Evolutionary and functional analysis of mulberry type III polyketide synthases

Han Li, Jiubo Liang, Hu Chen, Guangyu Ding, Bi Ma and Ningjia He*

Abstract

Background: Type III polyketide synthases are important for the biosynthesis of flavonoids and various plant polyphenols. Mulberry plants have abundant polyphenols, but very little is known about the mulberry type III polyketide synthase genes. An analysis of these genes may provide new targets for genetic improvement to increase relevant secondary metabolites and enhance the plant tolerance to biotic and abiotic stresses.

Results: Eighteen genes encoding type III polyketide synthases were identified, including six chalcone synthases (CHS), ten stilbene synthases (STS), and two polyketide synthases (PKS). Functional characterization of four genes representing most of the MnCHS and MnSTS genes by coexpression with 4-Coumaroyl-CoA ligase in Escherichia coli indicated that their products were able to catalyze p-coumaroyl-CoA and malonyl-CoA to generate naringenin and resveratrol, respectively. Microsynteny analysis within mulberry indicated that segmental and tandem duplication events contributed to the expansion of the MnCHS family, while tandem duplications were mainly responsible for the generation of the MnSTS genes. Combining the evolution and expression analysis results of the mulberry type III PKS genes indicated that MnCHS and MnSTS genes evolved mainly under purifying selection to maintain their original functions, but transcriptional subfunctionalization occurred during long-term species evolution. Moreover, mulberry leaves can rapidly accumulate oxyresveratrol after UV-C irradiation, suggesting that resveratrol was converted to oxyresveratrol.

Conclusions: Characterizing the functions and evolution of mulberry type III PKS genes is crucial for advancing our understanding of these genes and providing the basis for further studies on the biosynthesis of relevant secondary metabolites in mulberry plants.

Keywords: Chalcone synthase, Stilbene synthase, Functional analysis, Evolutionary analysis, Gene expression, Mulberry

Background

Mulberry (Morus spp.), belonging to the order Rosales, family Moraceae, and genus Morus, has been used as a food product and in herbal medicines in China for over 1,900 years [1]. Different plant parts, such as leaves, fruits, branches, bark, roots, and shoots, have a variety of pharmacological effects, including preventing headaches, hypertension, and diabetes, in addition to act as a diuretic [2]. The pharmacological effects of mulberry are attributed to bioactive compounds in tissues. Flavonoids and stilbenes are representative bioactive compounds with a wide range of activities, including UV protection, defense against herbivores and pathogens, and protection from abiotic stresses [3–5]. Their potential medicinal properties have also resulted in their use in treatments for a variety of diseases. For example, anthocyanins, which are a group of water-soluble pigments in mulberry fruits, are beneficial to humans in treatments for obesity, inflammation, and cancer [6]. Oxyresveratrol, which is a representative mulberry stilbene, has been used in cosmetics and for the treatment of hyperpigmentation disorders [7].

There are two pathways for the biosynthesis of flavonoids and stilbenes, which begins with the condensation of one CoA-ester molecule of cinnamic acid or derivatives such as coumaric or ferulic acid, and three molecules of malonyl-CoA (Additional file 1: Figure S1) [8]. The generated tetraketide intermediate is subsequently folded and cyclized to form a chalcone or stilbene ring structure depending on the polyketide synthase activities of chalcone.
synthase (CHS) or stilbene synthase (STS) [9]. Chalcone synthase catalyzes a C6 → C1 Claisen condensation of the elongated tetraketide intermediate to form naringenin chalcone, while STS catalyzes an alternative C2 → C7 aldol condensation of the same tetraketide intermediate to produce a stilbene backbone [10]. Both CHS and STS belong to the plant type III polyketide synthase superfamily, and share considerable similarities in their sequences and crystallographic structures, suggesting that STS independently evolved from CHS several times [11, 12]. Unlike CHS, which is present in all higher plants, STS is restricted to a few species such as peanut (Arachis hypogaea), Scots pine (Pinus sylvestris), and grapevine (Vitis vinifera) [13–15]. Besides CHS and STS, a growing number of functionally divergent CHS-like type III PKS genes have been cloned and characterized in plants. For example, 2-pyrone synthase (2-PS) from Gerbera hybrida catalyzes the formation of 6-methyl-4-hydroxy-2-pyrone [16]. Acriflavone synthase (ACS) from Ruta graveolens catalyzes three condensations of malonyl-CoA with an N-methylanthraniloyl-CoA starter to form a three-ring acridone skeleton of acridone alkaloids [17]. Other members in plant type III polyketide synthase superfamily, such as benzalacetone synthase (BAS), styrylpyrone synthase (SPS), and 4-coumaroyl triacetic acid synthase (CTAS), also catalyze iterative decarboxylative condensations of malonyl unit with a CoA-linked starter molecule to produce structurally diverse, pharmaceutically important plant secondary metabolites [10].

It is clearly evident that gene families arose by gene duplication and subsequent sequence divergence [18]. In some cases, more gene family members help to increase the abundance of enzymes and/or proteins [19]. In other instances, duplicate genes can lead to new functions [20].

Although flavonoids and stilbenes have important roles in mulberry plants, little is known about their biosynthesis in these plants. In this study, we identified and analyzed type III polyketide synthases in the genome of the mulberry tree Morus notabilis [21]. Our analyses of gene organization and expression, molecular evolution, and functions have generated new insights into the roles of mulberry type III polyketide synthases, and have laid the foundation for future studies on secondary metabolite biosynthesis in mulberry plants.

Results
Identification and phylogenetic analysis of mulberry type III polyketide synthases
Twenty-one putative type III polyketide synthase genes (PKS) were identified based on the recently sequenced M. notabilis genome (see Methods). The genes included six CHS (MnCHS1–6) genes, ten STS (MnSTS1–10) genes, two PKS genes (MnPKS1–2), and three genes with undetectable expression (MnCHSL1–3) (Fig. 1, Additional file 2: Table S1). In addition, six pseudogenes that were disrupted by stop codons, frame-shifts, and/or small deletions were also identified (data are not shown). For MnCHS gene family, MnCHS1 and MnCHS2 were identified and analyzed in one of our previous studies [22]. The deduced length of the enzymes encoded by these genes was 389 amino acids. Most of the remaining four MnCHS genes encoded a 391-amino acid enzyme,
except MnCHS4, whose enzyme was predicted to consist of 394 amino acids. The MnSTS genes encoded enzymes with 399 residues, and exhibited a highly conserved gene structure. All MnSTS genes contained two exons separated by an intron, following the GT/AG rule (Fig. 1).

Phylogenetic analyses revealed that mulberry CHS and STS clustered together with other plant CHS or STS homologs (Fig. 2). Each mulberry CHS and STS family can be divided into two subgroups, namely MnCHS-A, MnCHS-B, MnSTS-A, and MnSTS-B. Similar phylogenetic analysis result based on nucleotide sequences of mulberry type III PKSs also supported such clustering (Fig. 1). MnCHSL2 and MnCHSL3 located in plant CHS/STS clade were further clustered with mulberry STS proteins, suggesting they shared a closer evolutionary relationship. MnCHSL1, which was outside the CHS/STS clade, shared more sequences identical with MnCHSs (66.84 % ~ 68.21 %) than MnSTSs (56.52 % ~ 57.54 %). Two mulberry PKS proteins were located in a PKS-A/B clade containing three PKSs from Arabidopsis thaliana, which produce long-chain alkyl α-pyrones and participate in the biosynthesis of sporopollenin [23]. This PKS-A/B clade was distant to the plant CHS/STS cluster, but close to the Physcomitrella patens 2'-oxoalkylresorcinol synthase, which is considered the most recent common ancestor of the plant CHS family [24].

The amino acid sequences of MnPKS1 and MnPKS2 were 64.7 % identical, and were more similar to those of AtPKS-A (78.12 %) and AtPKS-B (82.35 %), respectively, than to those of other plant type III polyketide synthases (less than 40 %). MnPKS1 and MnPKS2 contained CHS active site residues, including Gly211, Phe215, Gly216, Gly256, Phe265, Ser338, and Pro375, as well as conserved catalytic residues such as Cys-164, His-303, and Asn-336 (Additional file 3: Figure S2) [8]. The Thr197 residue was substituted by Gly in MnPKS1 and MnPKS2 (Additional file 3: Figure S2), which was consistent with AtPKS-A and AtPKS-B sequences. This residue regulates product chain length by sterically modifying the active-site cavity, suggesting the function of MnPKS1 and MnPKS2 is related to that of the AtPKS-A and AtPKS-B enzymes [25–27].

