***De novo* transcriptome sequencing of genome analysis provides insights into *Solidago canadensis* invasive capability through** **photosynthesis**

Running title: Photosynthesis promote alien plant invasive

Yang Zhang1,2, Xiaojuan Wang1, Peng Nan3, Jinglong Li4, Alan C. Gange5, Liang Jin1

1Natural History Research Center, Shanghai Natural History Museum, Shanghai Science & Technology Museum, Shanghai 200041, China;

2Shanghai Association for Science & Technology, Jingan District, Shanghai 200070, China;

3Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, Institute of Biodiversity Science, Fudan University, Shanghai 200433, China

4School of Life Sciences, Fudan University, Shanghai 200433, China

5School of Biological Sciences, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK

Corresponding author:

Xiaojuan Wang

Natural History Research Center, Shanghai Natural History Museum,

Shanghai Science & Technology Museum,

399 Shanhaiguan Road, Jingan District,

Shanghai 200041,

P.R. China;

Tel: +86-13817638879

E-mail: wangxj@sstm.org.cn

**Abstract:** *Solidago canadensis* is one of the most destructive invasive plants in the East of China, yet nothing is known about its photosynthetic ability at the molecular level. In order to examine the mechanism aiding invasion from the photosynthetic and molecular perspective, transcriptome sequencing and *de novo* assembly of *S. canadensis* and its congener *S. decurrens* (native, non-invasive species) were compared in a bioassay experiment. Results showed that *S. canadensis* has higher biomass and net photosynthetic rates than those of the nativeplant, *S. decurrens* (*P*<0.05). Based on RNA-seq data, the genes which are closely related to light absorption and electron transfer in *S. canadensis* were observed to be up-regulated compared with those from *S. decurrens*, including the process of photosystems I and II, such as *PsbP*, *PsbQ*, *Psb27*, *PsbW*, *PsaG*, *PetE*, and light-harvesting complexes (LHCs). The RT-qPCR verification results of key genes were similar to those from transcriptome. Therefore, our results might partly explain the successful invasion of *S. canadensis* in China at the level of transcriptome, leading to its enhanced photosynthetic capability.

**Key words:** *Solidago canadensis*, *Solidago decurrens*, *de novo*, comparative transcriptomics, photosynthesis

**Introduction**

*Solidago canadensis* (Asteraceae) is a perennial clonal herb, originating from North America. It has been intentionally introduced into New Zealand, Australia, Europe and parts of Asia, and especially in the East of China (Li and Xie 2002; Dong et al. 2006). *S. canadensis* was first introduced into Jiangsu province, China in 1913 as an ornamental plant (Jin et al. 2004), and in the next 50 years, this species rapidly expanded into all types of land areas in the East of China and also the Central of China (Zhao et al. 2015). Currently, *S. canadensis* is abundant in many natural habitats, including roadsides, farmlands, woodlands, and city suburbs (Li & Xie 2002), presenting a huge threat to the biodiversity and abundance of the native vegetation.

Compared with its native congeneric species, *S. canadensis* can be adapted to many extreme environments, such as drought stress, salt stress and high temperature, which enables it to demonstrate high invasion capability (Hu et al. 2007). Previous studies have shown that there were several factors resulting in its successful invasion. Firstly, the potential capability of rapid growth and prolific reproduction (both sexual and asexual) clearly contribute to its successful invasion (Xu et al. 2006; Yang et al. 2007). Secondly, there is strong allelopathy effect from *S. canadensis* roots and rhizomes, which can inhibit the germination and seedling growth of native plant species (Zhang et al. 2014). Thirdly, soil microbes, such as arbuscular mycorrhizal fungi, could enhance the invasion of *S. canadensis* into new habitats, with a cumulative effect of AM fungi on the species over time (Jin et al. 2004). As an abiotic factor, light also could affect the competitive ability, growth, or stress-resistant ability of the invasive plant (Björn et al. 2016). However, there are still several unanswered questions about the light adaptation ability of *S. canadensis*.

Molecular tools and transcriptome sequences are cost-effective methods to support research on environmental responses in plants (Müller et al. 2012), and the RNA-seq method provides a direct and efficient way to compare the complex mechanism of photosynthesis (Xing et al. 2016). Thomas et al. (2016) demonstrated the generation and assembly of RNA sequences to explain the responses to environmental change in green ash (*Fraxinus pennsylvanica*), and emerald ash borer attack, which is a primary threaten to green ash trees. Yu et al. (2016) provided a solid foundation for identifying taproot thickening-related critical genes (*RsSPS1* and *RsSuSy1*) based on root transcriptome sequencing in *Raphanus sativus*. Furthermore, in order to explain the invasive mechanism of *Ipomoea cairica*, an invasive plant in southern China, Björn et al. (2016) sequenced and assembled the *de novo* transcriptomes from *I. cairica* and two related species (*I. digitata* and *I. nil*), and showed how some genes related to enzyme production (*pal*, *4cl*, *cad*, *chs*, and *chi*) in secondary metabolism played important roles in stress-resistant, growth and allelopathy processes.

Based on the above, this study was carried out to perform transcriptome sequencing and *de novo* assembly on two species, the invasive plant *S. canadensis* and the co-occurring native plant *Solidago decurrens* in order to: (1) compare the transcriptomes relating to photosynthesis to examine their role in invasion ability; (2) provide information for further functional and comparative genomic research about the invasive plant.

