TrichOME: A Comparative Omics Database for Plant Trichomes

Xinbin Dai, Guodong Wang, Dong Sik Yang, Yuhong Tang, Pierre Broun, M. David Marks, Lloyd W. Sumner, Richard A. Dixon, and Patrick Xuechun Zhao*

Plant Biology Division, The Samuel Roberts Noble Foundation, Ardmore, Oklahoma 73401 (X.D., D.S.Y., Y.T., L.W.S., R.A.D., P.X.Z.); Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China (G.W.); Nestlé R&D Center Tours, Plant Science and Technology, 37390 Notre-Dame D’Oé, France (P.B.); and Department of Plant Biology, University of Minnesota, St. Paul, Minnesota 55108 (M.D.M.)

Plant secretory trichomes have a unique capacity for chemical synthesis and secretion and have been described as biofactories for the production of natural products. However, until recently, most trichome-specific metabolic pathways and genes involved in various trichome developmental stages have remained unknown. Furthermore, only a very limited amount of plant trichome genomics information is available in scattered databases. We present an integrated “omics” database, TrichOME, to facilitate the study of plant trichomes. The database hosts a large volume of functional omics data, including expressed sequence tag/unigene sequences, microarray hybridizations from both trichome and control tissues, mass spectrometry-based trichome metabolite profiles, and trichome-related genes curated from published literature. The expressed sequence tag/unigene sequences have been annotated based upon sequence similarity with popular databases (e.g. Gene Ontology, Kyoto Encyclopedia of Genes and Genomes, and Transporter Classification Database). The unigenes, metabolites, curated genes, and probe sets have been mapped against each other to enable comparative analysis. The database also integrates bioinformatics tools with a focus on the mining of trichome-specific genes in unigenes and microarray-based gene expression profiles. TrichOME is a valuable and unique resource for plant trichome research, since the genes and metabolites expressed in trichomes are often underrepresented in regular non-tissue-targeted cDNA libraries. TrichOME is freely available at http://www.planttrichome.org/.

Plant trichomes are epidermal tissues located on the surfaces of leaves, petals, stems, petioles, peduncles, and seed coats depending on species. By virtue of their physical properties (size, density), trichome hairs can directly serve to protect buds of plants from insect damage, reduce leaf temperature, increase light reflectance, prevent loss of water, and reduce leaf abrasion (Wagner, 1991; Wagner et al., 2004). Although the morphology of trichomes varies greatly, they can be generally classified into two types: simple trichomes (STs) and glandular secreting trichomes (GSTs; Wagner et al., 2004). STs of Arabidopsis (Arabidopsis thaliana) have been chosen as models for studying cell fate and differentiation (Wagner, 1991; Breuer et al., 2009; Marks et al., 2009). In Arabidopsis, STs on leaves consist of a unicellular structure with a stalk and three to four branches (Fig. 1B). Although the STs are referred to as “nonglandular” (presumably nonsecreting), expression of genes involved in anthocyanin, flavonoid, and glucosinolate pathways can nevertheless be detected in STs, indicating the roles of STs in the biosynthesis of secondary compounds and defense (Wang et al., 2002; Jakoby et al., 2008). GSTs are found on about one-third of vascular plants. GSTs have a multicellular structure with a stalk terminating in a glandular head (Fig. 1, A and C–G). GSTs are initiated from a single protodermal cell that undergoes vertical enlargement and multiple divisions to give rise to fully developed trichomes. GSTs often produce and accumulate terpenoid and phenylpropanoid oils (Wagner et al., 2004). However, alkaloids, the third major class of plant secondary compounds, are not common in GST exudates (Laue et al., 2000). The amount of exudates produced by GSTs may reach 30% of mature leaf dry weight, as found in certain Australian desert plants (Dell and McComb, 1978). Plant GSTs can impact pathogen defense, pest resistance, pollinator attraction, and water retention based on the phytochemicals they secrete.

GSTs on the aerial organ surfaces have a unique capacity for synthesis and secretion of chemicals (largely plant secondary metabolites), and they have been described as “chemical factories” for the production of high-value natural products (Mahmoud and Croteau, 2002; Wagner et al., 2004; Schilmiller et al., 2008). Secondary metabolites play important roles in...
protecting the plant against insect predation and other biotic challenges (Peter and Shanower, 1998), and they are potential sources for pharmaceutical and nutraceutical product development. For example, the trichome-borne artemisinin from *Artemisia annua* is still the most effective drug against malaria, and the early steps of its biosynthetic pathway have been extensively studied (Duke et al., 1994; Arsenault et al., 2008). Recently, the mechanisms by which plant glandular trichomes make, transport, store, and secrete a great variety of unique compounds, especially terpenoids and flavonoids, have received extended research interest because of the potential use of these compounds in pharmaceutical and nutraceutical applications. Seminal studies have reported the assignment of gene functions to specific metabolic pathways in glandular trichomes of several plant species, including mint (*Mentha × piperita*; Alonso et al., 1992; Rajaonarivony et al., 1992; Lange et al., 2000), basil (*Ocimum basilicum*; Gang et al., 2002; Iijima et al., 2004; Xie et al., 2008), *Artemisia* (Teoh et al., 2006; Zhang et al., 2008), tomato (*Solanum lycopersicum*; Fridman et al., 2005; Besser et al., 2009; Schilmiller et al., 2009), and hop (*Humulus lupulus*; Nagel et al., 2008; Wang et al., 2008). Table 1 summarizes the trichome structure, classification, and secondary metabolites in these species.