Expansion patterns of mulberry type III PKS gene families

MnCHS6 shared a substantial colinear region with MnPKS1, MnCHSL1, and MnCHS2/MnCHS1 (Fig. 3a and c), suggesting these genes are linked by segmental duplication events. Conserved colinear genes surrounding MnPKS1, MnCHS2/MnCHS1, and MnCHS3/MnCHS4 were also detected. Less conserved colinear genes were detected between segments D and F. Although segments H and G contained a less conserved colinear region, nine of ten MnSTS genes originated in tandem arrays, suggesting tandem duplications were primarily responsible for generating the MnSTS family. MnSTS genes located on scaffold-643 were mixed with pseudogenes and some relics of transposable elements. The predicted synteny relationships and levels of synonymous substitution (Ks) in duplicated pairs (Additional file 2: Table S2) indicated that MnPKS1 and MnCHS6 arose from a single ancient duplication event, and MnCHSL1 arose after this duplication event. A detectable colinear relationship among the D, E, and A segments suggests MnCHS2/CHS1 and MnCHS3/CHS4 evolved from MnPKS1 after the appearance of MnCHSL1. Segments G and H, as outliers, lacked significant colinear relationships with the other segments (Additional file 2: Table S2). Additionally, based on the Ks values of all pairs of protein-coding genes among all genomic fragments, we used synonymous substitution rate of plant nuclear genes (5 × 10⁻⁹) to estimate the time when the segmental duplication events took place [28]. MnCHSL1 and MnCHS6 diverged about 192 million years ago (mya), which is approximate the divergence of dicots and monocots (around 200 mya) and is much earlier than the divergence times of the remaining MnPKS genes, including MnPKS1 and MnCHS2/MnCHS1 (117 mya), MnCHS3/MnCHS4 and MnPKS1 (93 mya), MnCHS2 and MnSTS10 (77 mya), and MnSTS3 and MnSTS8 (37 mya) (Fig. 3b, Additional file 2: Table S2) [29].

Selection constraints on the coding sequence of mulberry type III PKS

To explore how selective constraints evolved following the duplication of mulberry type III PKS genes, we calculated the $d_N/d_S$ ratio ($\omega$) using various codon substitution models, including the Branch, Clade, and Branch site models. Two data sets were analyzed independently. One set includes PKS, CHS and STS sequences from different plant species (Additional file 4: Figure S3). The other set is restricted to the type III PKS sequences identified in the mulberry genome (Fig. 1). As presented in Table 1, comparisons of complex Branch models with simpler models (i.e., MnCHS and M0) and nested models (i.e., MnCHS&MnSTS and MnCHS-MnSTS) indicated that the $\omega$ values for MnCHS and MnSTS sequences in mulberry sequence set (MSS) were $<1$: \(MnCHS \quad \omega = 0.0689\) [likelihood ratio test (LRT) \(P\) value approximately 1.721 × 10⁻⁵] and \(MnSTS \quad \omega = 0.1630\) (LRT \(P\) value approximately 8.220 × 10⁻⁸) in simpler models; \(MnCHS \quad \omega = 0.0703\) and \(MnSTS \quad \omega = 0.1613\) (LRT \(P\) value approximately 2.290 × 10⁻⁷) in nested models. These results were confirmed by the branch model analyses of the other specie sequence set (OSSS). According to these data, the $\omega$ of \(MnSTS\) was higher than that of \(MnCHS\), which was consistent with the situation for other specie sequences \(\omega\) other \(STS = 0.11261\), \(\omega\) other \(CHS = 0.03729\), all the LRT \(P\) values were much lower than 0.05, except the M0 model in OSSS (\(P\) value = 0.142]). Because Branch models make the unrealistic assumption of among-site homogeneity [30], we used the Clade and Branch site models in our
Fig. 2 (See legend on next page.)
analyses. The Branch site model detected a very small proportion of p2a (purifying-to-positive selection) and p2b (neutral-to-positive selection) sites in MnCHS (MSS: p2a = 0.04242, p2b = 0.00262; OSSS: p2a = 0.05719, p2b = 0.00151) and MnSTS (MSS: p2a = 0.12927, p2b = 0.00787; OSSS: p2a = 0.12654, p2b = 0.00331). Notably, the foreground ω of the MnCHS and MnSTS gene family in two data sets was 1, which was in agreement with the null model with fixed ω2 = 1. Similarly, Clade model C revealed that the majority of sites in the MnCHS and MnSTS sequences were under purifying selection. In the MSS, about 51.52 % of the p2 sites corresponded to codons that had experienced divergent selection pressures in different partitions. The ω2 of the class sites were 0.12011 for MnCHS, 0.28349 for MnSTS, and 0.13883 for the background lineages. In the OSSS analysis, this category included 46.76 % of the sites in MnCHS&MnSTS model with ω2 values of 0.17857 (MnCHS) and 0.28333 (MnSTS). The remaining two alternative hypothesis models also supported these results. The ω2 of p2 sites in MnCHS&coCHS model were 0.15212 for MnCHS and 0.06353 for other CHS. While in MnSTS&coSTS model, the ω2 of p2 sites were 0.26885 for MnSTS and 0.19651 for other STS. These results were consistent with the branch model results presented above and statistically significant when the LRT was completed using the M2a_rel null model (all P values were much less than 0.05). Moreover, according to the Branch site model, the foreground branch of MnCHSL2 lacked p2a or p2b sites in

![Fig. 2](See figure on previous page.)

**Fig. 2** Phylogenetic analysis of mulberry type III polyketide synthases and other type III polyketide synthases. The indicated scale represents 0.1 amino acid substitutions per site. ALS, aloesone synthase; BBS, bibenzyl synthase; BPS, benzophenone synthase; STCS, stilbene carboxylic synthase; VPS, valerophenone synthase; ORS, 2′-oxoalkylresorcinol synthase. The asterisk indicates the branch that is the most recent common ancestor to the plant CHS family. The ω-ketoacyl carrier protein synthase III (FABH) of *Escherichia coli* was used as an outgroup.

![Fig. 3](See figure on previous page.)

**Fig. 3** Evolutionary relationships among mulberry type III polyketide synthases. **a** Genes represented by a series of triangles of the same color are from the same genomic fragment. Genes from the type III PKS superfamily are indicated by black triangles. Triangles also indicate gene orientation on strands. The colinear homologous genes are linked by gray lines. **b** Evolutionary relationships were determined to clarify the order of duplication events for the type III polyketide synthase superfamily. The black vertical line indicates tandem duplication. Numbers indicate when (million years ago, MYA) the segmental duplication events occurred. **c** Organization of the mulberry type III PKS gene clusters. The long terminal repeat retrotransposons are presented in yellow. Right and left arrows indicate whether a retrotransposon is located on the + or − strand, respectively. RLG: retrotransposon belongs to the Gypsy superfamily; RLL: retrotransposon belongs to the Lard superfamily. Numbers represent the sequence ID.
Table 1  Parameter Estimates of Codon-Substitution Evolutionary Models for MnCHS and MnSTS family

| Model          | Proportion of sites | dN/dS ratios   | Likelihood          |
|----------------|---------------------|----------------|---------------------|
|                |                     |                | lnL                 | P-value             |
| Other specie sequence set |                     |                |                     |
| Branch Model   |                     |                |                     |
| MnCHS          | \( \omega_b = 0.07855 \) STS + CHSL + oCHS | \( \omega_f = 0.09698 \) MnCHS | \( -22140.77149 \) | 0.1418015 (MnCHS V.S M0) |
| MnSTS          | \( \omega_b = 0.07363 \) CHS + CHSL + oSTS | \( \omega_f = 0.16568 \) MnSTs | \( -22222.54872 \) | 5.192e-010 (MnSTS V.S M0) |
| oCHS           | \( \omega_b = 0.09719 \) STS + CHSL + MnCHS | \( \omega_f = 0.03729 \) oCHS | \( -22110.96154 \) | 3.886e-015 (oCHS V.S M0) |
| oSTS           | \( \omega_b = 0.07264 \) CHS + CHSL + MnSTS | \( \omega_f = 0.11261 \) oSTS | \( -22132.52527 \) | 1.570e-005 (oSTS V.S M0) |
| MnCHS&MnSTS    | \( \omega_b = 0.07079 \) CHSL + oCHS + oSTS | \( \omega_f = 0.09885 \) MnCHS | \( -22119.93911 \) | 3.595e-011 (MnCHS&MnSTS V.S M0) |
| MnCHS-MnSTS    | \( \omega_b = 0.07047 \) CHSL + oCHS + oSTS | \( \omega_f = 0.13084 \) MnCHS-MnSTS | \( -22124.16361 \) | 2.721e-009 (MnCHS-MnSTS V.S M0) |
| M0             | \( \omega_0 = 0.08027 \) |                     | \( -22141.85065 \) |                     |
| Clade Model C  |                     |                |                     |
| MnCHS&MnSTS    | \( p_0 = 0.51866 \) \( p_1 = 0.01377 \) \( p_2 = 0.46757 \) | \( \omega_0 = 0.22173 \) CHSL + CHS + oSTS \( \omega_f_1 = 0.22173 \) MnCHS \( \omega_f_2 = 0.22173 \) MnSTs | \( -21704.83735 \) | 1.321e-006 (MnCHS&MnSTS V.S M2a_rei) |
| MnCHS&oCHS     | \( p_0 = 0.46637 \) \( p_1 = 0.01890 \) \( p_2 = 0.51473 \) | \( \omega_0 = 0.02016 \) CHSL + STS \( \omega_f_1 = 0.02016 \) MnCHS \( \omega_f_2 = 0.02016 \) oCHS | \( -21695.89064 \) | 1.719E-10 |
| MnSTS&oSTS     | \( p_0 = 0.47783 \) \( p_1 = 0.01786 \) \( p_2 = 0.50430 \) | \( \omega_0 = 0.01889 \) CHSL + CHS \( \omega_f_1 = 0.01889 \) MnSTs \( \omega_f_2 = 0.01889 \) oSTS | \( -21693.48253 \) | 1.547E-11 |
### Table 1 Parameter Estimates of Codon-Substitution Evolutionary Models for MnCHS and MnSTS family (Continued)