**Materials and Methods**

**Sample preparation**

The plants of *S. decurrens* and *S. canadensis* were collected from the Purple Mountain in Nanjing, Jiangsu Province, China on May 20th, 2016 (E118o48’00’’-118o53’04’’, N32o01’57’’-32o06’15). Equal sized pieces of rhizome (5 cm long) from the samples were cultivated in natural conditions (day length 13h and night length 11h, temperature 16-25°C) in Zealquest Scientific Technology Co., Ltd., Shanghai, China. One rhizome of *S. canadensis* or *S. decurrens* was planted in each of 6 plastic pots (25 cm diameter) containing 3 kg of loamy soil (dry weight). The loamy soil originated from the farmland of *S. decurrens* and *S. canadensis* natural habitats. The nutrient content of the substrate was 15.6-21.8 μg g−1available P, 68.2-75.1μg g−1 NO3-N, 3.5-4.2% organic matter, 183-227μg g−1 potassium, and a pH of 7.5. The plants were grown for three months prior to this study. Each species were performed with three replicates (A1-A3, for *S. canadensis*, B1-B3, for *S. decurrens*).

**Photosynthetic analysis**

The photosynthetic parameters of the two species were tested for prior to destructive sample collection. Chlorophyll ﬂuorescence measurements (*F*v/*F*m) were obtained by recording ﬂuorescence induction curves with a Dual-PAM-100 measuring system (Walz GmbH, Effeltrich, Germany). Before determination, the leaves were kept in clip curettes for dark adaptation for 30 min. The fluorescence parameters *F*0 (initial fluorescence level) and *F*m (maximum fluorescence level) were measured and the ratio *F*v/*F*m (where *F*v=*F*m-*F*0) was determined according to Rosenqvist & van Kooten’s method (2003). Rapid light curve (α value) was also measured with Dual-PAM-100.

To characterize light-induced shifts in carbon acquisition, instantaneous gas exchanges were measured on fully expanded leaves, the photosynthetic rates were determined using the Portable Gas Exchange Fluorescence System GFS-3000 (Walz GmbH, Effeltrich, Germany). Two leaves of each replicate from the same site of both *S. canadensis* and *S. decurrens* were selected for sampling. The mean value of two leaves was used as the replicate for statistical analysis. All the three replicates of each plant were determined. All of the measurements were completed from 08:00 to 15:30 in August 27th, 2016.

After testing and collection, each plant was washed and collected, including all the shoots and roots, and then oven dried for the measurement of shoots biomass and roots biomass (65°C for 48 h).

**RNA extraction**

RNA from the leaves of both *S. canadensis* and *S. decurrens* was collected*.* Total RNA was extracted from the tissues using TRIzol® Reagent according to the manufacturer’s instructions (Invitrogen, Shanghai, China) and genomic DNA was removed using DNase I (Takara, Dalian, China). Then the RNA quality was determined by Agilent 2100 Bioanalyser and quantified using the ND-2000 (NanoDrop Technologies, Thermo, USA). Only the high-quality RNA samples (OD260/280=1.8~2.2, OD260/230≥2.0, RIN≥6.5, 28S:18S≥1.0, >10μg) were used to construct the sequencing library.

**Library preparation and Illumina Hiseq 4000 sequencing**

Firstly, the RNA-seq transcriptome library was prepared following TruSeqTM RNA sample preparation Kit from Illumina (San Diego, CA, USA). Messenger RNA was isolated according to polyA selection method by oligo (dT) beads and then fragmented by fragmentation buffer. Secondly, double-stranded cDNA was synthesized using a SuperScript double-stranded cDNA synthesis kit (Invitrogen, CA, USA) with random hexamer primers (Illumina). Then the synthesized cDNA was subjected to end-repair, phosphorylation and ‘A’ base addition according to Illumina’s library construction protocol. Libraries were size selected for cDNA target fragments of 200-300 bp on 2% Low Range Ultra Agarose (LRUA) followed by PCR amplified using Phusion DNA polymerase (NEB, Beijing, China) for 15 PCR cycles. After quantified by TBS380 (YPH-Bio, Beijing, China), paired-end RNA-seq sequencing library was sequenced with the Illumina HiSeq 4000 (2×150bp read length).

***De novo* assembly and annotation**

The raw paired end reads were trimmed and quality controlled by SeqPrep (https://github.com/jstjohn/SeqPrep) and Sickle (https://github.com/najoshi/sickle) with default parameters. Next, clean data from the samples were used to do *de novo* assembly with Trinity (http://trinityrnaseq.sourceforge.net/) (Grabherr et al. 2011). All the assembled transcripts were searched against the NCBI protein nonredundant (NR), String, and KEGG databases using BLASTX to identify the proteins that had the highest sequence similarity with the given transcripts to retrieve their function annotations and a typical cut-off *e*-value less than 1.0×10−5 was set. BLAST2GO (http://www.blast2go.com/b2ghome) program was used to get GO annotations of unique assembled transcripts for describing biological processes, molecular functions and cellular components (Conesa et al. 2005). Metabolic pathway analysis was performed using the Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.genome. jp/kegg/) (Goto & Kanehisa 2000).

**Detection of homologous genes**

OrthoMCL v2.0.2 was used to identify gene family clusters in *S. canadensis* and *S. decurrens* (Li et al. 2003). Pairwise sequence similarities between all input protein sequences were calculated by all-by-all BLASTp with an *e*-value cut-off of 1*e*-05 and a minimum match length of 50%. To define ortholog cluster structure, a Markov clustering algorithm was applied with an inflation value (-I) of 1.8 (default value in OrthoMCL) (Kim et al. 2017). Putative splice variants were removed from the data set, the longest protein sequences were kept and subsequently filtered for premature stop codons and incompatible sequences.

