Transcriptome analysis in *Cucumis sativus* identifies genes involved in multicellular trichome development

Jun-Long Zhao, Jun-Song Pan, Yuan Guan, Jing-Tao Nie, Jun-Jun Yang, Mei-Ling Qu, Huan-Le He, Run Cai *

School of Agriculture and Biology, Shanghai Jiao Tong University, 800 Dongchuan Road, Minhang District, Shanghai 200240, China

ABSTRACT

The regulatory gene network of unicellular trichome development in *Arabidopsis thaliana* has been studied intensively, but that of multicellular remains unclear. In the present study, we characterized cucumber trichomes as representative multicellular and unbranched structures, but in a spontaneous mutant, *mict* (*micro-trichome*), all trichomes showed a micro-size and stunted morphologies. We revealed the transcriptome profile using Illumina HiSeq 2000 sequencing technology, and determined that a total of 1391 genes exhibited differential expression. We further validated the accuracy of the transcriptome data by RT-qPCR and found that 43 genes encoding critical transcription factors were likely involved in multicellular trichome development. These 43 candidate genes were subdivided into seven groups: homeodomain, MYB-domain, WRKY-domain, bHLH-domain, ethylene-responsive, zinc finger and other transcription factor genes. Our findings also serve as a powerful tool to further study the relevant molecular networks, and provide a new perspective for investigating this complex and species-specific developmental process.

1. Introduction

Trichomes are highly specialized structures developed from the protodermal cells of most terrestrial plants. They play important biotic and abiotic roles in many aspects, such as insect, herbivore and microbe deterrence; water regulation through transpiration; light reflectance (including UV) [1,2]; pollen collection and dispersal; absorption of water and nutrients [3]; secretion of ions and pollutant metals [4,5]; reduced mechanical abrasion; and regulation of surface temperature [2]. Trichomes can be categorized according to whether they are unicellular or multicellular, glandular or glandless, and branched or unbranched [3]. They also provide a suitable model system for researching cell differentiation at the single-cell level, including cell fate determination, cell cycle developmental control, and cell morphogenesis [6,7].

In *Arabidopsis thaliana*, the unicellular trichome differentiation is thought to be regulated by a competitive system comprising promoting and limiting activities. The promoting activity up-regulates both activities, and the limiting activity spreads via cell-to-cell communication and inhibits trichome differentiation [8–10]. Critical positive transcription factors include GL1 (GLABRA1), which acts as an R2R3 MYB protein [11]; GL3 (GLABRA3) and its homolog EGL3 (ENHANCER OF GLABRA3), which act as basic helix-loop-helix proteins [12]; and TTG1 (TRANSPARENT TESTA GLABRA1), which acts as a WD40-repeat protein [13]. In contrast, negative transcription factors include TRY (TRIPTYCHON), CPC (CAPRICE), and ETC (ENHANCER OF TRY AND CPC), which belong to the small R3 single-repeat MYB family [14,15]. In this system, GL1, GL3/EGL3 and TTG1 constitute the promoting activity, TRY/CPC/ETC, GL3/EGL3 and TTG1 constitute the limiting activity, and GL2 (GLABRA2), a homeodomain protein, acts as a quantitative factor directly regulated by both activities [6,7,9,16].

Cucumber (*Cucumis sativus* L., 2n = 2x = 14), an annual sprawling herbaceous plant, is one of the most commercially important vegetable crops worldwide. Cucumber also serves as a model plant for sex determination studies due to its diverse sex types [17]. Trichomes are widely found on leaves, stems, flowers, tendrils and fruits of the wild type cucumber plants. The trichomes on the fruits are commonly called "fruit spines" (Fig. 1A, B). Cucumber fruits are economically valuable and fruit spines directly affect the appearance quality. Here, a spontaneous mutant, which originated from North China inbred line 06-1, presented a glabrous phenotype on leaves, stems, flowers, tendrils and fruits (Fig. 1C, D). Scanning electron microscopy revealed that all trichomes in this mutant exhibited a micro-size and stunted morphologies, and are only visible under at least 20 times magnification, thus, we named this gene "*Mict* (*Micro-trichome*)" and this line "mict". Compared with unicellular trichomes, the regulatory mechanisms of multicellular trichome development in plants are much less understood. In this study, we characterized cucumber trichomes as multicellular and unbranched. We further used Illumina HiSeq 2000 sequencing technology to reveal the transcriptome changes between the *mict* mutant and wild type, and identified a series of candidate genes encoding critical transcription factors which are likely involved in multicellular trichome development.