“Omics” approaches are being extensively employed for systematically studying the molecular aspects of trichome development, metabolism, and secretion at the “systems biology” level, and data are being rapidly generated; however, currently, only limited omics data for plant trichomes are publicly available in the literature and scattered databases. For example, the National Center for Biotechnology Information (NCBI) dbEST (Boguski et al., 1993) hosts approximately 130,000 ESTs sampled from trichome cDNA libraries of 13 species (as of February 2009) or mixed tissues including trichomes. The public Minimum Information About a Microarray Experiment-compliant (Brazma et al., 2001) repository ArrayExpress (Parkinson et al., 2009) hosts 16 microarray hybridization experiments for Arabidopsis nonglandular trichomes, but there are no glandular trichome microarray data available in public databases so far. Lastly, we could not find large-scale metabolite profile data for plant trichomes from public omics databases, such as the AraCyc and MedicCyc databases (Mueller et al., 2003; Urbanczyk-Wochniak and Sumner, 2007).

The omics data in publicly available databases are usually far from comprehensive and not integrated, as most of these data are in raw format. For example, EST/unigene and microarray data sets are not interlinked, and comparative analysis between trichome and nontrichome tissues for the identification of trichome-specific genes is absent. Therefore, to mine useful information such as trichome specificity of gene expression, signal strengths need to be normalized among hybridization experiments from both trichome and nontrichome tissues; to search a specific gene and deduce its potential function(s), sequences from cDNA libraries need to be annotated based on metabolic pathways and further cross-linked to metabolite profiles by catalytic reaction. Such data collection, integration, and mining processes pose great challenges to biologists interested in trichomes.

To overcome the above limitations, we have developed TrichOME, an integrated genomic database for plant trichomes. TrichOME integrates omics data, such as EST/unigene sequences, microarray gene expression...
profiles, metabolite profiles, and literature-mining results. To assemble unigenes and pursue comparative analysis of gene expression, the database has integrated a large amount of EST sequence and gene expression profile data from nontrichome tissues as well. We processed, curated, and annotated the hosted data by

<table>
<thead>
<tr>
<th>Species</th>
<th>Trichome Structure and Classification</th>
<th>Metabolites</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Arabidopsis thaliana</em></td>
<td>Nonglandular trichome, large, single epidermal cells with a stalk and three or four branches on the surface of most shoot-derived organs</td>
<td>Artemisinin, an endoperoxide sesquiterpene lactone</td>
</tr>
<tr>
<td><em>Artemisia annua</em></td>
<td>Biseriate 10-celled glandular trichome, head including three apical cell pairs (Duke and Paul, 1993)</td>
<td>Labdane-type components, such as ent-3’-acetoxy-13-epi-manoyl oxide and ent-13-epi-manoyl oxide (Falara et al., 2008)</td>
</tr>
<tr>
<td><em>Cistus creticus</em></td>
<td>Glandular trichome composed of a long multicellular stalk of over 200 µm topped by a small glandular head cell; two types of nonglandular trichome: multicellular stellate and simple unicellular spike (Gülf et al., 1996)</td>
<td>Essential oil, including myrcene, humulene, and caryophyllene; bitter acids, including humulones and lupulones; prenylfalvonoids, including xanthohumol and desmethylxanthohumol (Wang et al., 2008)</td>
</tr>
<tr>
<td><em>Humulus lupulus</em></td>
<td>Peltate glandular trichome with a glandular head consisting of 30 to 72 cells, four stalk cells, and four basal cells; bulbous glandular trichome, consisting of four (occasionally eight) head glandular cells, two stalk cells, and two basal cells; nonglandular trichome (cystolith hair) with a hard calcium carbonate structure at base of a hair (Oliveira et al., 1988; Nagel et al., 2008)</td>
<td>Essential oils, such as p-menthanes; monoterpenes, including menthone and menthol</td>
</tr>
<tr>
<td><em>Medicago sativa</em></td>
<td>Erect glandular trichome containing multicellular stalk typically over 200 µm long topped by a glandular head composed of a few cells with a diameter of approximately 15 µm; nonglandular trichome composed of a short base cell and a unicellular elongated shaft (Ranger and Hower, 2001)</td>
<td>N-(3-Methylbutyl) amide of linoleic acid (Ranger et al., 2005)</td>
</tr>
<tr>
<td><em>Mentha × piperita</em></td>
<td>Peltate glandular trichome consisting of a basal cell, a stalk cell, and disc of eight glandular cells approximately 60 µm in diameter (McCaskill et al., 1992)</td>
<td>Essential oils, such as p-menthanes; monoterpenes, including menthone and menthol</td>
</tr>
<tr>
<td><em>Nicotiana tabacum</em></td>
<td>Tall glandular trichome, a multicellular stalk topped by unicellular or multicellular head; short glandular trichome, a unicellular stalk topped by multicellular head (Akers et al., 1978)</td>
<td>Labdene-diol diterpenes and amphipathic sugar esters (Lin and Wagner, 1994)</td>
</tr>
<tr>
<td><em>Ocimum basilicum</em></td>
<td>Peltate glandular trichome consisting of a base cell, stalk cell, and a four-celled head; capitate glandular trichome, consisting of a single base and stock cell and one- to two-celled head; multicellular nonglandular spiked trichome (Werker et al., 1993)</td>
<td>Phenylpropanoid eugenol, monoterpanoid linalool, and phenylpropanoid methylcinnamnate (Xie et al., 2008)</td>
</tr>
<tr>
<td><em>Salvia fruticosa</em></td>
<td>Type I consisting of one to two stalk cells and one to two enlarged, rounded to pear-shaped secretory head cells; type II consisting of one to two stalk cells and one elongated head cell as narrow as the stalk cells at its base and slightly enlarged above; type III consisting of two to five elongated stalk cells and rounded head in young leaves, which becomes cup shaped in mature leaves (Werker et al., 1985)</td>
<td>Essential oils, such as α-pinene, 1,8-cineole, camphor, and borneol (Arikat et al., 2004)</td>
</tr>
<tr>
<td><em>Solanum habrochaites</em></td>
<td>Type I glandular trichome, large in size with a multicellular base; type II intermediate in size with a single basal cell; type IV a short, multicellular stalk that secretes droplets of sticky exudate at the tip; type V, short, slender, one to four celled; type VI, short with a two- to four-celled glandular head; type VII, 0.05- to 0.1-mm smaller glandular hair with a four- to eight-celled glandular head; type VI is particularly abundant (Reeves, 1977)</td>
<td>Mainly α-santalene, α-bergamotene, and β-bergamotene; small amounts of α-humulene and β-caryophyllene (Besser et al., 2009)</td>
</tr>
<tr>
<td><em>Solanum lycopersicum</em></td>
<td>Same as above but possesses high density of type IV glandular trichome (Lenke and Mutschler, 1984)</td>
<td>Monoterpenes (Besser et al., 2009)</td>
</tr>
<tr>
<td><em>Solanum pennelli</em></td>
<td>Same as above but possesses high density of type IV glandular trichome (Lenke and Mutschler, 1984)</td>
<td>2,3,4-Triacylglycerols (Goffreda et al., 1989)</td>
</tr>
</tbody>
</table>
referring to multiple popular biological databases to ensure the highest quality. TrichOME provides an interactive Web site (http://www.planttrichome.org/) for searching trichome-related ESTs/unigenes, microarray hybridizations, metabolites, and literature information.