| Model | Parameters | Estimates | Log Likelihood |
|-------|------------|-----------|----------------|
| **M2a_rel** | **Branch Site Model** | | |
| MnCHS | \( \omega^0 = 0.07409 \) | \( \omega_1^1 = 1.00000 \) | \( -22057.84405 \) |
| MnCHS(NULL) | | | | |
| MnSTS | \( \omega^0 = 0.07147 \) | \( \omega_1^1 = 1.00000 \) | \( -22040.53516 \) |
| MnSTS(NULL) | | | | |

| Model | Parameters | Estimates | Log Likelihood |
|-------|------------|-----------|----------------|
| **MnCHS (NULL)** | | | | |
| **MnSTS (NULL)** | | | | |

| Model | Parameters | Estimates | Log Likelihood |
|-------|------------|-----------|----------------|
| **Mulberry sequence set** | **Branch Model** | | |
| MnCHS | \( \omega^0 = 0.07834 \) | \( \omega_1^1 = 1.00000 \) | \( -9829.594816 \) |
| MnSTS | \( \omega^0 = 0.07834 \) | \( \omega_1^1 = 1.00000 \) | \( -9745.715191 \) |

| Model | Parameters | Estimates | Log Likelihood |
|-------|------------|-----------|----------------|
| **Clade Model C** | | | | |
| MnCHS&MnSTS | \( \omega^0 = 0.02006 \) | \( \omega_1^1 = 1.00000 \) | \( -9734.263087 \) |
| MnCHS-MnSTS | \( \omega^0 = 0.010035 \) | \( \omega_1^1 = 1.00000 \) | \( -9900.160494 \) |

| Model | Parameters | Estimates | Log Likelihood |
|-------|------------|-----------|----------------|
| **M2a_rel** | **Branch Site Model** | | |
| MnCHS | \( \omega^0 = 0.07834 \) | \( \omega_1^1 = 1.00000 \) | \( -9829.594816 \) |
| MnCHS(NULL) | | | | |
| MnSTS | \( \omega^0 = 0.07834 \) | \( \omega_1^1 = 1.00000 \) | \( -9745.715191 \) |
| MnSTS(NULL) | | | | |

**Li et al. BMC Genomics (2016) 17:540**
Table 1 Parameter Estimates of Codon-Substitution Evolutionary Models for MnCHS and MnSTS family (Continued)

| Model       | p0     | p1     | p2a    | p2b    | ω0b    | ω1b    | ω2ab   | ω2bb,2af,2bf | lnL     | p-value     |
|-------------|--------|--------|--------|--------|--------|--------|--------|--------------|---------|-------------|
| MnCHS(NULL) | 0.81332| 0.04954| 0.12927| 0.00787| 0.06667| 1.00000| 0.06667| 1            | 9829.594816 | 9813.87732  |
| MnSTS       | 0.81332| 0.04954| 0.12927| 0.00787| 0.06667| 1.00000| 0.06667| 1            | 9813.87732  | 9813.87732  |
| MnSTS(NULL) | 0.84861| 0.06802| 0.07719| 0.00619| 0.05978| 1.00000| 0.05978| 1            | 8214.284472 | 3.626E-06   |
| MnCHS-A     | 0.83646| 0.06502| 0.09141| 0.00711| 0.05774| 1.00000| 0.05774| 1            | 8225.011056 |             |
| MnSTS-A     | 0.83361| 0.06959| 0.08934| 0.00746| 0.05962| 1.00000| 0.05962| 1            | 8219.791085 | 3.684E-02   |
| MnCHS-A(NULL)| 0.81036| 0.06741| 0.11285| 0.00939| 0.05731| 1.00000| 0.05731| 1            | 8221.968483 |             |

The LTR tests are as follow: Branch Modl. One ratio Model 0 vs Two ratio Model 2. Clade Modl. C, M2α_rel vs. CmC. Branch Site Modl. Model A null (w is fixed as 1) vs. Model A. w = dN/dS. b = background; f = foreground; o = other; p0, p1, p2 = proportion of sites with dN/dS ratios = w0, w1, w2, respectively. lnL = ln of the likelihood; p-value = p-value of the likelihood ratio test.
both data sets, suggesting selective constraints did not decrease (Additional file 2: Table S3). The ω for the foreground branch of MnCHSL1 (MSS: ωMnCHSL1 = 0.1162; OSSS: ωMnCHSL1 = 0.1302) was higher than that of MnCHS lineages (MSS: ωMnCHS = 0.0689; OSSS: ωMnCHS = 0.0970) according to the Branch model. The proportion of p2 sites (p2 = 1 - p0 - p1) for this foreground branch (MSS: 0.1; OSSS: 0.11) was still higher than that of MnCHS lineages (MSS: 0.05; OSSS: 0.06) based on the Branch site model. Similarly, Branch model data indicated the ω for MnCHSL3 was much higher than that of MnSTS genes, and MnCHSL3 had a higher ω (MSS: 2.885; OSSS: 1.9191) in the Branch site model. The LRT values calculated with this model were not significantly higher than those of the null model, which means positive selection did not occur in this branch. These results indicate that the selective constraints of MnCHSL1 and MnCHSL2 were relaxed compared to those of the other genes in the same clades (Fig. 1).

Because STS may have evolved independently from CHS several times, positive selection was suspected along the branch ancestral to the MnSTS family clade, which corresponds to the foreground branch in our tests (branch b, Fig. 1). Branch site analyses revealed the ancestral branch of the MnSTS family experienced strong positive selection (ω approximately 210.99; LRT P values <0.05). Bayes empirical Bayes methods were used to calculate the posteriori probability of sites that experienced positive selection. There were seven amino acid sites in branch b with a posteriori probability >0.9 (Additional file 3: Figure S2). ConSurf was then used to calculate the evolutionary conservation of amino acid positions in MnSTS enzymes [31]. All seven positive selection sites were highly conserved, and the majority of these sites existed on the external surface of the MnSTS dimer (Fig. 4b). The positive selection sites Ser-212 and Pro-277 (numbering is based on MnSTS) are associated with CoA-binding and polyketide synthase functional diversity, respectively. This suggests they likely have major roles in MnSTS evolution [8].

**Functional divergence of the mulberry type III PKS**

Functional divergence of genes is related to the site-specific change in the corresponding protein sequence during evolution [32]. Therefore, functional divergence among the clusters of mulberry type III PKS was inferred by posterior analysis using the DIVERGE program [33]. Both θ1 of type I and θ2 of type II functional divergence for the MnSTS subfamily were almost 0 (Additional file 2: Table S4). The results indicated that the coefficients of type I functional divergence (θ1) between the MnCHS and MnSTS families were statistically significant (P <0.01), with a θ1 value of 0.493. Type II functional divergence (θ2) between the MnCHS and MnSTS families was also evident with a θ2 value of 0.176 (P <0.01). A total of seven potential type I sites were detected (posterior probability >0.75). These sites were well conserved in MnCHS clusters, but highly variable in MnSTS clusters (Additional file 3: Figure S2). Of the 31 potential type II sites (posterior probability >4), six were also under positive selection (Fig. 4a and b).