* Corresponding author. Fax: +86 21 34206938.
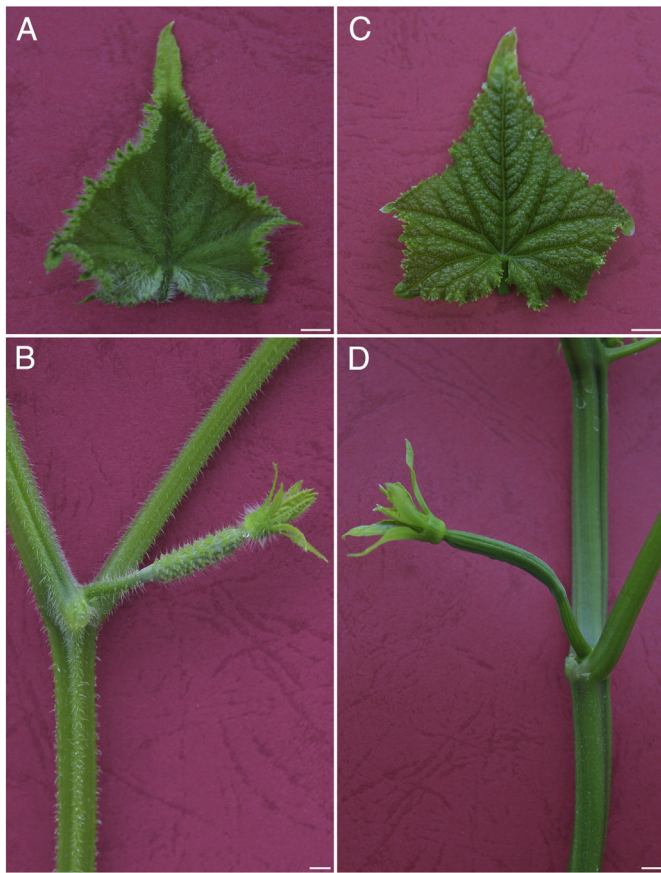 *E-mail address:* cairun@sjtu.edu.cn (R. Cai).

Fig. 1. Trichome phenotypes between the wild type and *mict* mutant. (A, B) Line 06-1 (wild type). (C, D) Line 06-2/mict (the *mict* mutant). The absence of normal trichomes can be seen in the leaf, stem, branch, flower and fruit of the *mict* mutant. Scale bars represent 5 mm.

2. Results

2.1. Trichome morphologies controlled by Mict

Scanning electron microscopy imaging showed that the wild type cucumber trichomes consisted of three distinct cell types: (a) apical cell: the apical cell could be classified into two types, the majority was a single-celled, non-glandular, and pyramid-shaped cell; the minority was a glandular secreting head (Fig. 2A); (b) stalk cell: the stalk was generally composed of two to four elongated cylindrical-shaped cells; and (c) base cell: the base was pie-shaped at the bottom connected with the epidermal cells (Fig. 2C). The trichomes in the *mict* mutant, however, were much smaller and only visible under a microscope (at least 20 times magnification). They presented two morphologies, the majority (type i) had only a small papillar-shaped head, and the minority (type ii) divided abnormally, resulting in a structure consisted of one to five rounded cells, but without the pyramid-shaped head and the pie-shaped base (Fig. 2B, D, F). The fruit spine had a similar structure with the leaf trichome, but the base was much more inflated and divided into multiple spherical cells (Fig. 2E). We also characterized the root hairs as single-celled, unbranched, elongated and soft structure with numerous small tumors attached, but there was no difference between the wild type and *mict* mutant (Fig. 2G, H), indicating that *Mict* functions in trichome differentiation of the aerial organs, such as leaves and fruits, rather than the underground organ root.