DATA COMPILATION, INTEGRATION, AND MINING

EST Sequences

Plant trichome-related EST sequences were downloaded from NCBI dbEST and were grouped by cDNA libraries and species using an in-house Perl script. The cDNA libraries from nontrichome tissues and full-length cDNAs were collected for the purpose of unigene assembly and in silico expression analysis. The cDNA libraries of low quality, smaller libraries (<500 EST members), and nontrichome cDNA libraries (due to some errors in the dbEST database) were manually removed.

The collected ESTs were further processed prior to assembly in order to remove the remaining cloning vectors, adaptors, and contaminating sequences (Lee et al., 2005). For each cDNA library, the cross_match (Ewing and Green, 1998) program was used to search against the corresponding vector sequence. Subsequently, the rigorous Smith-Waterman SSearch program (Pearson and Lipman, 1988) was employed to clean short adapters, restriction endonuclease sites, and PCR primers provided in the original cDNA library descriptions. After the above procedures, TrichOME currently hosts about one million cleansed EST sequences from both trichome and corresponding nontrichome control tissues of 13 species (A. annua, Cistus creticus, H. lupulus, Medicago sativa, Medicago truncatula, M. piperita, Nicotiana benthamiana, Nicotiana tabacum, O. basilicum, Salvia fruticosa, Solanum habrochaites, S. lycopersicum, and Solanum pennellii).

The cleansed EST sequences were further assembled into unigenes, which consist of singletons and tentative consensuses by species, using The Institute for Genomic Research Gene Index clustering tools (Lee et al., 2005). Some unigenes were further refined by referring to the available full-length cDNA from the same species in order to improve assembly quality.

The unigene and EST sequences were annotated using BLASTX queries against UniProtKB and Swiss-Prot databases with a cutoff E-value of less than 1e-04. The top five query results were listed as valid annotations. Using the same protocol, the unigene/EST sequences were further annotated using the Gene Ontology database (Ashburner et al., 2000), Kyoto Encyclopedia of Genes and Genomes Pathway Database (KEGG; Kanehisa et al., 2008), Transporter Classification Database (TCDB; Saier et al., 2006), and plant transcription factor database (PlantTFDB; Guo et al., 2008). The conserved domains of sequences were searched by InterProScan with its default cutoff E-value (Hunter et al., 2009).

Microarray Hybridizations

The microarray data sets in TrichOME were collected from three different sources: ArrayExpress (Parkinson et al., 2009), Arabidopsis Gene Atlas (Schmid et al., 2005), and our internal experiments. TrichOME currently hosts microarray-based expression profiles for two model plants, Arabidopsis (using the Affymetrix ATH1 GeneChip) and M. truncatula, as well as the forage crop M. sativa (alfalfa; using the Affymetrix Medicago GeneChip). Our expression analyses with three biological replicates were performed on glandular and nonglandular trichome tissues as well as nontrichome tissues (as a control), such as stem, flower, root, and nodule (Schmid et al., 2005; Benedito et al., 2008; Wang et al., 2008).

For each Affymetrix array hybridization, the resultant image .cel file was exported using the GeneChip Operating Software version 1.4 (Affymetrix) and then imported into Robust Multiarray Average software for global normalization (Irizarry et al., 2003). Presence/absence call for each probe set was analyzed using dCHIP software (Li and Wong, 2001) for its high reliability.