**Differential expression analysis and functional enrichment**

To identify DEGs (differential expression genes) between two different samples, the expression level of each transcript was calculated according to the fragments per kilobase of exon per million mapped reads (FRKM) method. RSEM (http://deweylab. biostat.wisc.edu/rsem/) was used to quantify gene abundances (Li & Dewey 2011). The R statistical package edgeR (Empirical analysis of Digital Gene Expression in R, http://www.bioconductor.org/packages/2.12/bioc/html/edgeR.html) was utilized for differential expression analysis (Robinson et al. 2010). In addition, functional -enrichment analysis including GO and KEGG was performed to identify which DEGs were significantly enriched in GO terms and metabolic pathways at Bonferroni-corrected *P*-value ≤0.05 compared with the whole transcriptome background. GO functional enrichment and KEGG pathway analysis were carried out by Goatools (https://github.com/tanghaibao/Goatools) and KOBAS (http://kobas.cbi. pku.edu.cn/home.do) (Xie et al. 2011).

**Quantitative real-time PCR analysis**

Gene specific RT-qPCR primers were described by Afonina et al. (2007), which are shown in Table S1. The cDNA was synthesized from the two species with total RNA in a 10μL reaction system using PrimeScript II RTase (TaKaRa Biotechnology, Dalian, China). RT-qPCR was performed in ABI Step-One Real-Time thermal cycler (USA) using power SYBR Green qPCR master mix (ABI, UK) following manufacturer's instructions. PCR ampliﬁcation was carried out with initial denaturation of 94°C for 10min, 40 cycles each of 30s at 95°C, 30s at respective annealing temperatures and extension at 72°C for 30s. All reactions were performed in three replicates. The fold change in expression was calculated and transformed to log2 scale.

Statistical analysis

All the datasets were checked for normality prior to further analysis, and none of them were significantly different from a normal distribution, so no data transformation was required. Means of graphical representation showing chlorophyll ﬂuorescence, ﬂuorescence induction curves, rapid light curve, biomass, the transcriptome and RT-qPCR expression profiles of 14 selected photosynthetic genes were calculated according to the data using SPSS (16.0). Student’s t-test was used to compare the means of graphical representation showing chlorophyll ﬂuorescence, ﬂuorescence induction curves, rapid light curve, biomass, the transcriptome and RT-qPCR expression profiles of 14 selected photosynthetic genes. The results were expressed as mean ± SE, and the statistical significance was accepted when *P*<0.05.

**Results**

**Photosynthetic and growth data**

Based on morphological observations, the invasive species, *S. canadensis*, showed a potential clear competitive advantage over its congener, *S. decurrens* (Fig. 1A, B). In addition, the photosynthetic parameters of the invasive species were significantly higher than those of the native species(*P*<0.05), such as chlorophyll ﬂuorescence (*F*v/*F*m), ﬂuorescence induction curves (α value) and rapid light curve (Fig. 1C-E). Results of the biomass were the same to photosynthetic data (Fig. 1F).

***De novo* sequencing and assembly**

After filtering out the low quality reads, a total of 88,751,170, 63,647,626, 84,054,340 (A1-A3, for *S. canadensis*) and 57,141,142, 58,065,410, 50,215,998 (B1-B3, for *S. decurrens*) clean reads were obtained. Based on trinity assembly, 177,737/176,128 (*S. canadensis*/*S. decurrens*) contigs with an average length of 739.93/812.53 bp were obtained. The N50 values were 1221/1198 (Table 1). 128,038/127,820 unigenes were generated by extracting the longest transcripts, and the GC contents for *S. canadensis* and *S. decurrens* unigenes were 41.91% and 44.38%, respectively.

**Transcript annotation**

Unigenes of both species were annotated by BLASTx alignment against NR, Swissprot, Pfam, and KEGG databases with an *e*-value threshold of 1*e*-5 (Table 2). In total, 55,435 (43.3%) and 56,762 (44.4%) unigenes returned values when searched against the NR nucleotide database, 37,302 (29.1%) and 37,861 (29.6%) mapped to KEGG database. Among unigenes that had a BLASTx-hit, the top three best hits for the two species were *Vitis vinifera* (6847, 5.34% for *S. canadensis*, 7208, 5.64% for *S. decurrens*) (unigenes number, the percentage of the hit unigenes), *Coffea canephora* (4988, 3.90% for *S. canadensis*, 5287, 4.14% for *S. decurrens*), and *Theobroma cacao* (3380, 2.65% for *S. canadensis*, 3351, 2.62% for *S. decurrens*).

All-unigenes were further classified using GO assignments after functional annotation. Based on homologous genes, all-unigenes (128,038/127,820) were categorized into 47/56 GO terms under three domains: biological processes, cellular components, and molecular function. From the 128,038/127,820 unigenes, 46,397 (36.2%)/46,714 (36.6%) were assigned a GO term at least once (Fig. S1). A high degree of consistency was found in GO analysis between *S. canadensis* and *S. decurrens*. For the category of biological process, the three most represented groups were metabolic processes (31,649, 25.0%/31,976, 25.0%), cellular processes (27,081, 21.2%/28,956, 22.7%), and single-organism processes (22,462, 17.5%/24,628, 19.3%). For the category of cellular components, the three most represented groups were cell (16,711, 13.1%/19,134, 15.0%), cell parts (16,711, 13.1%/19,133, 15.0%), and organelles (9681, 7.56%/11,413, 8.93%). Lastly, for the category of molecular function, the three most represented groups were catalytic activity (27,267, 21.3%/26,947, 21.1%), binding (24,035, 18.8%/25,514, 20.0%), and transporter activity (3452, 2.70%/3995, 3.13%).

There were 12,951 unigenes (10.1%)/14,594 unigenes (11.4%) were annotated into 25 categories in *S. canadensis* and *S. decurrens*, respectively (Fig. S2). The largest annotation category for both were assigned to the cluster of “general function prediction” (1590, 1.24%/1483, 1.16%), followed by “signal transduction mechanisms” (1522, 1.19%/1382, 1.08%), “posttranslational modification, protein turnover, chaperones” (1510, 1.06%/1302, 1.02%).