**Functional characterization of selected type III polyketide synthases**

The activities of selected CHS and STS enzymes were assayed *in vivo* by the coexpression with 4-coumaroyl-CoA ligase (4CL) in *Escherichia coli*. 4CL activates 4-coumaric acid to produce 4-coumaroyl-CoA (Additional file 1: Figure S1). When adding phenolic acid starter units to the medium containing transformed *E. coli*, phenylpropionic acids can be efficiently converted to flavonoids or stilbene compounds and secreted into the medium [34, 35]. Four genes, MnCHS2, MnCHS6, MnSTS7, and MnSTS8, were selected for functional analyses, which represented MnCHS-A, MnCHS-B, MnSTS-A, and MnSTS-B groups, respectively (Additional file 5: Figure S4). The addition of p-coumaric acid to the growth medium containing *E. coli* cells harboring Mn4CL and MnCHS resulted in the production of naringenin. Analyses of the medium using LC-MS revealed that the mass of a parent ion (M+H) matched an authentic naringenin standard (m/z 271.1) (Fig. 5). Similarly, the expected product, resveratrol (m/z 227.1), was also detected after p-coumaric acid was added to the growth medium containing *E. coli* cells harboring Mn4CL and MnSTS. Naringenin and resveratrol peaks were not observed for the medium in which *E. coli* cells carrying only Mn4CL were grown (control) (Fig. 5). The MnPKS1 and MnPKS2 enzymatic activities were examined using the same method. p-Coumaric acid was not converted to the corresponding naringenin or resveratrol by either MnPKS1 or MnPKS2 (Additional file 6: Figure S5).

In an effort to know the subcellular localization of these genes, their proteins fused with EGFP were transiently expressed in tobacco leaf epidermal cells. As shown in Additional file 7: Figure S6, the fluorescent signal of EGFP alone was detected in the nucleus and cytoplasm. In the cells transformed with MnPKS1-EGFP, MnPKS2-EGFP, MnCHS2-EGFP, MnCHS6-EGFP, MnSTS7-EGFP and MnSTS8-EGFP, fluorescence signals were only detected in the cytoplasm, suggesting that these six proteins were localized to the cytoplasm.

**Expression of mulberry type III PKS genes in different tissues**

All expressed mulberry type III PKS genes were grouped into five clusters using K-medians clustering (Fig. 6).
The genes in clusters A, C, D, and E were more highly expressed in five tissues than the genes of cluster B, which were expressed more in leaves or male flowers. *MnCHS3* in cluster A was expressed in leaves, male flowers, and winter buds. *MnCHS6* in cluster E was highly expressed in roots and the bark. Mulberry CHS genes in cluster D (i.e., *MnCHS2, MnCHS4,* and *MnCHS5*) exhibited relatively high expression levels in roots, male flowers, and the bark. The majority of *MnSTS* genes (i.e., 8 of 10) were expressed in roots and the bark. Only two *MnSTS* genes had a distinct expression profile, namely *MnSTS7* in cluster A with abundant expression in male flowers and leaves, and *MnSTS5* in cluster C with high levels of expression in the bark and male flowers.

**Effect of UV-C irradiation on the expression of mulberry type III PKS genes and biosynthesis of the corresponding compounds in mulberry leaves**

The expression of type III PKS genes was analyzed in leaves treated with ultraviolet (UV) light. The expression of *MnPKS1* was relatively stable following exposure to UV-C radiation, with a slight decrease 18 h after exposure to the dark (Fig. 7). *MnPKS2* expression increased slightly following UV-C irradiation, and remained stable in the dark. For the *MnCHS* family, UV irradiation resulted in a considerable increase in the expression of three genes in leaves, followed by a decrease after 12 h, especially for *MnCHS5*, whose expression increased about $3.5 \times 10^{-5}$-fold 12 h after UV-C treatment. The expression of *MnCHS4* was highest 24 h after UV-C treatment, and exposure to the dark may have also increased the expression of this gene. In contrast, exposure to UV-C and the dark resulted in a gradual decrease in *MnCHS3* expression. *MnCHS1* was the only gene whose expression was barely detectable even after UV-C exposure. The expression pattern of *MnSTS1* differed from that of the other *MnSTS* genes, with fluctuating profiles that peaked at 24-h after UV-C treatment. The expression levels for the other *MnSTS* genes were highest 12 h after exposure to UV-C radiation. Additionally, treatment in the dark induced the expression of *MnSTS* genes.
Fig. 5 In vivo characterization of CHS and STS by coexpression with 4-coumaroyl-CoA ligase in Escherichia coli. a In vivo assay of CHS. 4-Coumaric acid (1) and naringenin (2) were used as standard compounds. Mn4CL: E. coli-expressed Mn4CL. Mn4CL + MnCHS2: E. coli-coexpressed Mn4CL and MnCHS2. Mn4CL + MnCHS6: E. coli-coexpressed Mn4CL and MnCHS6. Selected ion chromatograms generated during liquid chromatography-electrospray ionization mass spectrometry analyses of the compounds are provided in the small panels on the right: 4-coumaric acid, m/z = 163.1; naringenin, m/z = 271.1. b In vivo assay of STS. Resveratrol (3) was used as the standard compound. Mn4CL: E. coli-expressed Mn4CL. Mn4CL + MnSTS7: E. coli-coexpressed Mn4CL and MnSTS7. Mn4CL + MnSTS8: E. coli-coexpressed Mn4CL and MnSTS8. Selected ion chromatograms of the compound are provided in the small panels on the right: resveratrol, m/z = 227.1.

Fig. 6 Heat maps of hierarchical clustering of mulberry type III polyketide synthase superfamily genes. Data were adjusted by log transformation and the mean center method. Hierarchical clustering with average linkages was used to calculate K-medians with five clusters.
Exposure to UV-C radiation resulted in a significant increase (P < 0.001) in the leaf flavonoid content 12 h after treatment, followed by a decrease in abundance (Fig. 7). The highest total flavonoid content of UV-C treated leaves was 1.6-fold higher than that of control leaves. Additionally, exposure to dark conditions resulted in a decrease in flavonoid abundance. Regarding stilbene content following UV-C irradiation, we measured the accumulation of three representative stilbenes in mulberry: resveratrol (trans-3’,4’,5’-trihydroxystilbene), oxyresveratrol (trans-2,3’,4,5’-tetrahydroxystilbene, which can be formed by 2’-hydroxylation of resveratrol), and mulberroside A (oxyresveratrol diglycopyranoside, where 4-OH and 3’-OH are replaced by two glucose molecules). Mulberroside A was undetectable in leaves after exposure to UV-C radiation and dark conditions. In contrast, oxyresveratrol content increased 6–18 h after UV-C treatment, peaking at 18 h. Resveratrol was almost undetectable in control samples, even after a slight increase 6–18 h after UV-C irradiation.

Fig. 7 Effect of UV-C irradiation on type III PKS gene expression and the biosynthesis of compounds. a The MnSTS2 real-time PCR primers could anneal to the other three STS genes (i.e., MnSTS3, MnSTS8, and MnSTS9), although all primers were designed to match the most variable STS regions, including the 3’ untranslated regions. Relative gene expression levels were normalized against a mulberry actin gene (MnACTIN3). Data are provided as the mean ± standard deviation. b Variability among compounds following UV-C irradiation. Data are provided as the mean ± standard deviation.
Discussion
Expansion of the mulberry type III PKS gene family
Expansion of multigene families is a common occurrence during evolution, with gene duplications playing an important role [36]. The genomes of terrestrial plants have undergone several gene duplications throughout their evolutionary histories, including tandem, segmental, and whole genome duplications [37, 38]. In this study, six MnCHS, ten MnSTS, and two MnPKS genes were identified in the mulberry genome. MnPKS1 and MnPKS2 arose early as ancestors of mulberry type III PKS genes. The contention that the number of MnCHS genes increased mainly through tandem and segmental duplications is supported by the results of microsynteny analysis. Additionally, tandem duplication is a major reason for the generation of the MnSTS family. The transposable elements distributed throughout the MnSTS cluster likely had important activities during the formation of tandem arrays. These transposable elements are present in interspersed repetitive elements throughout the genome, and promote crossing over between non-homologous segments during meiosis or recombinational repair [39, 40]. Such unequal recombination may have contributed to the formation of pseudogenes, and a putative gene with undetectable expression, MnCHSL2, gained an additional DNA fragment containing a CHS/STS active site (R-[LIVMFYS]-x-[LIVM]-x-[QHG]-x-G-C-[FYNA]-[GAPV]-G-[GAC]-[STAVK]-x-[LIVMF]-[RAL]) during this process.