To annotate the unigenes and array data, the trichome-related unigenes were mapped to Affymetrix GeneChip probe sets using a probe sets remapping PERL script developed by Affymetrix. Meanwhile, the Affymetrix probe set target sequences were mapped to the Gene Ontology database (Ashburner et al., 2000), KEGG gene and metabolic pathway database (Kanehisa et al., 2008), TCDB (Saier et al., 2006), and PlantTFDB (Guo et al., 2008) by performing BLASTX searching against the reference sequences with a cutoff E-value of less than 1e-04.

Mass Spectrometry-Based Metabolite Profiles

TrichOME hosts mass spectrometry (MS)-based metabolite profiles for plant trichomes. Currently, the database hosts gas chromatography (GC)-MS data obtained for potato leafhopper-susceptible and -resistant lines of M. sativa. The data are composed of triplicate profiles obtained by multiple published methods (Broeckling et al., 2005). These methods included GC-MS of a polar extract, GC-MS of a nonpolar extract with lipid hydrolysis, and GC-MS of nonpolar extracts without lipid hydrolysis. Metabolite profiling methods have been described in the details page of each experiment. Polar GC-MS metabolite profiles include over 300 components composed of amino acids, organic acids, alcohols, nucleotides, sugars, and sugar phosphates. Nonpolar, hydrolyzed GC-MS metabolite profiles include over 300 components composed of fatty acids, long-chain alcohols, waxes, terpenoids, and sterols. In particular, fatty acid amides were reported, as these have been recently suggested to contribute to leafhopper resistance (Ranger et al., 2005). Additional experimental data, including solid-phase microextraction-GC-MS of volatiles and ultra-
performance liquid chromatography/quadrupole time-of-flight MS for profiling secondary metabolites, have been collected and will be added to the database in the near future.

Raw GC-MS metabolite profiles were deconvoluted using AMDIS software (the Automated Mass Spectral Deconvolution and Identification System from the National Institute of Standards and Technology; http://www.amdis.net/). Deconvoluted peaks were identified using AMDIS spectral matching with a custom spectral database generated from authentic standards (Broeckling et al., 2005). The unidentified peaks were further analyzed through spectral comparisons with data in MassBank (http://www.massbank.jp/) with a similarity score cutoff of 0.85 or greater. The relative peak areas of metabolites were extracted using custom MET-IDEA software (Broeckling et al., 2006) and normalized against the peak areas of the internal standard. Normalization allows for the quantitative comparison of accumulated metabolites or tentative peaks.

The annotated peaks were linked to corresponding compounds in the PubChem database (http://pubchem.ncbi.nlm.nih.gov/) and KEGG compound database by manually matching compound names. We further mapped the identified compounds to trichome-related unigenes and probe set targets. This was performed by downloading the KEGG mapping files of compounds, reactions, and enzymes as well as standard KEGG Enzyme Commission (EC) number or KEGG Orthology (KO) sequence data sets from the KEGG ftp site. By referring to these mapping files, compounds were mapped to EC/KO number through reaction numbers. The compounds were further associated with unigenes and probe set target sequences by BLASTX searching these sequences against the downloaded standard EC/KO reference sequences (E-value ≤ 1e-04). TrichOME allows the user to download raw data by experiments or explore details of the peak data (Fig. 2).

**Literature Mining of Trichome-Related Genes and Proteins**

TrichOME hosts trichome-related genes and proteins curated from the published literature. The genes were initially searched using information such as gene/protein name, alias names, and annotations from the NCBI Gene Database (http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene), followed by extensive human curation. Hosted genes were further associated with trichome-related unigenes/EST sequences as well as microarray probe set target sequences by BLASTX searching these sequences against the downloaded standard EC/KO reference sequences (E-value ≤ 1e-04). TrichOME allows the user to download raw data by experiments or explore details of the peak data (Fig. 2).

**System Implementation and Web Interfaces**

The TrichOME system consists of a back-end database and a front-end Web site developed using Java. The front-end Web site interfaces integrated AJAX technology powered by YUI libraries (http://developer.yahoo.com/yui/) to improve user-server interactions.

The ESTs/unigenes and associated cDNA library information are available for batch download by species and libraries (Fig. 3A). Web tree views and comprehensive search interfaces were also implemented to facilitate the interactive queries. For example, the KEGG tree view allows users to explore trichome-related ESTs/unigenes and present the sequences by metabolic pathways (Fig. 3B). A BLAST interface enables users to search user-submitted sequences against the trichome-related sequences in the database. TrichOME allows users to download the sequences filtered by query individually or in batch. It also includes an online data submission page to collect trichome-related omics data from the research communities.

TrichOME integrates an in silico gene expression analysis algorithm to search the unique or highly expressed unigenes in trichome tissue (Fig. 4). The algorithm counts EST members of each unigene across multiple cDNA libraries to calculate an R value (Stekel et al., 2000), which represents the library specificity of the unigene.

For each hybridization experiment, TrichOME hosts both raw and normalized data, and it provides a description page with links for batch downloading. In addition, the database provides a series of “compare by...” functions for the query of hybridization signals by probe set identifiers, annotation of probe set target sequences, associated trichome-related unigenes, or the ratio of signal strength between trichome and nontrichome tissues.

The trichome-related unigenes, microarray probe sets, metabolites, and curated trichome-related genes have been mapped to each other to allow users to gain an overall systems understanding of trichome biology and biochemistry.

