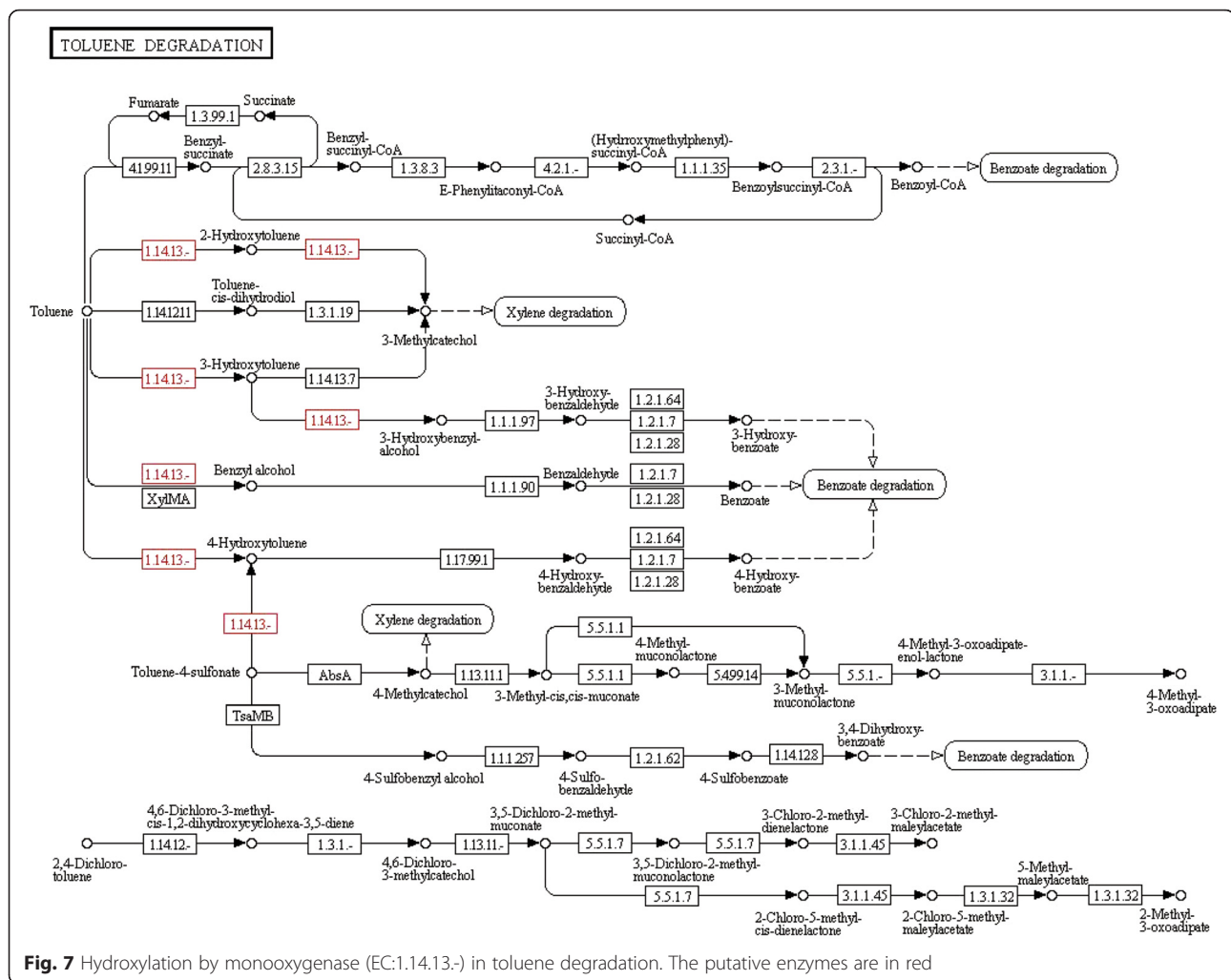


Fig. 7 Hydroxylation by monooxygenase (EC:1.14.13.-) in toluene degradation. The putative enzymes are in red

the low concentration of gastrodin in vegetative propagation corms. The induction of gene expression in response to fungi symbiosis might increase the translation efficiency or mRNA stability of the two key enzymes to increase gastrodin production in juvenile tubers. In response to stress, the translation repression of upstream ATGs/upstream ORFs could be significantly reduced [92–94] or mRNA stability could be increased [95, 96]. In response to fungi invasion, the repression reduction of the translation efficiency of both monooxygenases (unigene TRINITY_DN54282_c0_g1) and glycosyltransferase (unigene TRINITY_DN50323_c0_g1) genes in *G. elata* might also increase the accumulation of gastrodin in juvenile tubers, as fungal infection could be a biotic stress to *G. elata*.

Conclusions

The molecular basis of gastrodin biosynthesis in *G. elata* was clarified based on *de novo* transcriptome sequencing in the present study. Two putative monooxygenase (unigene TRINITY_DN54282_c0_g1) and glycosyltransferase

(unigene TRINITY_DN50323_c0_g1) genes associated with the gastrodin biosynthesis pathway were identified. The genes of the two key enzymes involved in gastrodin biosynthesis might be applied as the target genes for plant gene transformation in future studies to obtain transgenic plants or microbial hosts with gastrodin production. Moreover, this transcriptome dataset also provides important information to accelerate future gene expression and functional genomics studies in *G. elata*.