2.2. Genetic analysis of Mict

A cross between the wild type and *mict* mutant generated F_1 descendants that all presented a wild type trichome phenotype, and in their F_2 segregating population, 2032 of 7936 exhibited the mutated phenotype ($\chi^2 = 1.548 < \chi^2_{0.05,1} = 3.84$); in addition, a test cross yielded 122 descendants comprising 59 mutated individuals ($\chi^2 = 0.131 < \chi^2_{0.05,2} = 5.991$), which closely fit the segregation ratios of 3:1 and 1:1, respectively, indicating that the mutation is recessive and *Mict* acts as a single dominant nuclear gene.

2.3. Mapped reads and annotated genes

We sequenced the apical leaf (1.5 cm length, 21 days old, trichomes attached with epidermis) transcriptome between the *mict* mutant and its wild type background line on the Illumina HiSeq 2000 platform. Sequences of the two cDNA libraries generated 102.33 and 105.40 million high quality reads, respectively, and the average read length was 100 bp. A total of 93.89 (91.75%) and 98.16 (93.13%) million reads were mapped to the cucumber genome, including 88.09 (93.82%) and 92.03 (93.76%) million unique reads, and 5.80 (6.18%) and 6.12 (6.24%) million multiple reads, in which 91.23 (82.47%) and 95.67 (83.22%) million were mapped to genes, 88.99 (97.54%) and 93.17 (97.38%) million were mapped to exons, and 19.39 (17.53%) and 19.29 (16.78%) million were mapped to intergenic regions, respectively (Supplementary Table S1). All useful reads were assembled by Cufflinks program and annotated via BLAST against the cucumber database. As a result, a total of 23,454 genes were predicted including 20,040 annotated and 3414 unannotated (Supplementary dataset).

2.4. eggNOG functional category analysis

We used eggNOG (evolutionary genealogy of genes: Non-supervised orthologous groups) to classify orthologous genes with functional descriptions. A total of 14,857 (74.14%) genes were categorized into 25 eggNOG (Fig. 3). Among all eggNOG, unfortunately, "Function unknown" and "General function prediction only" still represented the largest clusters in cucumber species, which had 4020 (27.06%) and 2627 (17.68%) genes, respectively. "Signal transduction mechanisms" 1059 (7.13%), "Posttranslational modification, protein turnover, chaperones" 987 (6.64%), and "Transcription" 789 (5.31%) clusters were following. "Extracellular structures" 11 (0.07%) and "Cell motility" 7 (0.05%) clusters had the fewest orthologous genes (Supplementary Table S2).

2.5. Differential expression, function and pathway enrichment analyses

Gene expression levels were calculated by baseMean values and differential expression was defined by statistical parameters ($P < 0.05$ and fold change >2 or <-2). As a result, a total of 1391 genes exhibited differential expression, including 966 up-regulated and 425 down-regulated (Fig. 4, Supplementary dataset).

To explore the biological functions of these differentially expressed genes, GO (Gene ontology) enrichment analysis was carried out. Among all 53 GO terms, "Sequence-specific DNA binding transcription factor activity" ($P = 1.19E-09$, 29 up-regulated, 7 down-regulated) was the most enriched cluster, "Extracellular region" ($P = 1.77E-04$, 7 up-regulated, 7 down-regulated) and "External encapsulating structure" ($P = 1.26E-03$, 8 up-regulated, 4 down-regulated) clusters were also significantly enriched (Fig. 5, Supplementary Table S3).

We also performed KEGG (Kyoto encyclopedia of genes and genomes) enrichment analysis to determine whether multicellular trichome-related genes were involved in specific pathways, and a total of 4720 genes were annotated into 24 KEGG categories. The most enriched category was "Biosynthesis of other secondary metabolites" ($P = 5.33E-03$, 106 genes including 7 differentially expressed), "Environmental adaptation" ($P = 5.53E-02$, 103 genes including 5 differentially expressed) and "Lipid metabolism" ($P = 6.25E-02$, 249 genes including 9 differentially expressed) categories were following (Fig. 6, Supplementary Table S4).