**Case Studies**

**Case Study 1: Mining of Putative Trichome-Specific Genes by in Silico and Microarray Gene Expression Analysis**

Because *M. truncatula* has the most (52 in total) comprehensive nontrichome cDNA libraries available as controls, this organism was chosen as an example to search for putative trichome-specific or highly preferentially expressed sequences. In the “in silico expression” of the “EST analysis” section, we selected “MT_TRI” as trichome cDNA library and all of the nontrichome cDNA libraries as a control. After clicking the Analyze button, the server calculates R values of all unigenes by comparing the abundance of gene transcripts among cDNA libraries and returns 111 unigenes with significantly higher expression levels in the trichome cDNA library than in the nontrichome cDNA libraries (R > 4; Supplemental File S1). The sequences are available for batch download by clicking the “download all” link on the same page. Users
may also select the “show unigenes only detected in trichome” option to view the unigenes only present in the trichome library; the R value statistical result will be ignored under this option.

To cross-validate the above in silico expression analysis results, the 111 unigene identifiers were copied into the “compare by unigene” section under the “microarray analysis” menu to retrieve expression signals of corresponding probe sets from all tissue-specific microarray hybridization experiments (mean values). Among the 111 unigenes, 82 unigenes, for which corresponding probe set targets were expressed at least two times higher in signal strength in trichome tissue than in nontrichome tissue (Supplemental File S2), were identified as putative trichome-related genes in *M. truncatula* for future experimental analysis.

Remarkably, among the 20 genes that were between 10- and 758-fold preferentially expressed in trichomes compared with nontrichome tissue, 16 were annotated as being involved in biotic or abiotic stress responses (Table II). These include pathogenesis-related proteins of the PR1a, -4, and -5a (thaumatin) classes as well as chitinases and endoglucanases, all of which are associated with induced responses to pathogen attack (Rigden and Coutts, 1988). Chitinases and glucanases likely modify the cell walls of invading pathogens and are often potent inhibitors of fungal growth (Mauch et al., 1988; Zhu et al., 1994). Two genes annotated as germin-like proteins were preferentially expressed 72- and 78-fold in *Medicago* trichomes. Germ and germin-like proteins may exhibit oxalate oxidase or superoxide dismutase activity, have been implicated in the generation of hydrogen peroxide during plant defense responses, and have been shown genetically to play important roles in plant defense (Lou and Baldwin, 2006; Godfrey et al., 2007). Two lipid transfer protein genes were preferentially expressed by 32- and 37-fold in the trichomes. These genes are often among the most highly expressed in plant trichomes (Lange et al., 2000; Aziz et al., 2005; Wang et al., 2008). Although they may play roles in the transfer of lipid metabolites between cell compartments, they have also been shown to act in plant defense, either through direct antimicrobial activity (Kader, 1997) or as signal transduction components (Maldonado et al., 2002). The Ser protease inhibitor gene that was preferentially expressed over 750-fold in the *Medicago* trichomes is likely involved in defense against herbivory, whereas...
the gene annotated as encoding a desiccation-related protein may help protect against abiotic stress. Genes with similar potential functions (e.g., dehydrins) are also expressed in trichomes of other species (Aziz et al., 2005).

On the basis of this preliminary analysis, the M. truncatula glandular trichomes appear to provide the plant with a defensive barrier through primary expression of high levels of proteins with direct defensive properties. This is in contrast to the trichomes from other species, which, although expressing some of the same types of defense genes (Supplemental File S3), primarily appear to harness their gene expression toward the biosynthesis and accumulation of antimicrobial secondary metabolites. For example, genes associated with terpene metabolism are among the most highly expressed in trichomes of mint, hop, and tomato (Lange et al., 2000; Wang et al., 2008; Besser et al., 2009).

We also compared the transcriptional profiles in nonglandular trichomes (of Arabidopsis) and leaves from which trichomes had been removed using the Arabidopsis ATH1 GeneChip. Six transcription factor genes (three Myb genes, two WRKY genes, and one homeodomain/Leu zipper [HD-ZIP] gene) were found in the top 20 probe sets genes that were highly expressed in trichome tissue (Table III). Furthermore, a portion of these genes or their family members are reportedly involved in trichome development (Jakoby et al., 2008). For example, AT1G79840 (HD-ZIP, GLABRA2 [GL2] gene) and AT2G37260 (WRKY, TRANSPARENT TESTA GLABRA2 gene) are well-studied regulators of leaf trichome formation (Rerie et al., 1994; Johnson et al., 2002). Interestingly, the top 20 most highly expressed genes in nonglandular trichomes are quite different from the genes that are highly expressed in glandular trichomes (Table II). The latter include many more genes that are involved in biotic or abiotic stress responses, and no transcription factor was found in the latter list.

