Genomics 102 (2013) 388–396

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Transcriptome analysis reveals novel genes potentially involved in photoperiodic tuberization in potato

Jianwei Shan, Wei Song, Jun Zhou, Xiaohuan Wang, Conghua Xie, Xiaoxi Gao, Tingting Xie, Jun Liu *

National Center for Vegetable Improvement (Central China), People's Republic of China
Key Laboratory of Horticultural Plant Biology (Huazhong Agricultural University), Ministry of Education, People's Republic of China
Potato Engineering and Technology Research Center of Hubei Province, People's Republic of China
Huazhong Agricultural University, Wuhan, Hubei 430070, People's Republic of China

ARTICLE INFO

Article history:
Received 28 March 2013
Accepted 2 July 2013
Available online 13 July 2013

Keywords:
Solanum tuberosum
Tuber
Photoperiod
Transcriptome

ABSTRACT

Potato microtuber produced in vitro provides a model system to investigate photoperiod-dependent tuberization. However, the genes associated with potato tuberization remain to be elucidated. The present research involved three potato clones with distinct tuberization response to changes of photoperiod. Digital Gene Expression (DGE) Tag Profiling analysis of the short-day-sensitive clone identified 2218 genes that were regulated by day length. Both GO and KEGG pathway analysis provided insights into predominant biological processes and pathways, and enabled the selection of 56 genes associated with circadian rhythmicity, signal transduction, and development. Quantitative transcriptional analysis in the selected clones revealed 5 genes potentially associated with photoperiodic tuberization, which were predicted to encode a DOF protein, a blue light receptor, a lectin, a syntaxin-like protein, and a protein with unknown function. Our results strongly suggest that potato tuberization may be largely controlled by the homologs of genes shown to regulate flowering time in other plants.

1. Introduction

Potato (*Solanum tuberosum* L.), a member of the Solanaceae, ranks the fourth most important food crop after maize, rice and wheat (http://www.fao.org/). Potato tubers—the modified underground stems that are rich in starch, protein, antioxidants, and vitamins—are the most important organ of the species. Tuber formation has thus attracted considerable attention, with a view to improve yields and provide a model system to look into the development of modified organs in plants.

Potato tuberization is affected by day length and temperature. Cultivated potatoes (*S. tuberosum* ssp. *tuberosum*) derived from Chilean landraces are capable of tuberization under long day conditions. However, some potato species, such as *S. tuberosum* ssp. *andigena*, only tuberize under short day conditions [1]. Under favorable environmental conditions, the stolon arrests longitudinal elongation and triggers radial swelling at the subapical region. This phenomenon was suggested to be controlled by specific tuber-formation stimuli in the plant under certain conditions of photoperiod and temperature [2,3]. Although extensive literature is available, the mechanism of tuberization is still not fully understood.

Interspecific grafting and genetic transformation suggest that tuberization in potato and flowering in many plants are regulated by similar or even identical long-distance signals [1,4,5]. Therefore, it was hypothesized that the response of potato tuber initiation to photoperiod may involve the circadian clock that enables plants to sense the changes of day length. Photoreceptors, especially phytochromes and cytochromes, function as prominent inputs to the clock, and entrain its phase to diurnal day/night cycles [1,5]. It is speculated that under inductive day lengths, potato leaves generate long-distance systemic signals. These signals are loaded into the phloem system and transported to the subapical region of the underground stolon to modulate the expression of genes that induce the differentiation of stolon meristem cells to a tuber fate [5,6]. However, the identity of the transmissible signal molecule or molecules remains elusive.