To better establish links and assign a high-level function of genes in the genome, KEGG analyses were performed. 49,019 (38.3%) and 51,100 (40.0%) transcripts were successfully mapped to 385 and 390 metabolic pathways, respectively, broadly classified under five categories (A, Metabolism; B, Genetic Information Processing; C, Environmental Information Processing; D, Cellular Processes; E, Organismal Systems) (Fig. S3 for *S. canadensis*, Fig. S4 for*S. decurrens*).

**Pathway enrichment analysis of DEGs in** **photosynthesis of *S. canadensis* and *S. decurrens***

Based on the maximum number of differentially expressed transcripts, 1089 major genes (286 metabolic pathways) were found to be potentially involved in five major categories: (A) Metabolism, (B) Genetic Information Processing, (C) Environmental Information Processing, (D) Cellular Processes, (E) Organismal Systems. To categorize the biological functions of the DEGs of photosynthesis, a KEGG pathway enrichment analysis were performed. Twelve genes were found to be involved in photosynthesis (ko00195) from energy metabolism of metabolism (Fig. 2A), and eight genes were found to be involved in photosynthesis-antenna proteins (ko00196) from energy metabolism of metabolism (Fig. 2B). Our results showed that photosystem I genes (*PsaE*, *PsaG*, *PsaK*), photosystem II genes (*PsbP*, *PsbQ*, *PsbW*, *PsbY*, *Psb27*), photosynthethic electron transport genes (*PetE*, *PetF*, *PetH*) and F-ATPase delta to be up-regulated in photosynthesis (ko00195) of *S. canadensis* (Fig. 3). Genes of light-harvesting chorophyll protein complex (LHC) family (*Lhca1*, *Lhca2*, *Lhca4*, *Lhcb1*, *Lhcb3*, *Lhcb4*, *Lhcb6*, *Lhcb7*) were also found to be up-regulated in photosynthesis-antenna proteins (ko00196) of *S. canadensis* (Fig. 3).

**Validation of RNA-Seq Expression Data by RT-qPCR**

To verify the reliability of RNA-Seq differential expression analyses, quantitative realtime PCR was utilized to explore 20 selected genes between *S. canadensis* and *S. decurrens* in photosynthesis (Fig. 4). Finally, 14 selected genes were successfully validated by RT-qPCR. The 14 genes which belonged to photosynthesis (ko00195) and photosynthesis-antenna proteins (ko00196) were found to be up-regulated in *S. canadensis*. Furthermore, the majority of the genes expressed were similar to those found with RNA-Seq analysis.

**Discussion**

Photosystem II (PSII) catalyzes the initial step of photosynthesis, the light-dependent oxidation of water that yields molecular oxygen and reduced plastoquinone (Hou et al. 2015). It is the essential process in photosynthesis. The photosynthetic data (ratio *F*v/*F*m) is proportional to the potential maximal quantum yield of PSII, and the significant difference between *S. canadensis* and *S. decurrens* (Fig. 1C) might be attributed to the efficiency of PSII that is generally related to photoinhibition (Krause & Somersalo 1989) or alteration in the levels of chlorophyll biosynthetic pathways (Song et al. 2014). Combined with the results of RNASeq (Fig. 2 and Fig. 3) and RT-PCR (Fig. 4), the up-regulated genes are associated with PSII, which would explain the advantage of hypersensitivity to light in *S. canadensis*. The three genes, *PsbP*, *PsbQ* and *Psb27*, serve to optimize oxygen evolution and efficiency of PSII (Schmidt et al. 2016), and play an essential role in enabling plants to adapt to fluctuating light intensity (Ifuku et al. 2005; Hou et al. 2015). *PsbW* is located close to the minor antennae of the PSII complex, which would associate with the PSII reaction center and the LHC complex (Plöchinger et al. 2016). The protein, PsbY, has played a supporting role in stabilizing the binding of PsbE and PsbF (Plöchinger et al. 2016).

Photosystem I (PSI) is a multiprotein complex consisting of a number of subunits, which catalyzes the electron transfer from plastocyanin or cytochrome *c*6 to NADP+ through ferredoxin and ferredoxin: NADP+ oxidoreductase (Ozawa et al. 2010). In higher plants, PSI and peripheral light-harvesting complex I (LHCI, *Lhca1*-*Lhca5*) together form the PSI-LHCI supercomplex and drive oxygenic photosynthesis. The key genes, *PsaG* and *PsaK*, may be integrated after the LHCI is assembled into the PSI core complex (Ozawa et al. 2010). In addition, LHCI and PSI are significantly correlated with F-ATPase and C4 cycle (Xing et al. 2016). Therefore, the complex consisting of LHCI, PSI, and F-ATPase plays an important role in the coordinated expression between cyclic electron transport and the CO2-concentrating mechanism (Xing et al. 2016). In this study, several genes were found to be up-regulated in cyclic electron transport, such as *PsaG*, *PetE*, *PetF*, *Lhca1*, *Lhca2*, and *Lhca4* (Fig. 2), and the results of photosynthetic data (Fig. 1D, E) and qPCR (Fig. 4) are in good agreement with the transcription. Therefore, the competitive advantage of *S. canadensis* may be due to three factors: (i) the utilization of light for electron transport, especially in *PsaG*, *PetE*,*Lhca1*, *Lhca2*, and *Lhca4* (Brown et al. 2005; Ozawa et al. 2010; Xing et al. 2016); (ii) several key genes which associate with oxidative phosphorylation pathway which were observed to be up-regulated, such as F-ATPase (Allen, 2003); (iii) the adaptation of changing environment, eg. the overexpression of *PetF* could increase the efficiency of reactive oxygen species (ROS) scavenging in chloroplasts to confer heat tolerance (Lin et al. 2013).