**Case Study 2: Comparative Analysis of Monoterpenoid Biosynthesis Genes in Hop Trichomes**

Hop is a perennial, dioecious plant that belongs to the Cannabaceae family. Hop trichomes produce and accumulate large amounts of terpenoids (the mono-
terpene pinene and the sesquiterpenes humulene and caryophyllene), prenylflavonoids, and two important classes of bitter acids (humulone and lupulone derivatives), all of which give additive flavor to beer (Stevens et al., 1998; Wang et al., 2008). To study the monoterpenoid biosynthesis pathway, we first searched the hop trichome-related unigenes involved in the pathway by exploring the KEGG classification tree in the “EST analysis” section. Some of the returned unigenes, such as TCHL10609 and TCHL10281, are annotated as encoding pinene synthase (monoterpenoid synthase) and given the KEGG orthology term K07384. K07384 is located under the node representing monoterpenoid biosynthesis pathway (“Metabolism → Biosynthesis of Secondary Metabolites → Monoterpenoid Biosynthesis”). We then designed primers targeting these unigenes and successfully cloned the full-length cDNAs of the monoterpenoid synthases HlMTS1 and HlMTS2. The detailed comparative analysis of the monoterpenoid biosynthesis genes/enzymes in hop trichomes has been published (Wang et al., 2008).

Table II. Top 20 probe sets/unigenes highly expressed in trichome of M. truncatula

<table>
<thead>
<tr>
<th>Probe Set Identifier</th>
<th>Unigene Accession No.</th>
<th>Annotation</th>
<th>Ratioa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mtr.16277.1.S1_at</td>
<td>TCMT54757</td>
<td>Ser proteinase inhibitor</td>
<td>758.5</td>
</tr>
<tr>
<td>Mtr.49787.1.S1_at</td>
<td>TCMT59300 ES611317</td>
<td>Endo-1,3-β-glucanase</td>
<td>323.5</td>
</tr>
<tr>
<td>Mtr.43770.1.S1_at</td>
<td>TCMT49613</td>
<td>Pathogenesis-related protein 1a</td>
<td>242.7</td>
</tr>
<tr>
<td>Mtr.31496.1.S1_at</td>
<td>TCMT40799</td>
<td>Putative arsenate reductase</td>
<td>153.6</td>
</tr>
<tr>
<td>Mtr.22271.1.S1_at</td>
<td>TCMT48089</td>
<td>Foot protein 4 variant 2</td>
<td>153.28</td>
</tr>
<tr>
<td>Mtr.6179.1.S1_at</td>
<td>TCMT56621</td>
<td>Desiccation-related protein PCC13-62</td>
<td>144.43</td>
</tr>
<tr>
<td>Mtr.34454.1.S1_at</td>
<td>TCMT43273</td>
<td>Class lb chitinase</td>
<td>128.25</td>
</tr>
<tr>
<td>Mtr.32629.1.S1_at</td>
<td>TCMT42631</td>
<td>Thaumatin-like protein PR-5a</td>
<td>104.22</td>
</tr>
<tr>
<td>Mtr.22276.1.S1_at</td>
<td>TCMT59229</td>
<td>Unknown</td>
<td>94.63</td>
</tr>
<tr>
<td>Mtr.34878.1.S1_at</td>
<td>TCMT60298</td>
<td>Germin-like protein</td>
<td>77.63</td>
</tr>
<tr>
<td>Msr.1226.1.S1_at</td>
<td>CO513853</td>
<td>Germin-like protein</td>
<td>71.59</td>
</tr>
<tr>
<td>Mtr.44506.1.S1_x_at</td>
<td>TCMT59300</td>
<td>Endo-1,3-β-glucanase</td>
<td>66.05</td>
</tr>
<tr>
<td>Mtr.10325.1.S1_at</td>
<td>TCMT41963</td>
<td>Chitinase-related agglutinin</td>
<td>63.20</td>
</tr>
<tr>
<td>Mtr.35661.1.S1_at</td>
<td>TCMT59126</td>
<td>Probable lipid transfer protein family protein</td>
<td>54.78</td>
</tr>
<tr>
<td>MsrAffx.3538.1.S1_at</td>
<td>TCMT42628 TCMT40025</td>
<td>Putative senescence-associated protein</td>
<td>40.86</td>
</tr>
<tr>
<td>Mtr.8763.1.S1_at</td>
<td>ES61367</td>
<td>Thaumatin-like protein PR-5a</td>
<td>38.56</td>
</tr>
<tr>
<td>Mtr.34695.1.S1_at</td>
<td>TCMT55690</td>
<td>Plant lipid transfer/seed storage/trypsin-α-amylase inhibitor</td>
<td>37.48</td>
</tr>
<tr>
<td>Mtr.34695.1.S1_x_at</td>
<td>ESX25955</td>
<td>Plant lipid transfer/seed storage/trypsin-α-amylase inhibitor</td>
<td>32.14</td>
</tr>
<tr>
<td>Mtr.22715.1.S1_at</td>
<td>TCMT43191</td>
<td>elf4G eukaryotic initiation factor 4, Nic domain-containing protein</td>
<td>27.21</td>
</tr>
<tr>
<td>Mtr.39139.1.S1_at</td>
<td>TCMT52856</td>
<td>Pathogenesis-related protein 4a</td>
<td>19.37</td>
</tr>
</tbody>
</table>

aRatio represents the ratio of average signal strength between trichome tissue and nontrichome tissues.
DISCUSSION

The genes expressed in trichomes may be underrepresented in non-tissue-targeted libraries due to the distinct features of trichome tissue in secondary metabolism. That is to say, many of the trichome-related transcripts may not appear in regular EST databases because of their relative low expression levels in whole plant tissue. For example, 68% of ESTs in the *M. truncatula* trichome cDNA library are singletons or belong to trichome-specific unigenes, whereas the average value is 15% for all 52 nontrichome cDNA libraries. Therefore, the unique data in TrichOME are a resource to the trichome biology and plant defense research communities and are also expected to be a valuable resource for the annotation/reannotation of “model” plant genomes, especially for the model legume *M. truncatula*.