The MnCHS and MnSTS families are under strong purifying selection
The MnCHS2/MnCHS6 and MnSTS7/MnSTS8 genes were coexpressed with Mna4CL in E. coli, and the expected products were detected. It is unclear why the mulberry genome contains multiple MnCHS and MnSTS genes that encode enzymes with redundant catalytic activities. Analyses of evolutionary constraints revealed that no positive selection pressures were associated with MnSTS genes, which was consistent with the situation for MnCHS genes. Functional divergence within MnSTS lineages was also undetected. All studies indicated these two gene families were under strong purifying selection pressure to maintain their functions in the biosynthesis of secondary metabolites. Increasing the number of genes with redundant activities may have been necessary to ensure enough enzymes were produced to synthesize sufficient quantities of phytoalexins in response to abiotic or biotic stresses. According to the seminal theory concerning the fates of duplicate genes, the duplicated copy is shielded from the purifying selection pressure by the ancestral copy [41]. The pseudogenization of the duplicate gene is promoted by decreasing the selective constraint [40, 41]. Compared with those of the MnCHS and MnSTS families, the selective constraints of MnCHSL1 and MnCHSL3 gradually decreased, probably resulting in a lack of MnCHSL1 and MnCHSL3 expression.

Variability in MnCHS expression
Phenolic compounds can strongly absorb UV light, which protects plants from DNA damage caused by UV irradiation [13]. In a previous study of grapevine, the transcriptional responses of the VvSTS and VvCHS genes were diametrically opposed following exposure to UV-C radiation, suggesting that a flow of carbon between these two competing metabolic pathways was tightly regulated at the transcriptional level [42]. However, all MnSTS and most MnCHS genes were upregulated in mulberry leaves after UV-C treatment. The total flavonoid and stilbene contents exhibited a similar tendency. Only two MnCHS genes had distinct expression patterns. MnCHS1 was almost undetectable in leaves following UV-C treatment. It was expressed exclusively in male flowers, and likely participated in the fruit ripening process [22]. MnCHS3, which had the highest expression levels in leaves among MnCHS genes, was the only gene whose expression was strongly suppressed after UV-C treatment. The fact that the MnCHS3 expression levels coincided with total flavonoid contents suggests this gene may be important for flavonoid biosynthesis in mulberry leaves. Additionally, the MnCHS5 genes produced four distinct expression patterns in five tissues, which was in contrast to the MnSTS family members, with just two genes exhibiting distinct expression profiles. The evolution of these gene families may allow a fine spatial and temporal regulation of relevant metabolites biosynthesis under both normal and stress conditions, and a greater variability in the expression of MnCHS genes may favor plant growth and adaptation to environmental changes.

Substrate channeling in the stilbene biosynthetic pathway
According to their substrate preferences, STS enzymes are often classified as resveratrol synthase (catalyzes reactions involving 4-coumaroyl-CoA) and pinoresinol synthase (catalyzes reactions involving cinnamoyl-CoA) [14]. In mulberry seedlings of the 7th day to 20th day, the content of resveratrol is higher than that of oxyresveratrol in both leaves and roots, suggesting oxyresveratrol is probably transformed from resveratrol through oxidation [43]. MnSTS genes encode enzymes with resveratrol synthase activities, which are involved in producing specific metabolites. Further modifications (e.g., oxidation and O-methylation) of the aromatic ring derived from the phenylpropanoid-CoA substrates occur later [44]. Oxyresveratrol, and not resveratrol, can be detected in fruits, fruit marcs, and leaves of several mulberry species [45]. Additionally, we determined that 2′-hydroxylated resveratrol rapidly accumulated in mulberry leaves following UV-C treatment. An effective transformation
pathway converting resveratrol to oxyresveratrol may exist in mulberry plants (Additional file 1: Figure S1). The MnSTS enzymes contain six residues that experienced positive selection and type II functional divergence. All of these sites are highly conserved on the surface, suggesting they may be involved in interacting with other proteins to contribute to channeling resveratrol metabolites in mulberry plants.

Conclusion
Our results provide new insights into the evolutionary history and the role of mulberry type III PKS genes. MnCHS genes were expanded mainly through tandem and segmental duplications, while tandem duplications were primarily responsible for the generation of MnSTS family. Both MnCHS and MnSTS genes were under strong purifying selection to maintain their functions during such long-term evolution, suggesting genes with redundant activities may help to increase the abundance of relevant phytoalexins to keep the adaptability of plants in response to abiotic or biotic stresses. In addition, an effective transformation pathway converting resveratrol to oxyresveratrol may exist in mulberry plants. All these results providing the basis for further studies on the biosynthesis of relevant secondary metabolites in mulberry plants and may be useful for improving genomic tools and techniques for genetic manipulation.

Methods
Plant materials
The mulberry species, Morus notabilis C. K. Schneid., used to clone type III PKS genes was grown in Sichuan province, China. For UV-C treatments, mulberry plantlets were grown in a chamber at 25 °C with a 12-h photoperiod until the aerial parts grew to about 20 cm. Seedlings were exposed to UV-C irradiation for 30 min and then maintained in the dark. Control seedlings were kept in the dark without a prior UV-C treatment. Leaves were harvested after 0, 6, 12, 18, and 24 h.

Total RNAs from six tissues (roots, bark, shoots, leaves, male flowers, and winter buds) were extracted using RNAiso Plus (Takara, Otsu, Japan) according to the manufacturer’s instructions. cDNA was synthesized from 1 μg total RNA in a 25 μl reaction using the PrimeScript RT reagent kit (Takara).

Identification and cloning of mulberry type III PKS genes
Previously characterized type III polyketide synthase amino acid sequences were downloaded from UniProt (http://www.uniprot.org/) [46], and the consensus pattern of CHS/STS active site [PS00441] were obtained from PROSITE (http://prosite.expasy.org/) [47]. All sequences were then used as queries in a BLASTP search against the mulberry hypothetical protein database (http://morus.swu.edu.cn/morusdb/) [21]. Mulberry type III PKS genes were also identified by TBlastN analysis against the mulberry genome with an e-value cut-off of 1e-5. All hits were analyzed using the FgenesH++ program (http://www.softberry.com) [21, 48]. The predicted genes were manually corrected through comparisons with other type III PKS genes, and were further examined with the online domain analysis program InterProScan (http://www.ebi.ac.uk/interpro/scan.html) [49]. The predicted mulberry type III PKS genes were named based on scaffold locations.

Additional file 2: Table S5 lists the primers used in gene cloning. The purified PCR products were cloned into the pMD19-T simple vector (Takara), and the insertions were confirmed by sequencing.

Phylogenetic analysis of mulberry type III polyketide synthases
The sequences of the enzymes encoded by the type III PKS genes were aligned using MUSCLE v. 3.8.31 [50]. The alignment was manually corrected using GeneDoc [51]. Phylogenetic trees were constructed using the neighbor-joining algorithm [52] with the Poisson model and pairwise deletion in MEGA5 [53]. Tree topology was assessed by bootstrap analysis with 1,000 resampling replicates.

Expansion of the mulberry type III PKS superfamily and the timing of duplication events
All scaffolds that contained putative mulberry type III PKS genes were analyzed. The microsyntenic regions surrounding genes were detected using MicroSyn software [54]. The mean Ks value was calculated for all pairs of protein-coding genes among all genomic fragments using the method described by Nei and Gojobori [55]. The timing of gene duplication events was estimated using an established procedure, assuming that the average synonymous substitution rate per site per year for plant nuclear genes was about 5 × 10^{-9} [29], i.e., the Ks/(5 × 10^{-9} × 2). Any Ks values greater than 2 were discarded because of the risk of saturation [56]. The MntEdb was used to search for transposable elements [57].

Phylogenetic d_δ/d_δ ratio and functional divergence analysis
All full mulberry type III PKS coding sequences were aligned using PAL2NAL [58] based on the protein sequence alignment produced by MUSCLE v3.8.31. The phylogenetic trees was generated by Bayesian inference using MrBayes v3.2 [59], the best-fit models for mulberry gene sequence set and multispecies gene sequence set were GTR + G model and SYM + I + G model, respectively. Selection analyses of the type III PKS genes were completed using three models from the Codeml program of the PAML v4.3 package [60]. For protein-
coding genes, robust evidence for positive selection pressure was provided by a high nonsynonymous substitution rate $d_N$ relative to the synonymous rate $d_S$ ($\omega > 1$). A $\omega$ value of 0 or <1 indicated neutral evolution or purifying selection, respectively [61]. Likelihood ratio tests were used to identify the best model for $\omega$ changes between two hypotheses. For all LRTs, the null model was a simplified version of the selection model with fewer parameters, while the other model, as the alternative hypothesis, contained a priori foreground lineages [62, 63]. The functional divergence of type I and type II genes among the clusters of mulberry type III PKS genes was examined using DIVERGE v2.0 [33]. Type-I functional divergence referred to the evolutionary rate shift of site-specific amino acid after gene duplication, whereas Type-II functional divergence referred to the site-specific amino acid physiochemical property shift in the late phase when evolutionary rates were consistent. The MnCHS-group A subfamily and MnPKS family were excluded because groups with fewer than four sequences could not be analyzed using this method. A $\theta$ value >0 indicated altered selective constraints of amino acid sites after gene duplication.