**Acknowledgements**

This study was supported by Natural Science Foundation of Shanghai (18ZR1425400), National Natural Science Foundation of China (31800310) and the Research Funds for the Introduction of Talents of Shanghai Science and Technology Museum.

**References:**

Afonina I, Ankoudinova I, Mills A, Lokhov S, Huynh P, Mahoney W (2007) Primers with 5’ flaps improve real-time PCR. Biotechniques 43(6): 770-774

Allen JF. (2003) Cyclic, pseudocyclic and noncyclic photophosphorylation: new links in the chain. Trends Plant Sci 8:15-19

Björn LO, Geng Y, Chen LL, Lu H, Ning CJ, Li SS (2016) Metabolic characteristics of invasive plant *Ipomoea cairica* in South China by *de novo* transcriptomics. J Trop Subtrop Bot 24: 128-142

Brown NJ, Sullivan JA, Gray JC (2005) Light and plastid signals regulate the expression of the pea plastocyanin gene through a common region at the 5' end of the coding region. Plant J 43: 541-552

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21(18):3674-3676

Dong M, Lu JZ, Zhang WJ, Chen JK, Li B (2006) Canada goldenrod (*Solidago canadensis*): an invasive alien weed rapidly spreading in China. Acta Phytotaxon Sin 44: 72-85

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29: 644-652

Goto M, Kanehisa S (2000) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28: 27-30

Hou X, Fu A, Garcia VJ, Buchanan BB, Luan S (2015) PSB27: A thylakoid protein enabling *Arabidopsis* to adapt to changing light intensity. PNAS 112(5): 1613-1618

Hu TY, Fang F, Guo SL, Jiang HW (2007) Comparison of basic photosynthetic characteristics between exotic invasive weed *Solidago canadensis* and its companion species. J Zhejiang Univ (Agr Life Sci) 33 (4): 379-386

Ifuku K, Yamamoto Y, Ono T, Ishihara S, Sato F (2005) PsbP protein, but not PsbQ protein, isessential for the regulation and stabilization of photosystem II in higher plants. Plant Physiol 139: 1175-1184

Jin L, Gu YJ, Xiao M, Chen JK, Li B (2004) The history of *Solidago canadensis* invasion and the development of its mycorrhizal associations in newly-reclaimed land. Funct Plant Biol 31(10): 979-986

Kim YM, Kim S, Koo N, Shin AY, Yeom SI, Seo E, Park SJ, Kang WH, Kim MS, Park J, Jang I, Kim PG, Byeon I, Kim MS, Choi JH, Ko G, Hwang JH, Yang TJ, Choi SB, Lee JM, Lim KB, Lee J, Choi IY, Park BS, Kwon SY, Choi D, Kim RW (2017) Genome analysis of *Hibiscus syriacus* provides insights of polyploidization and indeterminate flowering in woody plants. DNA Res 24 (1): 71-80

Krause GH, Somersalo S (1989) Fluorescence as a tool in photosynthesis research: application in studies of photoinhibition, cold acclimation and freezing stress. Philos T R Soc B 323: 281-293

Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12: 323

Li L, Stoeckert CJ, Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13: 2178-2189

Li ZY, Xie Y (2002) Invasive alien species in China. China Forestry Publishing House, Beijing

Lin YH, Pan KY, Hung CH, Huang HE, Chen CL, Feng TY, Huang LF (2013) Overexpression of ferredoxin, PETF, enhances tolerance to heat stress in *Chlamydomonas reinhardtii*. Int J Mol Sci 14: 20913-20929

Müller T, Ensminger I, Schmid K J (2012) A catalogue of putative unique transcripts from douglas-fir (*Pseudotsuga menziesii*) based on 454 transcriptome sequencing of genetically diverse, drought stressed seedlings. BMC Genomics 13(1): 673

Ozawa S, Onishi T, Takahashi Y (2010) Identification and characterization of an assembly intermediate subcomplex of photosystem I in the green alga *Chlamydomonas reinhardtii*. J Biol Chem 285: 20072-20079

Plöchinger M, Schwenkert S, von Sydow L, Schröder WP, Meurer J (2016) Functional update of the auxiliary proteins PsbW, PsbY, HCF136, PsbN, TerC and ALB3 in maintenance and assembly of PSII. Front Plant Sci 7: 423-436

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139-140

Rosenqvist E, van Kooten O (2003) Chlorophyll fluorescence: a general description and nomenclature. In: DeEll JR, Toivonen PMA (eds) Practical applications of chlorophyll fluorescence in plant biology. Kluwer, Dordrecht, The Netherlands, pp 31-77

Schmidt SB, Powikrowska M, Krogholm KS, Naumann-Busch B, Schjoerring JK, Husted S, Jensen PE, Pedas PR (2016) Photosystem II functionality in barley responds dynamically to changes in leaf manganese status. Front Plant Sci 7: 1772-1784

Song A, Li P, Fan F, Li Z, Liang Y (2014) The effect of Silicon on photosynthesis and expression of its relevant genes in rice (*Oryza sativa* L.) under high-zinc stress. PLoS One 9(11): e113782

Thomas L, Teodora B, Nicole Z, Jack D, Nathan H, Xu Y, Koch J, Liang HY, McGraw J, Schuster S, Shim D, Coggeshall MV, Carlson JE, Staton ME (2016) The green ash transcriptome and identification of genes responding to abiotic and biotic stresses. BMC Genomics 17(1): 702

Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li CY, Wei L (2011) KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Res 39: 316-322