Since EST sequences are often short and of low quality, we have included additional sequences from EST libraries of nontrichome tissues as well as individual cDNA sequences to generate longer and more credible unigene sequences. These nontrichome EST libraries were also used as control libraries for in silico analysis of trichome-specific genes. We will routinely update the TrichOME database when more trichome or nontrichome EST sequences are available in public repositories, aiming to further improve the quality of unigene assembly and comparative analysis of trichome genes.

TrichOME hosts Metabolomics Standards Initiative (Sansone et al., 2007)-compliant data. The GC-MS-based metabolite profiles are much more reproducible than LC-MS with respect to retention time and spectra of peaks. The retention time index of each peak can be calibrated based on reference compounds and therefore enables comparison across experiments. Meanwhile, the spectrum data of each peak from GC-MS are less variable and are comparable across laboratories, because the common electrical ionization protocol (70 eV) is applied for spectrum generation. In the TrichOME database, some of the peaks were identified based on the standard electrical ionization spectrum libraries, for example, our internal library and the MassBank (http://www.massbank.jp/).

TrichOME also provides tools for the analysis of microarray data and unigenes in addition to hosting sequences, gene expression profiles, metabolite profiles, and curated trichome-related genes. The data from different sections have been carefully cross-mapped to each other to enable users to cross-reference the data in different sections. For example, TrichOME hosts the Affymetrix Medicago GeneChip hybridization results from two organisms. The probe sets are mapped with the curated trichome-related genes from the literature and are also linked to unigenes. These data and analytical functions enable users to identify the differentially expressed genes between trichome and nontrichome tissues.

Transcription factors and transporters are receiving extensive interest from the trichome research community, as the former may regulate secondary metabolic pathways and the latter function to transfer natural products across the plasma membrane and tonoplast. TrichOME annotates unigenes based on the published PlantTFDB and TCDB databases (Saier et al., 2006; Guo et al., 2008). These in-depth annotations facilitate the

<table>
<thead>
<tr>
<th>Probe Set Identifier</th>
<th>Gene Accession No.</th>
<th>Annotation</th>
<th>Signal Ratio (Trichome-Leaf)</th>
</tr>
</thead>
<tbody>
<tr>
<td>249408_at</td>
<td>At5g40330</td>
<td>Myb-related protein</td>
<td>2.65</td>
</tr>
<tr>
<td>245181_at</td>
<td>At5g12420</td>
<td>Putative protein similarity to various predicted proteins, contains ATP synthase-δ (OSCP) subunit signature AA211-230</td>
<td>2.61</td>
</tr>
<tr>
<td>257490_x_at</td>
<td>At1g01380</td>
<td>Myb homolog</td>
<td>2.55</td>
</tr>
<tr>
<td>260166_at</td>
<td>At1g79840</td>
<td>GL2, HD-ZIP protein, GL2 gene</td>
<td>2.40</td>
</tr>
<tr>
<td>260386_at</td>
<td>At1g74010</td>
<td>Putative strictosidine synthase</td>
<td>2.36</td>
</tr>
<tr>
<td>267459_at</td>
<td>At2g33850</td>
<td>Unknown protein</td>
<td>2.31</td>
</tr>
<tr>
<td>256359_at</td>
<td>At1g66460</td>
<td>Protein kinase, Plasm profile PF00069</td>
<td>2.31</td>
</tr>
<tr>
<td>265954_at</td>
<td>At2g37260</td>
<td>Putative WRKY-type DNA-binding protein</td>
<td>2.31</td>
</tr>
<tr>
<td>264387_at</td>
<td>At1g11990</td>
<td>Putative growth regulator protein (axi 1 gene)</td>
<td>2.24</td>
</tr>
<tr>
<td>253392_at</td>
<td>At4g32650</td>
<td>Potassium channel protein AtKC potassium channel</td>
<td>2.24</td>
</tr>
<tr>
<td>253595_at</td>
<td>At4g30830</td>
<td>M1 protein, Streptococcus pyrogenes, PIR2:S46489</td>
<td>2.23</td>
</tr>
<tr>
<td>263775_at</td>
<td>At2g46410</td>
<td>Putative MYB family transcription factor</td>
<td>2.23</td>
</tr>
<tr>
<td>263480_at</td>
<td>At2g04032</td>
<td>Putative root iron transporter protein</td>
<td>2.23</td>
</tr>
<tr>
<td>265008_at</td>
<td>At1g61560</td>
<td>Miq protein</td>
<td>2.21</td>
</tr>
<tr>
<td>248236_at</td>
<td>At5g3870</td>
<td>Putative phytoacterin/early nodulin-like protein</td>
<td>2.19</td>
</tr>
<tr>
<td>248896_at</td>
<td>At5g46350</td>
<td>WRKY-type DNA-binding protein</td>
<td>2.17</td>
</tr>
<tr>
<td>246000_at</td>
<td>At5g20820</td>
<td>Putative protein-predicted proteins</td>
<td>2.15</td>
</tr>
<tr>
<td>259410_at</td>
<td>At1g13340</td>
<td>Hypothetical protein</td>
<td>2.14</td>
</tr>
<tr>
<td>266544_at</td>
<td>At2g35300</td>
<td>Late embryogenesis abundant proteins</td>
<td>2.13</td>
</tr>
</tbody>
</table>
further study of genes involved in secondary meta-
bolic pathways.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental File S1. Highly expressed unigenes in the trichome cDNA library compared to nontrichome cDNA libraries in M. truncatula.