Methods

Materials

Plant materials used in this study were got from the Chinese medical farm. The voucher specimens were deposited at the herbarium of the Taiwan Endemic Species Research Institute (TAIE) and the voucher numbers are Hsu 17054 and 17055. The vegetative propagation corms and juvenile tubers of *G. elata* were, respectively, harvested 1 and 12 months after sowing at the Chinese medical farm (Hakusan City, Jilin Province, China) (Fig. 1b and d). *A. mellea* was

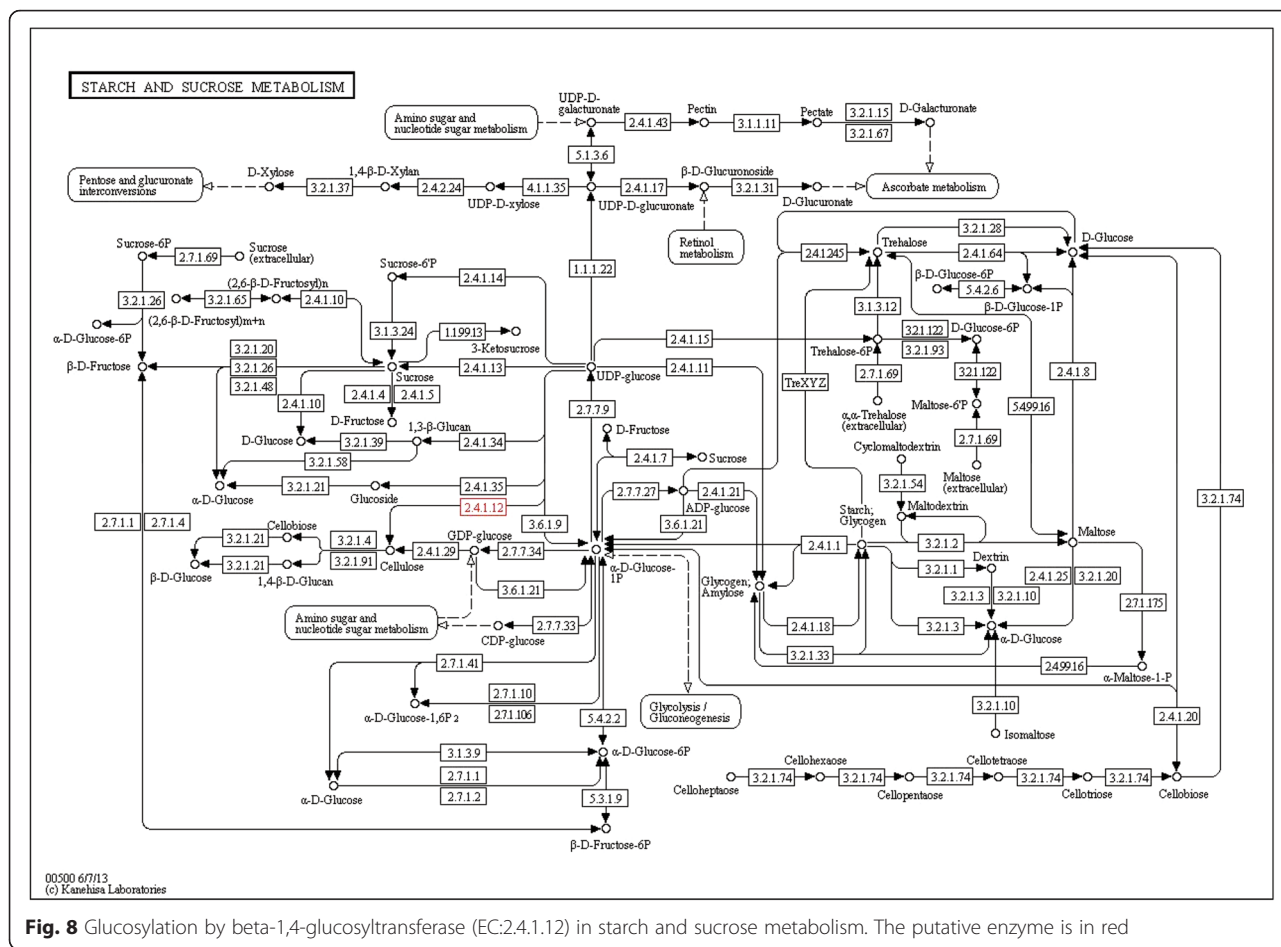


Fig. 8 Glucosylation by beta-1,4-glucosyltransferase (EC:2.4.1.12) in starch and sucrose metabolism. The putative enzyme is in red

isolated from the juvenile tuber of *G. elata* and identified based on ITS sequence of nuclear ribosomal DNA (nrDNA) (data not shown). The fungi were cultured on PDA medium according to Mishra and Dubey (1994) for further molecular studies (Fig. 1c) [97].

RNA isolation, cDNA library preparation, deep sequencing and de novo assembly

The vegetative propagation corms of *G. elata*, juvenile tuber of *G. elata* and *A. mellea* was separately harvested and extracted total RNA using Trizol® Reagent

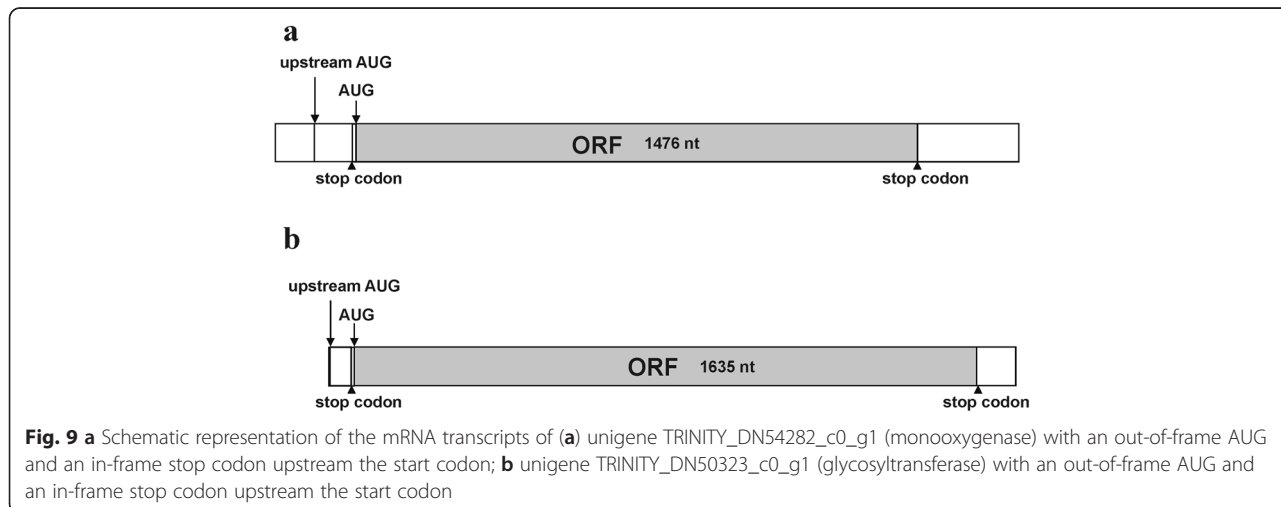


Fig. 9 a Schematic representation of the mRNA transcripts of (a) unigene TRINITY_DN54282_c0_g1 (monoxygenase) with an out-of-frame AUG and an in-frame stop codon upstream the start codon; (b) unigene TRINITY_DN50323_c0_g1 (glycosyltransferase) with an out-of-frame AUG and an in-frame stop codon upstream the start codon