In vivo characterization of a selection of mulberry type III polyketide synthases by coexpression with 4-coumaroyl-CoA ligase in Escherichia coli

Six putative type III PKS genes, namely MnPKS1, MnPKS2, MnCHS2, MnCHS6, MnSTS7, and MnSTS8, were cloned into a pET28a (+) vector (Novagen) to create pET28a-PKS expression constructs. A pCold-4CL construct was prepared using the pCold vector (Takara) and Mn4CL, which encodes an M. notabilis 4CL. The primers used to clone Mn4CL are listed in Additional file 2: Table S5. Aliquots (5 μl) of overnight cultures of E. coli containing pCold-4CL or pCold-4CL and pET28a-PKS were used to inoculate 50 ml Terrific Broth (TB) medium supplemented with ampicillin or kanamycin and ampicillin (100 mg ml$^{-1}$). Cells were grown at 37 °C until the optical density (600 nm) of the cultures reached 0.3. Samples were then cooled on an ice bath for 30 min, and cells were induced with 0.55 mM IPTG at 16 °C. The target recombinant proteins were detected on 12 % SDS–PAGE before processing for further experiments. The cultures were then supplemented with 1 mM 4-coumaric acid prepared in dimethyl sulfoxide. The culture medium was harvested at 24 and 48 h after the addition of the phenolic acid and centrifuged at 13,000 $\times$ g for 10 min. The supernatants were divided into two samples, with one (2 ml) acidified with 0.1 N HCl and treated with two volumes of ethyl acetate [35]. The pH of the other sample was adjusted to approximately 9.0 to enable the spontaneous conversion of naringenin chalcone to naringenin in vitro, and then treated with an equal volume of ethyl acetate [34, 64]. After evaporating the samples, two extracts were dissolved in 100 μl acetonitrile for analysis by LC-ESI-MS.

Subcellular localization

A linker (5’-TGAATTCTCTCTCCCTGATCCTCCTCCTCCT CCTGATCCTCCTCCTCC-3’) was individually introduced into six type III PKS genes (MnPKS1, MnPKS2, MnCHS2, MnCHS6, MnSTS7, and MnSTS8). The resultant sequences were then inserted into KpnI and BamHI sites of pLGNL-EGFP plasmid to create EGFP-target fusion plasmids. The EGFP-target fusion plasmids and the pLGNL-EGFP control plasmid were introduced into tobacco epidermal cells by Agrobacterium-mediated transformation according to a reported method [65]. The signals were detected and photographed under a fluorescent inverted microscope (Olympus IX73, Japan).

Expression of mulberry type III PKS genes

The RNA-Seq data (i.e., SRX504963, SRX504944, SRX504924, SRX504906, and SRX504893) of five mulberry tissues were downloaded from NCBI. The expression levels of all predicted mulberry type III PKS genes were determined according to the fragments per kilo base of transcript per million fragments mapped (FPKM) method [66]. The data were adjusted by log transformation and the mean center method. Hierarchical clustering with average linkages was used to calculate K-medians with five clusters. All data were analyzed and expressed graphically using Heatmap3 [67]. For quantitative reverse transcription PCR (qRT-PCR), the second leaves of mulberry saplings were used. Total RNA extraction and cDNA synthesis were completed as described earlier. Each qRT-PCR was completed using SYBR Premix EX Taq II (Takara) and the StepOnePlus Real-Time PCR System (Applied Biosystems, Foster City, CA, USA) according to the manufacturers’ instructions. Diluted cDNA (2 μl) was used as the template. The mulberry MnAC-TIN3 gene served as a control to normalize the target gene expression data. Relative expression was defined as $2^{[Ct (target gene) - Ct (control gene)]}$ [68]. Additional file 2: Table S6 lists the gene-specific primers used for qRT-PCR.

Extraction, identification, and quantification of compounds

Compounds were extracted from the third, fourth, and fifth leaves. Fresh leaf tissue (1 g) in 25 ml 60 % methanol was used in ultrasonic extractions for 40 min. The resulting supernatant was filtered through a membrane with 0.45-μm pores. All samples were prepared in triplicate. To measure the total flavonoid content, 2.5 ml sample was mixed with 1 ml 10 % aluminum chloride and 5 % sodium nitrite. A flavonoid-aluminum complex formed after the addition of 10 ml 1 M sodium hydrate
solution. The absorbance at 510 nm was determined after mixing the solution for 5 min [69]. Methanol instead of sample was used as a blank. The total flavonoid content was calculated using a standard curve of rutin [70].

An Acquity UPLC system (Waters, Milford, MA, USA) was used to analyze stilbene content. Separations were performed on an Acquity UPLC BEH C18 column (1.7 µm, 1.0 x 100 mm) at 40 °C. Three standards (i.e., mulberoside A, oxyresveratrol, and resveratrol) were accurately weighed and dissolved in 60 % methanol. Acetonitrile and 0.5 % (v/v) formic acid were used as mobile phases A and B, respectively, with the following elution profile: 0–5 min, 5–30 % A; 5–6 min, 100 % A; 6–7.5 min, 5 % A. Fractions were monitored at 320 nm. Components were identified by comparing the retention times of the eluting peaks with those of commercial standards under the same conditions. Dose-dependent calibration curves of the standards were used to determine the component concentrations.

The products of mulberry type III polyketide synthases were separated using acetonitrile and 0.5 % (v/v) acetic acid as mobile phases A and B, respectively. For the alkaline sample, the gradient elution conditions were as follows: 0–1 min, 30 % A; 1–4 min, 30–80 % A; 4–5 min, 80 % A; 5–7 min, 80–100 % A; 7–8 min, 100–30 % A; 8 min, 30 % A. For the acidic sample, the gradient elution conditions were as follows: 0–1 min, 30 % A; 1–4 min, 30–70 % A; 4 min, 100 % A; 5–6 min, 30 % A. The electrospray ionization mass spectrometer (Acquity QDa Detector) was operated in the negative mode, scanning the mass-to-charge ratio (m/z) between 100 and 300. Negative ion data for standard compounds were as follows: 4-coumaric acid, m/z = 163.1; naringenin, m/z = 271.1; resveratrol, m/z = 227.1.

Homology modeling of MnSTS and calculating the evolutionary conservation of amino acid positions
A dimer of the mulberry STS2 enzyme was modeled using MODELER [71]. Structural alignments were prepared using the following known structures: alfalfa (Medicago sativa) CHS enzymes (PDB 1CGZ) [72], peanut STS bound to resveratrol (PDB 1Z1F) [73], and Freesia hybrida CHS1 (4WUM) [74]. The generated structure was evaluated using PROCHECK from the Structure Analysis and Verification Server v.4 (http://services.mbi.ucla.edu/SAVES/) [75]. ConSurf (http://consurf.tau.ac.il/) was used to identify the probable evolutionarily conserved residues in the STS enzymes [31]. The analysis was conducted using the Bayesian algorithm with the JTT model.