Supplemental File S2. Putative trichome-related unigenes in M. truncatula.

Supplemental File S3. Cross-species comparison of preferentially expressed M. truncatula trichome-related unigenes.

ACKNOWLEDGMENTS

The authors thank Drs. Haiquan Li, Mohamed Bedair, and Zhentian Lei for their comments on mass spectral data analysis, Joshua Smith and Zhaozhong Zhuang for literature curation, and Kevin Staggs and Shane Porter for Web interface improvement.

Received August 4, 2009; accepted November 19, 2009; published November 25, 2009.

LITERATURE CITED

Ahlstrand G (1996) Low-temperature low-voltage scanning microscopy (LTLVSEM) of uncoated frozen biological materials: a simple alterna-


Alonso WR, Rajanarivony JF, Gershenson J, Croteau R (1992) Purifica-
tion of 45-limonene synthase, a monoterpene cyclase from the glandular
trichomes of peppermint (Mentha piperita) and spearmint (Mentha sppicata). J Biol Chem 267: 7582–7587

Arikan NA, Jawad FM, Karam NS, Shibii RA (2004) Micropropagation and
accumulation of essential oils in wild sage (Salvia fruticosa Mill.). Sci Hortic (Amsterdam) 100: 193–202


Aschner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis

Aziz N, Paiva NL, May GD, Dixon RA (2005) Transcriptome analysis of
allafia glandular trichomes. Planta 221: 28–38

Benedito VA, Torres-Jarez I, Murray JD, Andriankaja A, Allen S, Kakar
K, Wandle M, Verdis J, Zubieta C, Chen F, Beuerle T, Lewinsohn E, Noel JP,
Pichersky E (2002) Characterization of phenylpropene and terpenyltrans-
feresaes from sweet basil: facile change of substrate specificity and convergent evolution within a plant O-methyltransferase family. Plant Cell 14: 505–519

protein (VvGLP3) is closely linked to the site of Erysiphe necator infection: a possible role in defense? Mol Plant Microbe Interact 20: 1112–1125

Garland DR, Lavid N, Zubieta C, Chen F, Beuerle T, Lewinsohn E, Noel JP,
Pichersky E (2002) Characterization of phenylpropene O-methyltrans-
feresaes from sweet basil: facile change of substrate specificity and convergent evolution within a plant O-methyltransferase family. Plant Cell 14: 505–519


Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork


Summaries of Affymetrix GeneChips probe level data. Nucleic Acids Res 31: e15

Jakoby MJ, Falkenhagen D, Mader MT, Brininstool G, Wischnitzki E, Plat

GLABRA2, a trichome and seed coat development gene of
Arabidopsis. Plant Cell 14: 66–70

tion of 4S-limonene synthase, a monoterpene cyclase from the glandular

LVC, Chen J-C (2000) Fast track to the trichome: induc-
tion of N-acyl nornicotines precedes nicotine induction in Nicotiana
reppanda. Planta 210: 510–514


Dixon RA, Sumner LW (2005) Metabolic profiling of Medicago truncu-

Dell B, McComb A (1978) Plant resins: their formation, secretion and

aristolomin and arisitomin in fowei tissues of glanded and glandless


and the closely related AtGL1380 gene in controlling Arabidopsis trichome
patterning. Plant J 40: 860–869


Falarza F, Fotopoulos V, Margaritis T, Anastasaki T, Pateraki I, Bosabalidis
AM, Kafetzopoulos D, Demetzos C, Pichersky E, Kanelis AK (2008)
Transcriptome analysis approaches for the isolation of trichome-specific genes from the medicinal plant Cistus creticus subsp. creticus. Plant Mol Biol 68: 633–651

Fridman D, Wang J, Iijima Y, Frechlich JE, Gang DR, Ohlrogge J,
Pichersky E (2005) Metabolic, genomic, and biochemical analyses of
glandular trichomes from the wild tomato species Lycopersicon hirsutum
identify a key enzyme in the biosynthesis of methylketones. Plant Cell 17: 1252–1267

Gang DR, Lavid N, Zubieta C, Chen F, Beuerle T, Lewinsohn E, Noel JP,
Pichersky E (2002) Characterization of phenylpropene and terpenyltrans-
feresaes from sweet basil: facile change of substrate specificity and convergent evolution within a plant O-methyltransferase family. Plant Cell 14: 505–519

Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res 31: e15

Jakoby MJ, Falkenhagen D, Mader MT, Brininstool G, Wischnitzki E, Plat

GLABRA2, a trichome and seed coat development gene of
Arabidopsis, encodes the WRKY transcription factor. Plant Cell 14: 1359–1375


information to genomes and the environment. Nucleic Acids Res 36: D485–D484

Lange BM, Wildung MR, Staubier EJ, Sanchez C, Pouchnik D, Croteau R
expression of evaluated sequence tags from mint glandular trichomes. Proc Natl Acad Sci USA 97: 2934–2939

Laue G, Preston C, Baldwin L (2000) Fast track to the trichome: induc-
tion of N-acyl nornicotines precedes nicotine induction in Nicotiana
reppanda. Planta 210: 510–514

Lee Y, Tsai J, Sunkara S, Karamycheva S, Pertea G, Sultana R, Antonescu

TrichOME: a Comparative Omics Database for Plant Trichomes