(Invitrogen, Carlsbad, CA, USA) and followed by the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany). Purified RNA was quantified at OD₂₆₀ using a ND-1000 spectrophotometer (NanoDrop Technology, San Diego, CA, USA) and qualified using an Agilent Bioanalyzer 2100 with the RNA 6000 nano labchip kit (Agilent Technology, Santa Clara, CA, USA). The RNA Integrity Number (RIN) was identified 8.7, 9.1, and 8.0 for vegetative propagation corm, juvenile tuber, and *A. mellea*, respectively. Each of the RNA libraries was separately constructed using TruSeq RNA Library Preparation Kit v2 (Illumina Inc., San Diego, CA) with standard Illumina protocols. After quality-control using an Agilent Bioanalyzer 2100, each library was paired-end deep sequenced by Illumina HiSeq 2000 sequencer (Illumina Inc., San Diego, CA, USA) according to the manufacturer's instructions. The quality of the produced fastq sequences was assessed using FastQC program (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), see Additional files 12, 13 and 14: Figures S4, S5 and S6. Before the assembling stage, the reads were processed to trim adaptor sequences and low-quality ends via Trimmomatic version 0.32 [98] with parameters "ILLUMINACLIP: TruSeq3-PE.fa:2:30:10 SLIDINGWINDOW:4:5 LEADING:5 TRAILING:5 MINLEN:25". A single *de novo* transcriptome assembly was generated from high-quality short reads of all three samples using the Trinity software (<https://github.com/trinityrnaseq/trinityrnaseq/wiki/>) [99]. Trinity v2.1.0 (release 20140413p1) was employed with the default *k-mer* of 25 and minimum assembled contig length of 200. Candidate coding regions within transcript sequences were identified using TransDecoder (<http://transdecoder.github.io/>).

Transcript abundance estimation

Quantification of transcripts was estimated using the RNA-Seq by Expectation-Maximization (RSEM) software version 1.2.23 [100], which was bundled with the Trinity software distribution. The RSEM protocol uses the Bowtie software (<http://bowtie-bio.sourceforge.net/index.shtml>) [101] to align trimmed reads from each sample separately to the assembled transcripts, and then computes transcript abundance, estimates the number of aligned fragments corresponding to each transcript, including normalized expression values as FPKM for paired-end reads. In addition, RSEM computes 'gene-level' expression values using the Trinity component as a proxy for the gene. For comparing expression levels of different transcripts or genes across samples, a Trinity-bundled script invokes the EdgeR package to perform an additional TMM scaling normalization that aims to estimate differences in total RNA production across all samples [102, 103]. Transcripts with zero TMM-normalized FPKM values for all three samples were removed from the assembly and not counted into

the total transcript number, and the rest with longest length per component (i.e., gene locus) defined by Trinity were interpreted to represent "unigenes" for downstream analysis. The data presented in this publication have been deposited in NCBI's Gene Expression Omnibus [104] and are accessible through GEO Series accession number GSE73633 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE73633>).

Gene functional annotation, GO classification and KEGG pathway analysis

The functional annotation of the unigenes was conducted using the BLASTX algorithm of DIAMOND program [105] with E-values < 1.00E-5 and enabled 'sensitive mode' against NCBI non-redundant protein sequence database (nr, October 2015), which consisted of all non-redundant peptide sequences from GenBank CDS translations, RefSeq Proteins, PDB (Protein Data Bank) database, SwissProt database, PIR (Protein Information Resource) database, and PRF (Protein Research Foundation) database. Gene Ontology (GO) annotations of unigenes was performed by converting the GeneBank identifiers (gi) of hits from the BLASTX results into UniProt IDs through the IDmapping data files [106] downloaded from The Universal Protein Resource (UniProt; <http://www.uniprot.org/>), then the corresponding GO terms were retrieved via the UniProt IDs using home-made Perl scripts. KEGG (<http://www.kegg.jp/>) [107] pathway annotation of unigenes was performed using the BLASTX algorithm of DIAMOND program with E-values < 1.00E-5 and enabled 'sensitive mode' against KEGG gene peptide database (October, 2015), and then the corresponding KO identifiers, EC numbers and pathway categories were parsed using home-made Perl scripts.

Comparative analysis of differentially expressed genes

The number of unigenes per sample was counted if the corresponding TMM-normalized FPKM values of unigenes above zero. For differential expression analysis, each pairwise comparison was performed from the TMM-normalized FPKM values using the R package limma (Linear Models for Microarray and RNA-Seq Data) [108] with upper-quantile normalization. Zero TMM-normalized FPKM values were replaced to 0.001 from either sample to avoid problems associated with zero value. A number of statistics were calculated, including log₂-FC (fold changes) between the two samples, the *p*-values, and associated *q*-values (FDR-corrected *p*-values). Differentially expressed unigenes were selected using the criteria that the absolute value of the log₂-FC ≥ 1, the *q*-values < 0.05, and different TMM-normalized FPKM values under variant stringency. Genes with a *q*-value < 0.05 were considered to be differentially expressed [109].