CONSTANS (*CO*) encodes a nuclear protein similar to zinc finger transcription factors [6]. In *Arabidopsis*, the expression of *CO* was reported to be clock-regulated by GIGANTEA [7]. In the presence of light, both phytochrome and cryptochrome photoreceptors prevent the degradation of CO, which accelerates flowering time in long day conditions, at least partially by inducing the expression of *FLOWERING LOCUS T* (*FT*) [8]. In potato, orthologues of *CO* and *FT* were also identified and proposed to be involved in photoperiodic regulation of

Abbreviations: DGE, digital gene expression; DEG, differentially expressed gene; DOF, DNA binding with one finger; CDF, CYCLING DOF FACTOR; LOV, light-oxygen-voltage; FMN, flavin mononucleotide; LKP2, LOV KELCH REPEAT PROTEIN 2; JAME, jasmonic acid methyl ester; TA, tuberonic acid; SNARE, soluble *N*-ethylmaleimide-sensitive factor attachment protein receptors.

* Corresponding author at: Huazhong Agricultural University, Wuhan, Hubei 430070, People's Republic of China. Fax: +86 27 87286939.
E-mail addresses: chanyujianwei@126.com (J. Shan), songwei58@126.com (W. Song), zhoujun820830@yahoo.cn (J. Zhou), wxh123haha@163.com (X. Wang), xiech@mail.hzau.edu.cn (C. Xie), gaoxiaoxi214@hotmail.com (X. Gao), tingtingxie@yahoo.cn (T. Xie), liujun@mail.hzau.edu.cn (J. Liu).

0888-7543/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ygeno.2013.07.001

potato tuberization [4,6]. Furthermore, two mobile RNA molecules, *miR172* and *StBEL5*, were also reported to act as long-distance transmissible signals that were related to day-length control of tuber formation [9,10]. Phytochrome B is a photoreceptor that is stably accumulated in green leaves, and is likely to be involved in sensing day length [1]. Photoperiodic control of tuberization was abolished in *S. tuberosum* ssp. *andigena* plants in which phytochrome B levels were reduced by antisense-mediated suppression. These plants could tuberize under either short day conditions or long day conditions, as well as short day conditions supplemented with a night break [11]. Tuberization requires coordination of environment cues, plant photoreceptors, and the genes activated in downstream pathways. However, photoperiodic control of potato tuber formation is far from being fully elucidated. New genomics tools, especially high-throughput sequencing, should considerably facilitate the identification of candidate genes involved in the photoperiodic tuberization.

In past years, the next generation sequencing based genome-wide gene expression profilings, such as RNA-Seq and Digital Gene Expression (DGE), have dramatically improved the efficiency and speed of gene discovery. The main advantages of DGE have been addressed. Firstly, it is high-throughput, as a result, it allows to analyze genome wide gene expression including weakly expressed genes which cannot be assessed by microarrays [12]. Secondly, it is digital quantification, which is directly comparable across different experiments [13]. Thirdly, it is highly replicable and accurate. Finally, it is not requiring presynthesized oligonucleotide probes [14]. However, there are some disadvantages of DGE as it preferentially generates sequencing tags from the 3′end of the transcripts, it cannot detect alternative splice variants, and it is limited to the analysis of genes containing CATG site. Despite its disadvantages, DGE is a useful technology in the analysis of gene expression at an unprecedented level of sensitivity, especially in comparing two very similar samples [13]. In a direct comparison with RNA-seq, both methods provide similar assessments of relative transcript abundance. However, DGE better detects expression differences for poorly expressed genes and does not exhibit transcript length bias [15].

To identify genes that are potentially involved in photoperiodic control of potato tuberization with DGE, the present research used three potato clones which are derived from the same cross but tuberize differently in response to changes of day length. The results will deepen our understanding of the mechanism underlying potato photoperiodic tuberization.

2. Plant materials and methods

2.1. Plant materials

The offspring of the tetraploid potato (2n = 4x = 48) cross 393075.54 × 391679.12 (seeds were kindly offered by the International Potato Center) were previously characterized for tuberization ability in vitro. Three clones, E26 (tuberizes only under short day conditions), E109 (tuberizes under both short day conditions and long day conditions), and E20 (tuberizes neither under short day conditions nor under long day conditions), were selected to identify the genes related to tuberization.