Xing S, Kang L, Xu Q, Fan Y, Liu W, Zhu C, Song Z, Wang Q, Yan J, Li J, Sang T (2016) The coordination of gene expression within photosynthesis pathway for acclimation of C4 energy crop *Miscanthus lutarioriparius*. Front Plant Sci 7: 109-121

Xu HM, Li YJ, Lu YM (2006) Primary research on the biological characteristics of *Solidago canadensis*. Weed Sci (2): 32-33

Yang RY, Mei LX, Tang JJ, Chen X (2007) Allelopathic effects of invasive *Solidago canadensis* L. on germination and growth of native Chinese plant species. Allelopathy J 19(1): 241-248

Yu R, Liang X, Wei Z, Yan W, Luo X, Wang R, Zhu AW, Xie Y, Karanja B, Liu LW (2016) *De novo* taproot transcriptome sequencing and analysis of major genes involved in sucrose metabolism in radish (*Raphanus sativus* L.). Front Plant Sci 7: 1-12

Zhang DY, Jin Y, Lu B, Li J, Tian JJ, Dong LY (2014) Allelopathic effect and mechanism of extract solution from *Solidago canadensis* L. on wheat (*Triticum aestivum*). J Nanjing Agr Univ 37(5): 87-92

Zhao SY, Sun SG, Dai C, Gituru RW, Chen JM, Wang QF (2015) Genetic variation and structure in native and invasive *Solidago canadensis* populations. Weed Res 55(2): 163-172

Table 1 Transcriptome for sequencing and *de novo* assembling in *S. canadensis* and *S. decurrens*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *S. canadensis* | | *S. decurrens* | |
|  | Unigenes | Transcripts | Unigenes | Transcripts |
| Total sequence num | 128,038 | 177,737 | 127,820 | 176,128 |
| Total sequence base | 94,738,995 | 14,994,534 | 91,621,373 | 143,110,141 |
| Percent GC | 41.91 | 40.79 | 44.38 | 42.95 |
| Largest | 61,116 | 61,116 | 61,360 | 61,360 |
| Smallest | 201 | 201 | 201 | 201 |
| Average  N50 | 739.93  1107 | 843.64  1221 | 716.8  1097 | 812.53  1198 |

Table 2 The annotation list of transcripts in *S. canadensis* and *S. decurrens*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Pfam | Swissprot | KEGG | NR |
| *S. canadensis* | 4,8884 | 61,434 | 49,019 | 105,278 |
| Proportion(%) | 38.2 | 48.0 | 38.3 | 82.2 |
| *S. decurrens* | 51,687 | 51,100 | 63,349 | 107,569 |
| Proportion(%) | 40.4 | 40.0 | 49.6 | 84.2 |