Semi-quantitative RT-PCR, real-time RT-PCR analysis and SMART-RACE cDNA amplification

The cDNA products were diluted 20-fold with deionized water prior to use as a template in semi-quantitative RT-PCR and real-time PCR [110]. Semi-quantitative RT-PCR reactions were performed in 20 μ l reactions containing gene-specific primers [See Additional file 15: Table S9] and the ubiquitin gene primer as an internal control [108, 111]. Additional reaction components were 1X Red Taq Mastermix (RBC Bioscience, Taipei, Taiwan), 1 mM MgCl₂ and 1 μ M of the specific primers. Following PCR amplification, 5 μ l of the PCR products were separated on a 1 % TAE agarose gel containing ethidium bromide, and the bands were photographed under UV light using gel documentation system alpha imager EC (Alpha Innotech, Japan). Real-time RT-PCR was performed using the Power SYBR Green PCR Master Mix (Applied Biosystems, Foster City, CA, USA) and a 7900HT Fast Real-Time PCR System (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions. SDS2.2.2 software (Applied Biosystems, Foster City, CA, USA) was used for the comparative Δ C_t analysis, and the ubiquitin gene served as an internal control. The relative gene expression was calculated using the 2^{- $\Delta\Delta$ C_t} method [112]. In SMART-RACE cDNA amplification, the 5' and 3'-RACE (5' and 3'-rapid amplification of cDNA ends) was performed using the SMART-RACE cDNA amplification kit (Clontech, Palo Alto, CA, USA) according to the manufacturer's instructions. All primers used in the present study are listed in Additional file 15: Table S9.

Availability of supporting data

The data set supporting the results of this article are available in the NCBI GEO repository, with the accession numbers GSE73633 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE73633>). All the supporting data are included as additional files.

Additional files

Additional file 1: Figure S1. Gene ontology (GO) term classifications under the molecular function, biological process and cellular compartment categories at level 2 derived from the mapped unigenes using read data from (a) *Armillariella mellea*, (b) vegetative propagation corm of *Gastrodia elata*, and (c) juvenile tuber of *Gastrodia elata*. (PDF 227 kb)

Additional file 2: Figure S2. Unigenes identified solely or both for *Armillaria mellea* and juvenile tuber of *Gastrodia elata* in the comparative transcriptome analysis. (PDF 185 kb)

Additional file 3: Table S1. Sixty-nine unigenes were significantly up-regulated (\log_2 -FC ≥ 1 , q -value < 0.05 , TMM-normalized FPKM > 0.3) in *Armillaria mellea* compared with juvenile tuber of *Gastrodia elata*. (PDF 170 kb)

Additional file 4: Table S2. 223 unigenes were significantly down-regulated (\log_2 -FC ≤ -1 , q -value < 0.05 , TMM-normalized FPKM > 0.3) in *Armillaria mellea* compared with juvenile tuber of *Gastrodia elata*. (PDF 176 kb)

Additional file 5: Table S3. Mapping of KEGG biological pathways for up-regulated (\log_2 -FC ≥ 1 , q -value < 0.05 , TMM-normalized FPKM > 0.3) unigenes from *Armillaria mellea* compared to juvenile tuber of *G. elata*. (PDF 125 kb)

Additional file 6: Table S4. Mapping of KEGG biological pathways for down-regulated (\log_2 -FC ≤ -1 , q -value < 0.05 , TMM-normalized FPKM > 0.3) unigenes from *Armillaria mellea* compared to juvenile tuber of *G. elata*. (PDF 130 kb)

Additional file 7: Figure S3. Unigenes identified solely or both for vegetative propagation corm and juvenile tuber of *Gastrodia elata* in the comparative transcriptome analysis. (PDF 202 kb)

Additional file 8: Table S5. 298 unigenes were significantly up-regulated (\log_2 -FC ≥ 1 , q -value < 0.05 , TMM-normalized FPKM > 10) in juvenile tubers compared with vegetative propagation corms of *Gastrodia elata*. (PDF 183 kb)

Additional file 9: Table S6. 405 unigenes were significantly down-regulated (\log_2 -FC ≤ -1 , q -value < 0.05 , TMM-normalized FPKM > 10) in juvenile tubers compared with vegetative propagation corms of *Gastrodia elata*. (PDF 168 kb)

Additional file 10: Table S7. Mapping of KEGG biological pathways for up-regulated (\log_2 -FC ≥ 1 , q -value < 0.05 , TMM-normalized FPKM > 10) unigenes from juvenile tuber of *G. elata* compared to vegetative propagation corm of *G. elata*. (PDF 131 kb)

Additional file 11: Table S8. Mapping of KEGG biological pathways for down-regulated (\log_2 -FC ≤ -1 , q -value < 0.05 , TMM-normalized FPKM > 10) unigenes from juvenile tuber of *G. elata* compared to vegetative propagation corm of *G. elata*. (PDF 130 kb)

Additional file 12: Figure S4. *Armillaria mellea* fastq sequences checked by FastQC for (a) per base sequence quality and (b) per sequence quality scores. (a): The y-axis on the graph shows the quality scores. The higher the score the better the base call. The background of the graph divides the y axis into very good quality calls (green), calls of reasonable quality (orange), and calls of poor quality (red). The central red line is the median value. The yellow box represents the inter-quartile range (25-75 %). The upper and lower whiskers represent the 10 and 90 % points. The blue line represents the mean quality. (b): The most frequently observed mean quality below 38 is meaning that this equates to a 0.0158 % error rate. (PDF 108 kb)

Additional file 13: Figure S5. Vegetative propagation corm of *Gastrodia elata* fastq sequences checked by FastQC for (a) per base sequence quality and (b) per sequence quality scores. (a): The y-axis on the graph shows the quality scores. The higher the score the better the base call. The background of the graph divides the y axis into very good quality calls (green), calls of reasonable quality (orange), and calls of poor quality (red). The central red line is the median value. The yellow box represents the inter-quartile range (25-75 %). The upper and lower whiskers represent the 10 and 90 % points. The blue line represents the mean quality. (b): The most frequently observed mean quality below 39 is meaning that this equates to a 0.0126 % error rate. (PDF 109 kb)

Additional file 14: Figure S6. Juvenile tuber of *Gastrodia elata* fastq sequences checked by FastQC for (a) per base sequence quality and (b) per sequence quality scores. (a): The y-axis on the graph shows the quality scores. The higher the score the better the base call. The background of the graph divides the y axis into very good quality calls (green), calls of reasonable quality (orange), and calls of poor quality (red). The central red line is the median value. The yellow box represents the inter-quartile range (25-75 %). The upper and lower whiskers represent the 10 and 90 % points. The blue line represents the mean quality. (b): The most frequently observed mean quality below 39 is meaning that this equates to a 0.0126 % error rate. (PDF 109 kb)