Additional files

**Additional file 1: Figure S1.** General flavonoid and stilbene biosynthetic pathways in mulberry. The enzymes shown in these pathways are as follows: PAL, phenylalanine ammonia-lyase; C4H, cinnamate-4-hydroxylase; 4CL, 4-coumaroyl-CoA ligase; CHS, chalcone synthase; STS, stilbene synthase. (PDF 783 kb)

**Additional file 2: Table S1.** Type III polyketide synthase genes in the M. notabilis genome. Genes have been named based on scaffold locations. ORF length: open reading frame length confirmed by clone analysis. *: genes without expression data. Table S2: Estimation of the mean Ks values for the genes flanking mulberry type III PKS genes. The mean Ks value was calculated for each homologous pair of protein-coding genes between genomic fragments containing type III PKS genes if microsynteny was detected. Table S3: Parameter estimates of codon-substitution evolutionary models for three MnCHS-like genes with undetectable expression. The likelihood ratio tests are as follows: Branch model, one ratio model 0 vs two ratio model 2; Clade model C, M2a_rel vs CmC; Branch site model, model A null (ω is fixed as 1) vs model A, ω = dS/dN = background; f = foreground, o=other; p0, p1, and p2 = proportion of sites with dS/dN ratios = ω0, ω1, and ω2, respectively. In = in of the likelihood; P-value = P-values of the likelihood ratio tests. Table S4: Type I and type II functional divergence between MnSTS and MnCHS families as well as in MnSTS subfamilies. Table S5: Primers used for cloning mulberry type III PKS and 4CL genes. Four genes labelled with asterisk, including MnSTS2, MnSTS3, MnSTS8 and MnSTS9 shared the forward (from initiation codon ATG) and reverse (end with stop codon TAG) primers, while other primers matching to variable regions have been used to distinguish them. Sequences of primers used for quantitative reverse transcription PCR. (DOXX 36 kb)

**Additional file 3: Figure S2.** Comparison of mulberry type III polyketide synthase amino acid sequences. Conserved sites are shaded. The asterisk indicates the site contains a conserved Thr197 in CHS, but a Gly in PKS1 and PKS2. Purple arrow: positive selection and type II divergence sites; red arrow: positive selection sites; green arrow: type I divergence sites; yellow arrow: type II divergence sites. Red box: catalytic triad; blue box: co-enzyme A binding sites; yellow box: important residues for functional diversity. (PDF 2015 kb)

**Additional file 4: Figure S3.** The phylogenetic tree used to calculate phylogenetic dN/dS ratio for multispecies set. (PDF 1102 kb)

**Additional file 5: Figure S4.** Coexpression of M. notabilis 4CL and type III PKS in E. coli. The results were detected in a SDS-PAGE gel (12%). (A): soluble protein fraction; (B): insoluble protein fraction; Lane 1: pCold-4CL; Lane 2: pCold-4CL+pET28a-MnPKS1; Lane 3: pCold-4CL+pET28a-MnPKS2; Lane 4: pCold-4CL+pET28a-MnCHS2; Lane 5: pCold-4CL+pET28a-MnCHS6; Lane 6: pCold-4CL+pET28a-MnSTS2; Lane 7: pCold-4CL+pET28a-MnSTS8; Lane 8: pCold-4CL control (no IPTG); Lane 9: pCold-4CL+pET28a-MnPKS control (no IPTG); Lane 10: pCold+pET28a empty plasmids control. Marker: molecular weight standard. (PDF 1187 kb)

**Additional file 6: Figure S5.** In vivo characterization of MnPKS1 and MnPKS2 by coexpression with 4-coumaroyl-CoA ligase in Exchellia collaria. Mn4CL: E. coli-expressed Mn4CL, Mn4CL+MnPKS1: E. coli-coexpressed Mn4CL and MnPKS1, Mn4CL+MnPKS2: E. coli-coexpressed Mn4CL and MnPKS2, (A) 4-Coumaric acid (1) and naringenin (2) were used as standard compounds. Selected ion chromatograms generated during liquid chromatography-electrospray ionization mass spectrometry analyses of the compounds are provided in the small panel on the right side. 4-Coumaric acid, m/z = 163.1; naringenin, m/z = 271.1. (B) Resveratrol (3) was used as the standard compound. A selected ion chromatogram of the compound is provided in the small panel on the right side. resveratrol, m/z = 227.1. (PDF 941 kb)

**Additional file 7: Figure S6.** The subcellular localization of mulberry type III polyketide synthases. (A) The control vector pLGNL-EGFP, (B) pLGNL-MnPKS1-EGFP, (C) pLGNL-MnPKS2-EGFP, (D) pLGNL-MnCHS2-EGFP, (E) pLGNL-MnCHS6-EGFP, (G) pLGNL-MnSTS3-EGFP, and (F) pLGNL-MnSTS8-EGFP were transiently expressed in tobacco epidermal cells. Column (a), bright field; column (b), EGFP fluorescence; and column (c), merged fluorescence. (PDF 3312 kb)

**Abbreviations**
4CL, 4-coumaroyl-CoA ligase; CHS, chalcone synthase; IPTG, isopropyl β-D-1-thiogalactopyranoside; KS, synonymous silent substitutions per site; LC-ESI-MS, liquid chromatography-electrospray ionization mass spectrometry; LC-MS, liquid chromatography-mass spectrometry; LRT, likelihood ratio test; MSS, mulberry sequence set; MFA, million years ago; OS, other species sequence
set; PKS, polyketide synthase; STS, stilbene synthase; TB medium, Terrific Broth medium; UPLC, Ultra Performance Liquid Chromatography; UV-C, ultraviolet-C radiation; ω, the non synonymous/synonymous rate ratio.

Acknowledgments
Not applicable.

Funding
This project was funded by the research grants from the National Hi-Tech Research and Development Program of China (No. 2013AA100625-3), Natural Science Foundation of China (No. 31573233), China Postdoctoral Science Foundation funded projects (No. 2013 M540694 and No. 2014 T70845), and the “111” Project (B12006).

Availability of data and material
The phylogenetic tree (Fig. 1, Fig. 2 and Additional file 4: Figure S3) and corresponding alignment matrix data can be found on Treebase at [http://purl.org/phylo/treebase/phylows/study/TB2:S19415].

Authors’ contributions
HL and NH conceived and designed research. HL conducted experiments. GD provided new reagents or analytical tools. HL, BM and NH analyzed data. HL wrote the paper and NH revised the manuscript. All authors read and approved the manuscript.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Ethics approval and consent to participate
Not applicable.

Received: 26 January 2016 Accepted: 16 June 2016
Published online: 04 August 2016