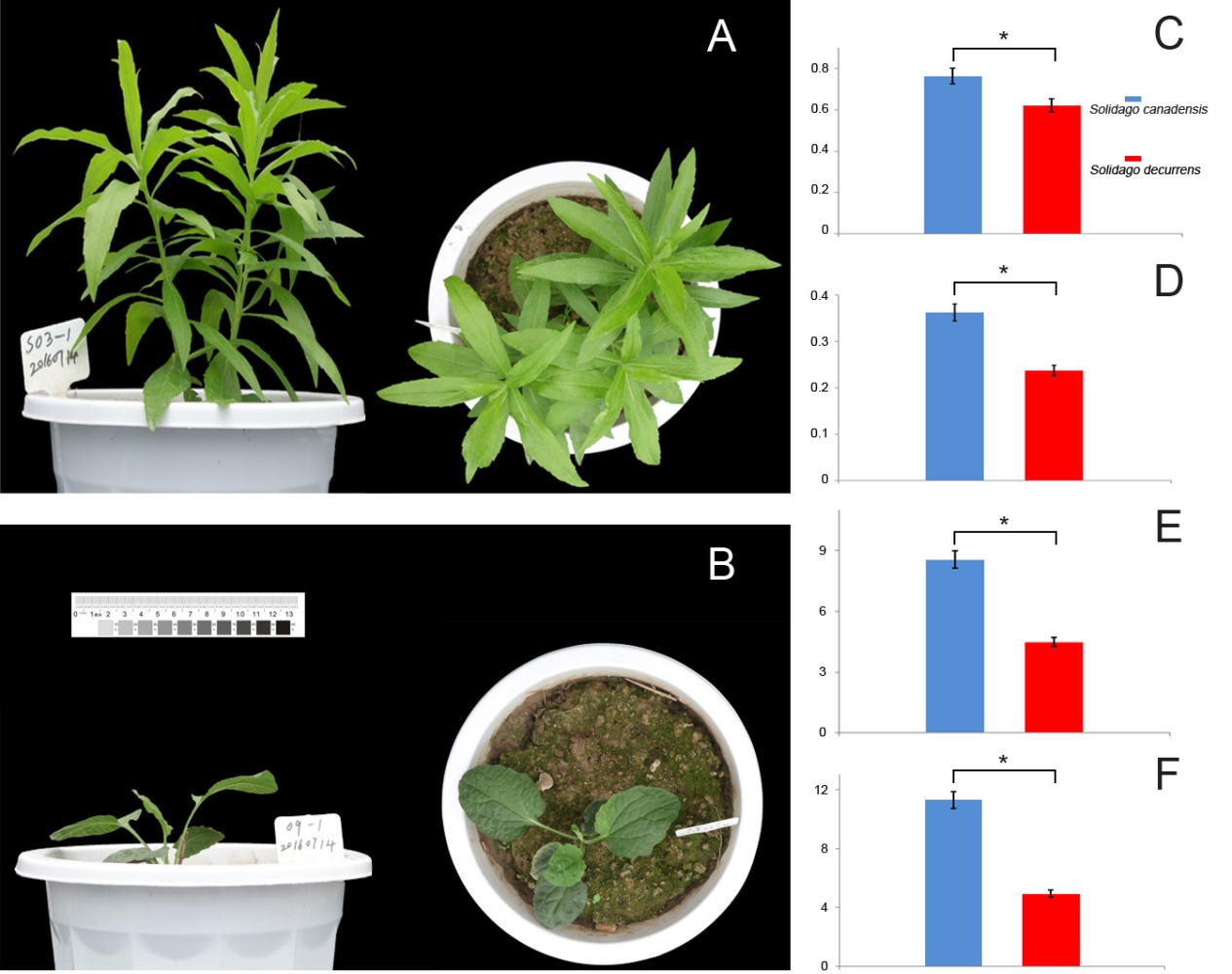


Fig. 1 Morphological observations and photosynthetic data

Note: A: *S. canadensis* for three months; B: *S. decurrens* for three months; C: graphical representation showing chlorophyll ﬂuorescence; D: ﬂuorescence induction curves; E: rapid light curve; F: biomass. \*: *P*<0.05, significant differences were determined relative to each species using a Student’s *t*-test, error bar represents the SE for replicated experiments.

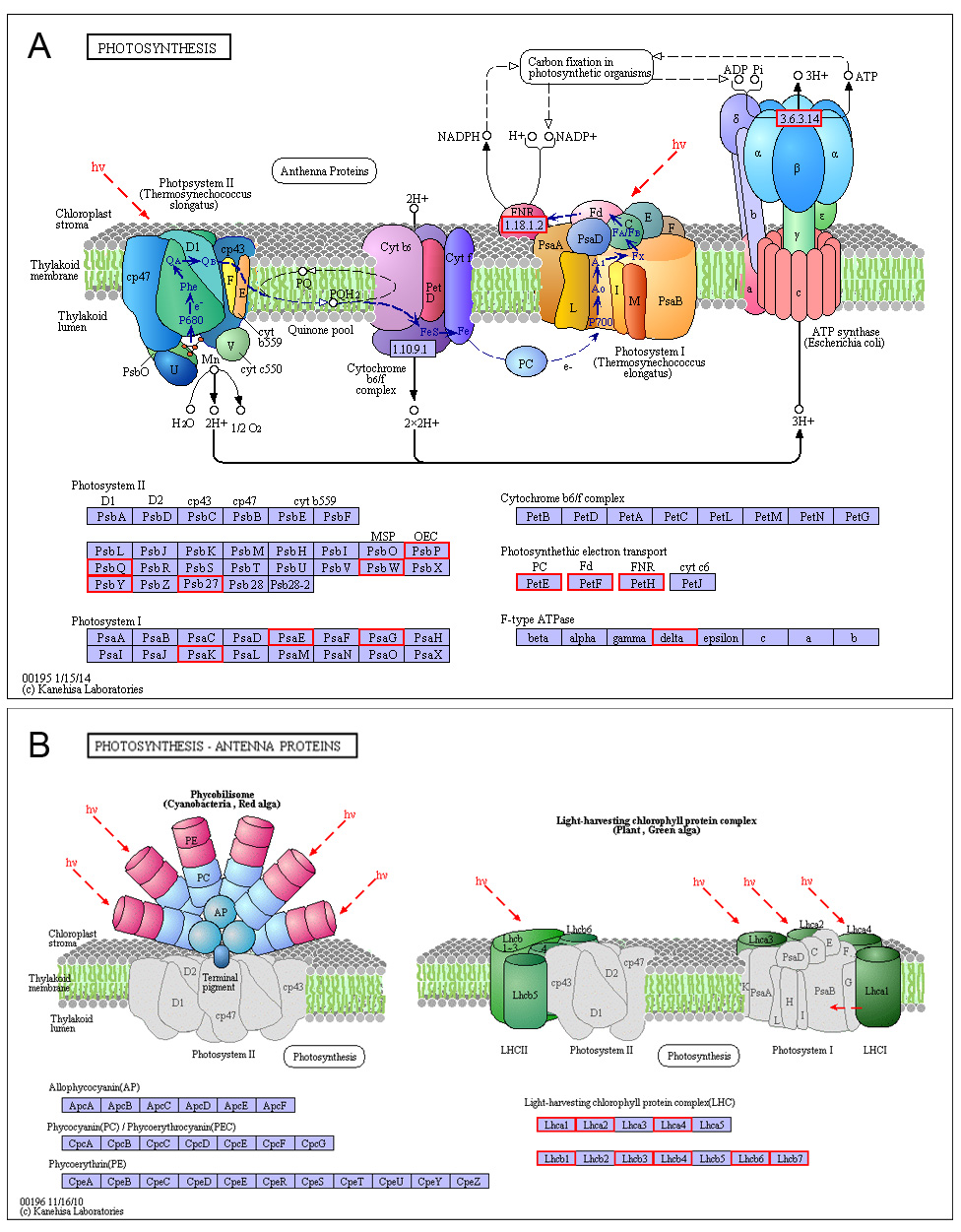


Fig. 2 Unigenes involved in photosynthesis pathways (A) and photosynthesis-antenna proteins pathway (B). The red boxes indicated up-regulated genes.