Additional file 15: Table S9. List of primers used for SMART-RACE and Real-time PCR analysis. (PDF 80 kb)

Abbreviations

CYP450: cytochrome P450; EC: enzyme commission; EST: expressed sequence tag; FPKM: fragments per kilobase of transcript per Million mapped reads; GAFFs: *Gastrodia* antifungal proteins; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; KOids: KEGG Ortholog identifiers;

NGS: next generation sequencing; PDB: protein data bank; PIR: protein information resource; PRF: protein research foundation; RIN: RNA integrity number; RSEM: RNA-seq by expectation-maximization; TAIE: Taiwan Endemic Species Research Institute; TMM: trimmed mean of M-values; TSA: transcriptome shotgun assembly.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

CCT and YCC performed experimental design and drafted the manuscript. CCT and CYH participated in the sample collection and carried out the RNA isolation. KMW and TYC performed bioinformatics analysis and drafted the manuscript. CYH and SJL provided experimental data. YCC and CHC participated in its design, and drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This study was dedicated to the memory of Dr. C.C. Tsai, who passed away on November 05, 2015 in an unexpected stroke. Dr. C.C. Tsai was a research scientist studying genetics and breeding at the Crops Improvement Division, Kaohsiung District Agricultural Research and Extension Station, Pingtung, Taiwan. This research was supported by funding from the Ministry of Science and Technology, Taiwan (NSC-100-2621-B-067E-001-MY3) to C. C. Tsai, (NSC-100-2621-B-110-001-MY3) to Y. C. Chiang, and from the China Medical University, Taichung, Taiwan to C. H. Chou.

Author details

¹Crop Improvement Division, Kaohsiung District Agricultural Improvement Station, Pingtung 900, Taiwan. ²Graduate Institute of Biotechnology, National Pingtung University of Science and Technology, Pingtung 912, Taiwan. ³Welgene Biotech. Co., Ltd., Taipei 115, Taiwan. ⁴Department of Life Science, Cheng-Kung University, Tainan 701, Taiwan. ⁵Research Center for Biodiversity, China Medical University, Taichung 404, Taiwan. ⁶Department of Biological Sciences, National Sun Yat-sen University, Kaohsiung 804, Taiwan. ⁷Department of Biomedical Science and Environment Biology, Kaohsiung Medical University, Kaohsiung 807, Taiwan.