References
1. Singáñ AB, El-Beshbishi HA, Yonekawa M, Nomura T, Fukai T. Hypoglycemic effect of Egyptian Morus alba root bark extract: effect on diabetes and lipid peroxidation of streptozotocin-induced diabetic rats. J Ethnopharmacol. 2005;103(3):333–8.
2. Nomura T, Fukai T, Hanyo Y. Constituents of the Chinese crude drug Sang-Bai-Pi (Morus Root Bark) [1] structure of a new flavanol derivative, Sanggenon A[2]. Planta Med. 1983;47(1):33–4.
3. Noel JP, Austin MB, Bomati EK. Structure-function relationship in plant phenylpropanoid biosynthesis. Curr Opin Plant Biol. 2005;8:249–53.
4. Schnée S, Vret O, Gindro K. Role of stilbenes in the resistance of grapevine to powdery mildew. Physiol Mol Plant Pathol. 2008;72:128–33.
5. Ververidis F, Trantas E, Douglas C, Vollmer G, Kretzschmar G, Panopoulos N. Biotechnology of flavonoids and other phenylpropanoid-derived natural products. Part I: Chemical diversity, impacts on plant biology and human health. Biotechnol J. 2007;10(2):124–34.
6. de Pascau-Teresa S, Sanchez-Ballesta MT. Anthocyanins: from plant to health. Phytochem Rev. 2008;7(2):281–99.
7. Kim YM, Yun J, Lee CK, Lee H, Min KR, Kim Y. Oxyresveratrol and hydroxystilbene compounds. Inhibitory effect on tyrosinase and mechanism of action. J Biol Chem. 2002;277(18):16340–4.
8. Abe J, Morita H. Structure and function of the chalcone synthase superfamily of plant type III polyketide synthases. Nat Prod Rep. 2010;27(7):809–38.
9. Ferrer JL, Austin MB, Stewart Jr C, Noel JP. Structure and function of enzymes involved in the biosynthesis of phenylpropanoids. Plant Physiol Biochem. 2008;46(3):356–70.
10. Austin MB, Bowman ME, Ferrer JL, Schröder J, Noel JP. An aldol switch discovered in stilbene synthases mediates cyclization specificity of type III polyketide synthases. Chem J. 2004;29(11):79–94.
11. Austin MB, Noel JP. The chalcone synthase superfamily of type III polyketide synthases. Nat Prod Rep. 2003;20(1):79–110.
12. Tropf S, Lanz T, Rensing SA, Schröder J, Schröder G. Evidence that stilbene synthases have developed from chalcone synthases several times in the course of evolution. J Mol Evol. 1994;38(6):610–8.
13. Dao TT, Linthurst HJ, Verpoorte R. Chalcone synthase and its functions in plant resistance. Phytochem Rev. 2011;10(3):397–412.
14. Monta H, Noguchi H, Schröder J, Abe I. Novel polyketides synthesized with a higher plant stilbene synthase. Eur J Biochem. 2001;268(13):3759–66.
15. Eckermann S, Schröder G, Schmidt I, et al. New pathway to polyketides in plants. Nature. 1998;396:387–90.
16. Koskela S, Söderholm PP, Ainaosa M, Wennberg T, Kikka KD, Ovharekko W, Kylanlaiti I, Auerma T, Yli-Kauhaluoma J, Pihlaja K, Vuorela PM, Teeri TTh. Polyketide derivatives active against Botrytis cinerea in Gerbera hybrida. Planta. 2011;233(1):37–48.
17. Baumert A, Maier W, Gröger D, Deutzmann R. Purification and properties of aciconine synthase from cell suspension cultures of Ruta graveolens L. Naturforh. C. 1994;49(1–2):26–32.
18. Flagel LE, Wendel JF. Gene duplication and evolutionary novelty in plants. New Phytol. 2009;183(3):557–64.
19. Holding DR, Meeley RB, Habezook J, Seidler D, Grues F, Jung R, Larks BA. Identification and characterization of the maize arogenate dehydrogenase gene family. J Exp Bot. 2010;61(13):3663–73.
20. Espinosa-Cantú A, Ascencio D, Barona-Gómez F, DeLuna A. Gene duplication and the evolution of moonlighting proteins. Front Genet. 2015;6:227.
21. He N, Zhang C, Qi X, Zhao S, Yao Y, Yang G, Lee TH, Wang X, Cai Q, Li D, et al. Draft genome sequence of the mulberry tree Morus notabilis. Nat Commun. 2013;4:2445.
22. Qi X, Shuai Q, Chen H, Fan L, Zeng Q, He N. Cloning and expression analyses of the anthocyanin biosynthetic genes in mulberry plants. Mol Genet Genomics. 2014;289(5):1–11.
23. Kim SS, Grienenerberg E, Lallambad B, Colpitts CC, Kim SY, Souza Cde A, Geofroy P, Heintz D, Heitz T, Suh DY, Legrand M, Douglas CJ. LARP5/POLYKETIDE SYNTHASE A and LAPS/POLYKETIDE SYNTHASE B encode hydroxylalkyl pyrone synthases required for pollen development and sporopollenin biosynthesis in Arabidopsis thaliana. Plant Cell. 2010;22(12):4045–56.
24. Kim SY, Colpitts CC, Wiedemann G, Jepson C, Rahimi M, Rothwell JR, McInnes AD, Hasebe M, Reski R, Sterenberg BT, Suh DY. Physcomitrella PpPORS, basal to plant type III polyketide synthases in phylogenetic trees, is a very long chain 2′-oxalylferesorico acid synthase. J Biol Chem. 2013;288(4):2767–77.
25. Mizuuchi Y, Shimokawa Y, Wabiuchi K, Noguchi H, Abe I. Structure function analysis of novel type III polyketide synthases from Arabidopsis thaliana. Biol Pharm Bull. 2009;32(11):2050–1.
26. Jez JM, Austin MB, Ferrer J, Bowman ME, Schröder J, Noel JP. Structural control of polyketide formation in plant-specific polyketide synthases. Chem Biol. 2000;7(12):919–30.
27. Monta H, Kondo S, Oguro S, Noguchi H, Sugio S, Abe I, Kohno T. Structural insight into chain-length control and product specificity of pentaketide choromone synthase from Aloe arborescens. Chem Biol. 2007;14(4):359–69.
28. Li WH. Molecular Evolution. Sunderland: Sinauer Association; 1999.
29. Wolfe KH, Gouy M, Yang YW, Sharp PM, Li WH. Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proc Natl Acad Sci U S A. 1989;86(16):6021–5.
30. Weadick CJ, Chang BS. Complex patterns of divergence among green-sensitive (Rh2a) African cichlid opsins revealed by Clade model analyses. BMC Evol Biol. 2012;12:206.
31. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N, Consurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res. 2010;38(13):W529–33.
32. Gu X. Maximum-likelihood approach for gene family evolution under functional divergence. Mol Biol Evol. 2001;18(4):453–64.
33. Gu X, Velden K. DIVERGE: Phylogeny-based analyses or Functional-structural divergence of a protein family. Bioinformatics. 2002;18:500–1.
34. Watts KT, Lee PC, Schmidt-Dannert C. Exploring recombinant flavonoid biosynthesis in metabolically engineered Escherichio coli. ChemBiochem. 2004; 5(4):500–7.
35. Watts KT, Lee PC, Schmidt-Dannert C. Biosynthesis of plant-specific stilbene polyketides in metabolically engineered Escherichia coli. BMC Biotechnol. 2006;6(1):1–12.
36. Xu GX, Guo CC, Shan HY, Kong HZ. Divergence of duplicate genes in exon-intron structure. Proc Natl Acad Sci U S A. 2012;109(4):1187–92.
et al. BMC Genomics (2016) 17:540

37. Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Solits DE, Solits PS, Wendel JF. Evolutionary genetics of genome merger and doubling in plants. Annu Rev Genet. 2008;42:443–61.
38. Zhu Y, Wu N, Song W, Yin G, Qin Y, Yan Y, Hu Y. Soybean (Glycine max) expansin gene superfamily origins: segmental and tandem duplication events followed by divergent selection among subfamilies. BMC Plant Biol. 2014;14.
39. Fiston-Lavie AS, Anxolabheire D, Quesneville H. A model of segmental duplication formation in drosophila melanogaster. Genome Res. 2007;17(10):1458–70.
40. Rodgers-Melnych E, Mane SP, Dhamawanadhana P, Slavov GT, Costa OR, Strauss SH, Brunner AM, Difazio SP. Contrasting patterns of evolution following whole genome versus tandem duplication events in populus. Genome Res. 2012;22(1):95–105.
41. Ohno S. Evolution by a gene duplication. New York: Springer; 1970.
42. Vannozzi A, Dry IB, Fasoli M, Zenoni S, Lucchin M. Genome-wide analysis of the grapevine stilbene synthase multigenic family: genomic organization and expression profiles upon biotic and abiotic stresses. BMC Plant Biol. 2012;12(4):1–22.
43. Zhou J, Li SX, Wang W, Guo XY, Lu XY, Huang D, Wei BY, Cao L. Variations in the levels of mulberroside A, oxyresveratrol, and resveratrol in mulberries in different seasons and during growth. Scientific World J. 2013;19:380692.
44. Hammerbacher A, Ralph SG, Bohlmann J, Fenning TM, Gershenzon J, Schmidt A. Biosynthesis of the major tetrahydroxystilbene in spruce, arstingin and isohoratin, proceeds via resveratrol and is enhanced by fungal infection. Plant Physiol. 2011;157(2):876–90.
45. Song W, Wang BJ, Buchell P, Zhang PF, Wei DZ, Lu YH. Phytochemical profiles of different mulberry (Morus sp.) species from China. J Agric Food Chem. 2009;57(19):9133–40.
46. The UniProt Consortium. Activities at the Universal Protein Resource (UniProt). Nucleic Acids Res. 2014;42(Database issue):D191–8.
47. Sigrist CJ, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L, Xenarios I. New and continuing developments at PROSITE. Nucleic Acids Res. 2013;41(Database issue):D344–7.
48. Solovyev V, Kosarev P, Seledsov I, Vorobyev D. Automatic annotation of eukaryotic genes, pseudogenes and promoters. Genome Biol. 2006; Suppl 1:S10.1–12.
49. Zdobnov EM, Apweiler R. InterProScan—an integration platform for the signature-recognition methods in InterPro. Bioinformatics. 2001;17(9):847–8.
50. Edgar RC. Muscle: multiple sequence alignment with high accuracy and speed. Genome Biol. 2004;5(2):R95.
51. Nicholas KB, Nicholas H, Deerfield D. Genedoc: Analysis and visualization of genetic variation. Embnew News. 1997;4:14.
52. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987;4(4):406–25.
53. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGAS: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011;28(10):2731–9.
54. Bai B, Yang X, Tuskan GA, Cheng ZM. Microsyn: a user friendly tool for detection of microsynteny in a gene family. BMC Bioinformatics. 2011;12:79.
55. Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol. 1986;3(5):418–26.
56. Guo X, Su B, Zhou Z, Sha J. Rapid evolution of mammalian X-linked tests microRNAs. BMC Genomics. 2009;10:97.
57. Ma B, Li T, Xiang Z, He N. MiTEDb, a collective resource for mulberry transposable elements. Database (Oxford). 2015;2015.
58. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 2006;34(Web Server issue):W690–12.
59. Ronquist F, Helsenbeek JP, Mirkayev J. Bayesian phylogenetic inference under mixed models. Bioinformatics. 2003;19(12):1572–4.
60. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24(8):1586–91.
61. Yang Z, Wong WSW, Nielsen R. Bayes empirical Bayes inference of amino acid site under positive selection. Mol Biol Evol. 2005;22:1107–18.
62. Yang Z, Nielsen R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol. 2002;19(5):908–17.