Single stem nodes of 4-week-old plantlets, from which the leaves were excised, were cultured for 3 weeks on MS [16] medium supplemented with 8% sucrose, 0.7% agar and 0.08% activated carbon under 16 h light/8 h darkness (light intensity 83 μmol m^{-2} s^{-1}) at 20 ± 1 °C. They were then subjected to different photoperiod treatments by moving to short day conditions (8 h light/16 h darkness) or long day conditions (16 h light/8 h darkness) at the same temperature. The entire plants were sampled at 0, 2 and 5 d, immediately frozen with liquid nitrogen and stored at −76 °C until RNA extraction. The remaining plantlets of each treatment were kept under the same conditions to confirm the tuberization response to photoperiod.

2.2. RNA extraction and DGE tag profiling

Samples of potato clone E26 were taken at 0 d (E26d0, before the photoperiod treatment), 2 d (E26S2 and E26L2, 2 d grown under short day conditions and long day conditions, respectively) and 5 d (E26S5 and E26L5, 5 d grown under short day conditions and long day conditions, respectively). They were used to select differentially expressed genes (DEGs) between the photoperiods. Total RNA was extracted from each sample as previously described [17]. The mRNA was isolated from 6 μg of total RNA using oligo(dT) magnetic beads, and then double-stranded cDNA was synthesized after reverse transcription using oligo(dT) primer bound to magnetic beads. Two restriction enzymes were used to generate sequencing tags. First the cDNA was digested with *Nla*III that recognizes and cleaves at CATG sites of cDNA, then the Illumina adapter 1 was annealed to the 5′ end of the purified cDNA fragment that was attached to the magnetic beads. The junction of Illumina adapter 1 and CATG generates the recognition site of *Mme*I. Thereafter, *Mme*I was used to cut the cDNA fragments 17 bp downstream of the CATG site. After removing the cDNA fraction by virtue of its association with magnetic beads, the Illumina adapter 2 was introduced to the 3′ends of the tags. Fifteen cycles of PCR amplification were carried out to enrich the sample for 21 bp tags. Subsequently, massively parallel sequencing-by-synthesis was performed using an Illumina Genome Analyzer. Sequencing data were deposited in the GEO database of NCBI under the submission number of GSE45008 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=dhajfcmakqemslm&acc=GSE45008).

2.3. Data processing and statistical analysis

Prior to bioinformatics analysis, empty reads, low quality tags (tags containing unknown bases and tags with more than 50% bases of quality value less than 10), tags that were either too long or too short, and single-copy tags were filtered out to generate clean tags. Thereafter, sequencing data was evaluated by assessing the distribution of tag expression, the saturation of sequencing data, and experimental reproducibility.

The potato genome was sequenced and assembled in 2011 [18]. In the present study, the clean tags were mapped to the reference genes of potato genome available at http://solgenomics.net/organism/Solanum_tuberosum/genome. To annotate clean tags, a preprocessed reference database of all possible CATG + 17 nucleotides tag sequences was created using the reference genes. Subsequently, all clean tags were mapped to the reference database using SOAP2 [19], and no more than 1 nucleotide mismatch was allowed. The number of clean tags mapped to each gene was calculated and then normalized to the number of transcripts per million clean tags (TPM) [20,21].

To identify differentially expressed genes between two samples, the rigorous algorithm developed by Audic and Claverie [22] was applied to the present research. Briefly, the copy number of gene A was denoted as *x*. As every gene's expression occupies only a small part of the library, the proportion of Gene A, $p(x)$, can be estimated by the Poisson distribution:

$$p(x) = \frac{e^{-\lambda}\lambda^{x}}{x!} (\lambda \text{ is the real transcripts of the gene}).$$

When sample 1 and sample 2 respectively have the total clean tags N1 and N2, if gene A is mapped by x tags from N1 and y tags from N2, theoretically, the probability of gene A expressed equally between the two samples can be calculated with:

$$p(y|x) = \left(\frac{N_2}{N_1}\right)^{y} \frac{(x+y)!}{x!y!\left(1+\frac{N_2}{N_1}\right)^{(x+y+1)}}.$$