Received: 6 August 2015 Accepted: 22 February 2016

Published online: 09 March 2016

References

- Zhang W, Li B. The biological relationship of *Gastrodia elata* and *Armillaria mellea*. *Acta Bot Sin.* 1980;22:57–62.
- Huang H, Liang Z, Wang W. Review of nutrient source studies on *Gastrodia elata* Bl growth up. *J Northwest Sci Tech Univ Agric For.* 2004;32:145–53.
- Park E, Lee W, Ahn J. In vitro propagation of myco-heterotrophic *Gastrodia elata*. *Hortic Environ Biotech.* 2012;53:415–20.
- Liu H, Luo Y, Liu H. Studies of mycorrhizal fungi of Chinese orchids and their role in orchid conservation in China-A review. *Bot Rev.* 2010;76:241–62.
- Park E, Lee W. In vitro symbiotic germination of myco-heterotrophic *Gastrodia elata* by *Mycena* species. *Plant Biotech Rep.* 2013;7:185–91.
- Xu J, Gao S. Retrospect on the research of the cultivation of *Gastrodia elata* Bl, a rare traditional Chinese medicine. *Chinese Med J.* 2000;113:686–92.
- Xu J, Ran Y, Mou C, Wang C, Cao J, Wang M, et al. A brief report on the nutritious sources of seed germination of *Gastrodia elata*. *Bull Chin Mat Med.* 1981;6:2.
- Xu J, Ran Y, Guo S. Studies on the life cycle of *Gastrodia elata*. *Acta Acad Med Sin.* 1989;11:237–41.
- Zhou J, Yang Y, Liang H, Liu C. Morphology of *Gastrodia elata*. *Peijin: Science Publish Center;* 1987. p. 1–122.
- Liu T, Li CM, Han YL, Chiang TY, Chiang YC, Sung HM. Highly diversified fungi are associated with the achlorophyllous orchid *Gastrodia flavilabella*. *BMC Genomics.* 2015;16:185.
- Fan L, Guo S. Interaction between protocorms of *Gastrodia elata* (orchidaceae) and *Mycena dendrobii* in symbiotic germination. *Mycosystema.* 1999;18:219–25.
- Fan L, Guo S, Xiao P. Interaction between protocorms of *Gastrodia elata* (orchidaceae) and *Mycena anoetochila* during symbiotic germination. *Mycosystema.* 2001;20:539–46.
- Fan L, Guo S, Cao W, Xiao P, Xu J. Isolation, culture, identification and biological activity of *Mycena orchidicola* sp. nov. in *Cymbidium sinense* (orchidaceae). *Acta Mycol Sin.* 1996;15:251–5.
- Guo S, Fan L, Cao W, Xu J, Xiao P. *Mycena anoetochila* sp. nov. isolated from mycorrhizal roots of *Anoetochilus roxburghii* from Xishuangbanna, China. *Mycologia.* 1997;89:952–4.
- Guo S, Fan L, Cao W, Chen X. *Mycena dendrobii*, a new mycorrhizal fungus. *Mycosystema.* 1999;18:141–4.
- Xu J, Guo X. Fungus associated with nutrition of seed germination of *Gastrodia elata*-*Mycena osmundicola* Lange. *Acta Mycol Sin.* 1989;8:221–6.
- Guo S, Xu J. Studies on the cell ultrastructure in the course of *Gastrodia elata* digesting *Mycena osmundicola* LANGE and *Armillaria mellea* FR. *Acta Mycol Sin.* 1990;9:218–25.
- Kusano S. *Gastrodia elata* and its symbiotic association with *Armillaria mellea*. *J Coll Agric Imp Univ Tokyo.* 1911;4:1–65.
- Zou N, Bai X, Lu J, Yang J, Xu G, Sun D. Study on symbiotic mechanism between *Gastrodia elata* Blume and *Armillaria mellea* in tissue culture system. *Med Plant.* 2010;1:91–2.
- Lan J, Xu J, Li J. Studies on the infecting process of labelled *Armillaria mellea* to *Gastrodia elata*. *Acta Agric Nucleatae Sinica.* 1986;10:123–5.
- Xu J. The changes of cell structure in the courses of *Armillaria mellea* penetrating the nutritional stems of *Gastrodia elata*. *Acta Acad Med Sin.* 2001;23:150–3.
- Xu J, Fan L. Cytodifferentiation of the seeds (protocorms) and vegetative propagation corms colonized by mycorrhizal fungi. *Acta Bot Sin.* 2001;43:1003–10.
- Tang W, Eisenbrand G. *Gastrodia elata* B1. Chinese drugs of plant origin. Berlin/Heidelberg: Springer-Verlag; 1992. p. 545–8.
- Bensky D, Gamble A. Chinese herbal medicine: materia medica. Revised Edition. Seattle: Eastland Press; 1993. p. 971–4.
- Jung T, Suh S, Lee H, Kim I, Kim H, Yoo H, et al. Protective effects of several components of *Gastrodia elata* on lipid peroxidation in gerbil brain homogenates. *Phytother Res.* 2007;21:960–4.
- Kim D, Kim J, Han Y. Alzheimer's disease drug discovery from herbs: neuroprotectivity from β -Amyloid (1–42) insult. *J Altern Complement Med.* 2007;13:333–40.
- Feng X, Chen Y, Yang J. Studies on constituents of Tian-Ma. *Gastrodia elata* Bl. *Acta Chim Sin.* 1979;37:175–81.
- Zhou J, Yang Y, Yang J. Chemistry of *Gastrodia elata* Bl. I. Isolation and identification of chemical constituents of *Gastrodia elata* Bl. *Acta Chim Sin.* 1979;37:183–8.
- Zhao Y, Cao Q, Xiang Y, Hu Z. Identification and determination of active components in *Gastrodia elata* Bl by capillary electrophoresis. *J Chromatogr.* 1999;849:277–83.
- Zhen J. Study on actions of tuber (tian-ma) and vanillin on counteracting epilepsy. *Chin J Physiol.* 1961;24:187.
- Huh K, Kim J, Kwon T. The mechanism of anticonvulsive effect of the rhizoma of *Gastrodia elata* in pentylenetetrazole treated rats. *Yakhak Hoechi.* 1998;42:330–5.
- Hsieh C, Tang N, Chiang S, Hsieh C, Lin J. Anticonvulsive and free radical scavenging actions of two herbs, *Uncaria rhynchophylla* (MIQ) Jack and *Gastrodia elata* Bl., in kainic acid-treated rats. *Life Sci.* 1999;65:2071–82.
- Ha J, Lee D, Lee J, Kim J, Yong C, Kim J, et al. 4-Hydroxybenzaldehyde from *Gastrodia elata* B1. is active in the antioxidation and GABAergic neuromodulation of the rat brain. *J Ethnopharmacol.* 2000;73:329–33.
- Hsieh C, Chang C, Chiang S, Li T, Tang N, Pon C, et al. Anticonvulsive and free radical scavenging activities of vanillyl alcohol in ferric chloride-induced epileptic seizures in Sprague-awley rats. *Life Sci.* 2000;67:1185–95.
- Ha JH, Shin SM, Lee SK, Kim JS, Shin US, Huh K, et al. In vitro effects of hydroxybenzaldehydes from *Gastrodia elata* and their analogues on GABAergic neurotransmission, and a structure-activity correlation. *Planta Med.* 2001;67:877–80.
- Kim H, Moon K, Oh S, Kim S, Lee S. Ether fraction of methanol extracts of *Gastrodia elata*, a traditional medicinal herb, protects against kainic acid-induced neuronal damage in the mouse hippocampus. *Neurosci Lett.* 2001;314:65–8.
- Taguchi H, Yosioka I, Yamasaki K, Kim I. Studies on the constituents of *Gastrodia elata* Blume. *Chem Pharm Bull.* 1981;29:55–62.
- Huang NK, Chern Y, Fang JM, Lin CI, Chen WP, Lin YL. Neuroprotective principles from *Gastrodia elata*. *J Nat Prod.* 2007;70:571–4.
- Yang X, Zhu J, Yang R, Liu J, Li L, Zhang H. Phenolic constituents from the rhizomes of *Gastrodia elata*. *Nat Prod Res.* 2007;21:180–6.

40. Gong X, Sucher N. Stroke therapy in traditional medicine (TCM): prospects for drug discovery and development. *TIPS*. 1999;20:191–6.
41. Jin W, Tian D. Studies on chemistry and pharmacology of *Gastrodia elata*. *Chin Tradit Drug Tech*. 2000;2:21–3.
42. Yang S, Lan J, Xu J. Progress of study on *Gastrodia elata*. *Chin Tradit Herb Drugs*. 2000;31:6–69.
43. Zhou J, Yang Y, Yang C. Chemical study on gastrodin and related compounds, the syntheses of gastrodin and related compounds. *Acta Chem Sin*. 1980;38:162–5.
44. Ye H, Wang J, Wang S, Tan D. Compare seed stem with commercial rhizoma *Gastrodia* in pharmacologic effect II. *Lishizhen Med Mat Med Res*. 2003;14:730–1.
45. Zhu H, Dai P, Zhang W, Chen E, Han W, Chen C, et al. Enzymic synthesis of gastrodin through microbial transformation and purification of gastrodin biosynthesis enzyme. *Biol Pharm Bull*. 2010;33:1680–4.
46. Peng CX, Gong JS, Zhang XF, Zhang M, Zheng SQ. Production of gastrodin through biotransformation of *p*-hydroxybenzyl alcohol using hairy root cultures of *Datura tatula* L. *Afr J Biotechnol*. 2008;7:211–6.
47. Zhang H, He G, Liu J, Ruan H, Chen Q, Zhang Q, et al. Production of gastrodin through biotransformation of *p*-2-hydroxybenzyl alcohol by cultured cells of *Armillaria luteovirens* Sacc. *Enzyme Microb Technol*. 2008;43:25–30.
48. Sun C, Li Y, Wu Q, Luo H, Sun Y, Song J, et al. *De novo* sequencing and analysis of the American ginseng root transcriptome using a GS FLX Titanium platform to discover putative genes involved in ginsenoside biosynthesis. *BMC Genomics*. 2010;11:262.
49. Kouko J, Conn E. The metabolism of aromatic compounds in higher plants: IV. Purification and properties of the phenylalanine deaminase of *Hordeum vulgare*. *J Biol Chem*. 1961;236:2692–8.
50. Heiden A, Kobel K, Komenda M, Koppmann R, Shao M, Wildt J. Toluene emissions from plants. *Geophys Res Lett*. 1999;26:1283–6.
51. Singer A, Crowley D, Thompson I. Secondary plant metabolites in phytoremediation and biotransformation. *TRENDS Biotech*. 2003;21:123–30.
52. Rasi S, Veijanen A, Rintala J. Trace compounds of biogas from different biogas production plants. *Energy*. 2007;32:1375–80.
53. Zhang H. Research on production of gastrodin through biotransformation by *Armillaria luteo-virens* Sacc. Master's thesis, Zhejiang University. 2010.
54. Carmona M, Zamarro M, Blazquez B, Durante-Rodriguez G, Juarez J, Valderrama J, et al. Anaerobic catabolism of aromatic compounds: a genetic and genomic view. *Microbiol Mol Biol Rev*. 2009;73:71–133.
55. Radniecki T, Gilroy C, Sempirni L. Linking NE1545 gene expression with cell volume changes in *Nitrosomonas europaea* cells exposed to aromatic hydrocarbons. *Chemosphere*. 2011;82:514–20.
56. Jones K, Trudgill P, Hopper D. Metabolism of *p*-Cresol by the Fungus *Aspergillus fumigatus*. *Microbiology*. 1993;59:1125–30.
57. Coon M. Cytochrome P450: nature's most versatile biological catalyst. *Annu Rev Pharmacol Toxicol*. 2005;45:1–25.
58. Morant M, Bak S, Moller B, Werck-Reichhart D. Plant cytochromes P450: tools for pharmacology, plant protection and phytoremediation. *Curr Opin Biotechnol*. 2003;14:151–62.
59. Metzker M. Sequencing technologies-the next generation. *Nat Rev Genet*. 2010;11:31–46.
60. Wakaguri H, Suzuki Y, Katayama T, Kawashima S, Kibukawa E, Hiranuka K, et al. Full-Malaria/Parasites and Full-Arthropods: databases of full-length cDNAs of parasites and arthropods. *Nucleic Acids Res*. 2009;37:D520–5.
61. Hua W, Zhang Y, Song J, Zhao L, Wang Z. *De novo* transcriptome sequencing in *Salvia miltiorrhiza* to identify genes involved in the biosynthesis of active ingredients. *Genomics*. 2011;98:272–9.
62. Wang W, Wang Y, Zhang Q, Qi Y, Guo D. Global characterization of *Artemisia annua* glandular trichome transcriptome using 454 pyrosequencing. *BMC Genomics*. 2009;10:465.
63. Barrero R, Chapman B, Yang Y, Moolhuijzen P, Keeble-Gagnère G, Zhang N, et al. *De novo* assembly of *Euphorbia fischeriana* root transcriptome identifies prostratin pathway related genes. *BMC Genomics*. 2011;12:600.
64. Hao D, Ge G, Xiao P, Zhang Y, Yang L. The first insight into the tissue specific taxus transcriptome via Illumina second generation sequencing. *PLoS ONE*. 2011;6, e21220.
65. Liu S, Li W, Wu Y, Chen C, Lei J. *De novo* transcriptome assembly in chili pepper (*Capsicum frutescens*) to identify genes involved in the biosynthesis of capsaicinoids. *PLoS ONE*. 2013;8, e48156.
66. Vega-Arreguin J, Ibarra-Laclette E, Jimenez-Moraila B, Martinez O, Vielle-Calzada J, Herrera-Estrella L, et al. Deep sampling of the Palomero maize transcriptome by a high throughput strategy of pyrosequencing. *BMC Genomics*. 2009;10:299.
67. Annadurai R, Neethiraj R, Jayakumar V, Damodaran A, Rao S, Katta M, et al. *De novo* transcriptome assembly (NGS) of *Curcuma longa* L. rhizome reveals novel transcripts related to anticancer and antimalarial terpenoids. *PLoS ONE*. 2013;8:e56217.
68. Barakat A, DiLoreto D, Zhang Y, Smith C, Baier K, Powell W, et al. Comparison of the transcriptomes of American chestnut (*Castanea dentata*) and Chinese chestnut (*Castanea mollissima*) in response to the chestnut blight infection. *BMC Plant Biol*. 2009;9:51.
69. Mizrahi E, Hefer C, Ranik M, Joubert F, Myburg A. *De novo* assembled expressed gene catalog of a fast-growing *Eucalyptus* tree produced by Illumina mRNA-Seq. *BMC Genomics*. 2010;11:681.
70. Alagna F, D'Agostino N, Torchia L, Servili M, Rao R, Pietrella M, et al. Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development. *BMC Genomics*. 2009;10:399.
71. Shi C, Yang H, Wei C, Yu O, Zhang Z, Jiang C, et al. Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds. *BMC Genomics*. 2011;12:131.
72. Wang Z, Fang B, Chen J, Zhang X, Luo Z, Huang L, et al. *De novo* assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweet potato (*Ipomoea batatas*). *BMC Genomics*. 2010;11:726.
73. Weber A, Weber K, Carr K, Wilkerson C, Ohlrogge J. Sampling the Arabidopsis transcriptome with massively parallel pyrosequencing. *Plant Physiol*. 2007;144:32–42.
74. Su C, Chao Y, Chang Y, Chen W, Chen C, Lee A, et al. *De novo* assembly of expressed transcripts and global analysis of the *Phalaenopsis aphrodite* transcriptome. *Plant Cell Physiol*. 2011;52:1501–14.
75. Feng C, Chen M, Xu C, Bai L, Yin X, Li X, et al. Transcriptomic analysis of Chinese bayberry (*Myrica rubra*) fruit development and ripening using RNA-Seq. *BMC Genomics*. 2012;13:19.
76. Kudapa H, Bharti A, Cannon S, Farmer A, Mulaosmanovic B, Kramer R, et al. A comprehensive transcriptome assembly of *Pigeonpea* (*Cajanus cajan* L.) using sanger and second-generation sequencing platforms. *Mol Plant*. 2012; 5:1020–8.
77. Ranjan A, Ichihashi Y, Farhi M, Zumstein K, Townsley B, David-Schwartz R, et al. *De novo* assembly and characterization of the transcriptome of the parasitic weed *Cuscuta pentagona* identifies genes associated with plant parasitism. *Plant Physiol*. 2014;166:1186–99.
78. Salgado L, Koop D, Pinheiro D, Rivalan R, Guen V, Nicolás M, et al. *De novo* transcriptome analysis of *Hevea brasiliensis* tissues by RNA-seq and screening for molecular markers. *BMC Genomics*. 2014;15:236.
79. Wang Y, Pan Y, Liu Z, Zhu X, Zhai L, Xu L, et al. *De novo* transcriptome sequencing of radish (*Raphanus sativus* L.) and analysis of major genes involved in glucosinolate metabolism. *BMC Genomics*. 2013;14:836.
80. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010;28:511–5.
81. Wang HX, Yang T, Zeng Y, Hu Z. Identification of the gastrodin gene *ga4B* in an achlorophyllous plant *Gastrodia elata* Orchidaceae and its expression pattern during life cycle. *Acta Bot Yunnanica*. 2007;29:345–50.
82. Hu Z, Yang Z, Wang J. Isolation and partial characterization of an antifungal protein from *Gastrodia elata* corm. *Acta Bot Yunnan*. 1988;10:373–80.
83. Shiau Y, Ho W. Method for extraction of gastrodin in *Anoectochilus formosanus* Hayata. *J Taiwan Agric Res*. 2012;61:259–68.
84. Bowles D, Lim E, Poppenberger B, Vaistij F. Glycosyltransferases of lipophilic small molecules. *Annu Rev Plant Biol*. 2006;57:567–97.
85. Gachon C, Langlois-Meurinne M, Saindrenan P. Plant secondary metabolism glycosyltransferases: the emerging functional analysis. *Trends Plant Sci*. 2005;10: 542–9.
86. Lairson L, Henrissat B, Davies G, Withers S. Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem*. 2008;77:521–55.
87. Kozak M. Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J Biol Chem*. 1991;266:19867–70.
88. Morris DR, Geballe AP. Upstream open reading frames as regulators of mRNA translation. *Mol Cell Biol*. 2000;20:8635–42.
89. Mendell JT, Sharifi NA, Meyers JL, Martinez-Murillo F, Dietz HC. Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat Genet*. 2004;36:1073–8.

90. Wittmann J, Hol EM, Jack H-M. hUPF2 silencing identifies physiologic substrates of mammalian nonsense-mediated mRNA decay. *Mol Cell Biol*. 2006;26:1272–87.
91. Yepiskoposyan H, Aeschmann F, Nilsson D, Okoniewski M, Muhlemann O. Autoregulation of the nonsense-mediated mRNA decay pathway in human cells. *RNA*. 2011;17:2108–18.
92. Watatani Y, Ichikawa K, Nakanishi N, Fujimoto M, Takeda H, Kimura N, et al. Is regulated by the 5'-untranslated region stress-induced translation of ATF5 mRNA. *J Biol Chem*. 2008;283:2543–53.
93. Spriggs KA, Bushell M, Willis AE. Translational regulation of gene expression during conditions of cell stress. *Mol Cell*. 2010;40:228–37.
94. Barbosa C, Romao L. Translation of the human erythropoietin transcript is regulated by an upstream open reading frame in response to hypoxia. *RNA*. 2014;20:594–608.
95. Zhao C, Datta S, Mandal P, Xu S, Hamilton T. Stress-sensitive regulation of IFRD1 mRNA decay is mediated by an upstream open reading frame. *J Biol Chem*. 2010;285:8552–62.
96. Hatano M, Umemura M, Kimura N, Yamazaki T, Takeda H, Nakano H, et al. The 5'-untranslated region regulates ATF5 mRNA stability via nonsense-mediated mRNA decay in response to environmental stress. *FEBS J*. 2013;280:4693–707.
97. Mishra A, Dubey N. Evaluation of some essential oils for their toxicity against fungi causing deterioration of stored food commodities. *Appl Environ Microbiol*. 1994;60:1101–15.
98. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
99. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011;29:644–52.
100. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011;12:323.
101. Langmead B, Trapnell C, Pop M, Salzberg S. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10:R25.
102. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*. 2010;11:R25.
103. Dillies MA, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, Servant N, et al. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief Bioinform*. 2013;14:671–83.
104. Edgar R, Domrachev M, Lash AE. Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. 2002;30:207–10.
105. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods*. 2015;12:59–60.
106. Huang H, McGarvey PB, Suzek BE, Mazumder R, Zhang J, Chen Y, et al. A comprehensive protein-centric ID mapping service for molecular data integration. *Bioinformatics*. 2011;27:1190–1.
107. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2007;28:27–30.
108. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43, e47.
109. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc B Methodol*. 1995;57:289–300.
110. García-Vallejo J, Van het Hof B, Robben J, Van Wijk J, Van Diea I, Joziassse D, et al. Approach for defining endogenous reference genes in gene expression experiments. *Anal Biochem*. 2004;329:293–9.
111. Pfaffl M, Tichopad A, Prgomet C, Neuvians T. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper – Excel-based tool using pair-wise correlations. *Biotechnol Lett*. 2004;26:509–15.
112. Livak K, Schmittgen T. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods*. 2001;25:402–8.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