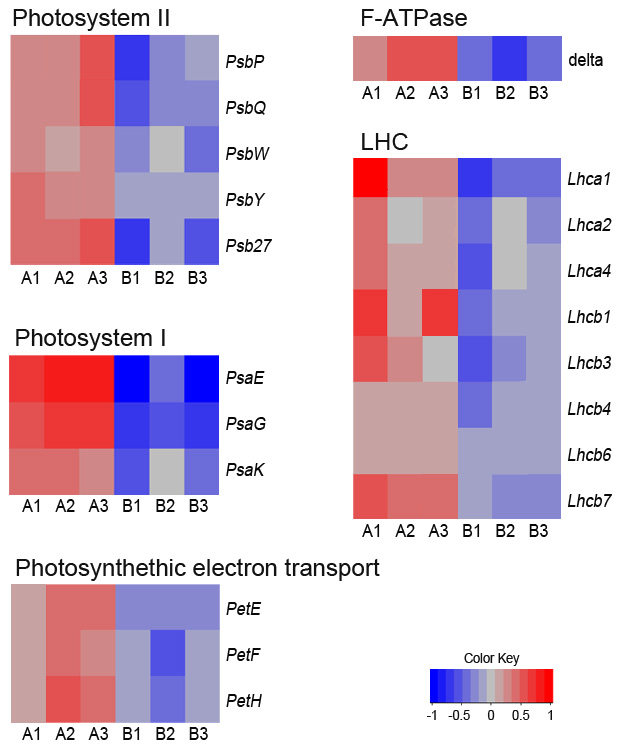


Fig. 3 Heatmap showing transcript abundance of photosynthetic genes in *S. canadensis* (A1-A3) and *S. decurrens* (B1-B3). Blue indicates down-regulated expression and red represents up-regulated expression.

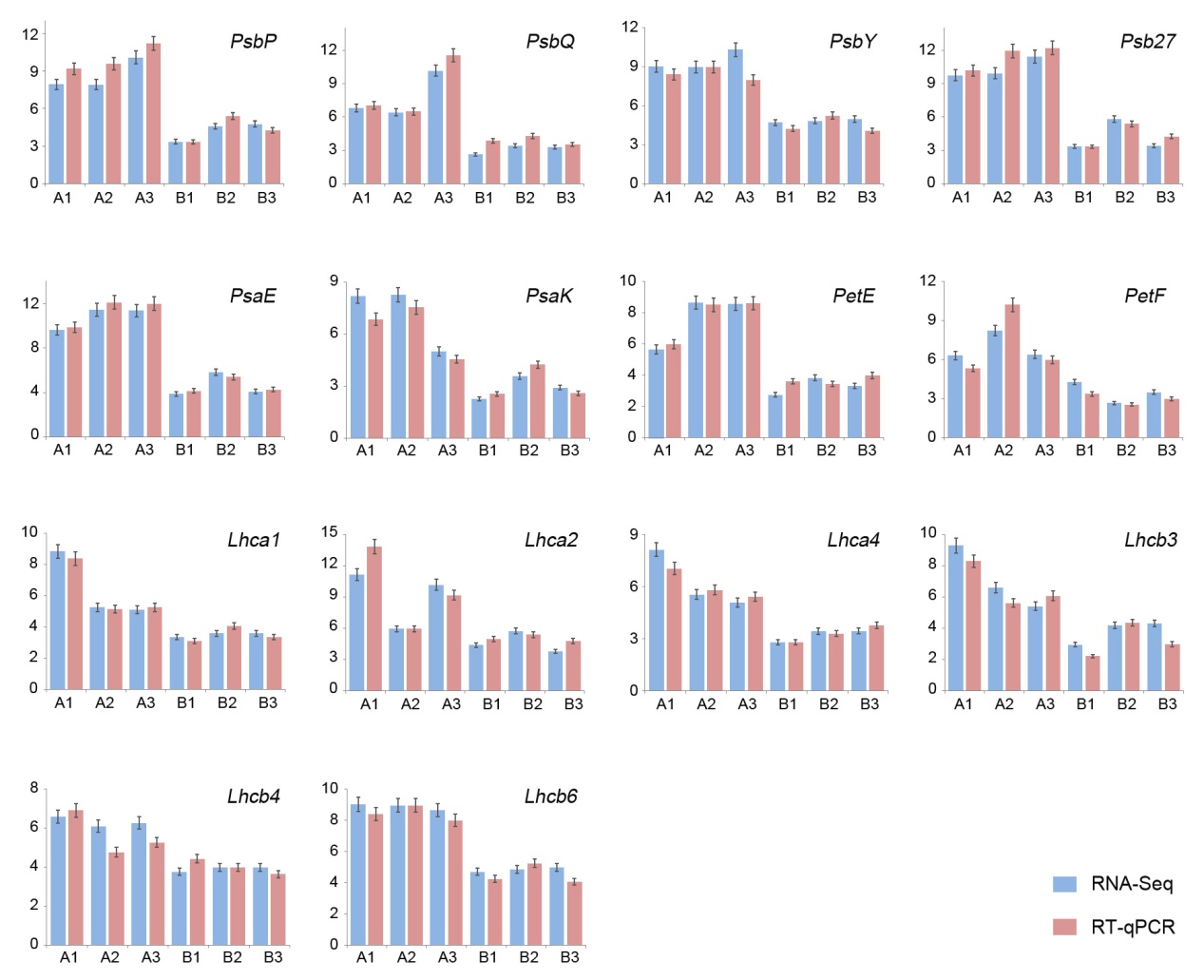


Fig. 4 Comparison between transcriptome and RT-qPCR expression profiles of 14 selected photosynthetic genes (A1-A3, for *S. canadensis*; B1-B3, for *S. decurrens*). Error bar represents the SE for replicated experiments.

Fig. S1 Frequencies of transcripts matching GO terms.

Fig. S2 Histogram presentation of clusters of orthologous groups (COG) classification.

Fig. S3 Pathway assignment of *Solidago canadensis* based on KEGG. (A) metabolism; (B) genetic information processing; (C) environmental information processing; (D) cellular processes; (E) environmental adaptation.

Fig. S4 Pathway assignment of *Solidago decurrens* based on KEGG. (A) metabolism; (B) genetic information processing; (C) environmental information processing; (D) cellular processes; (E) environmental adaptation.

Table S1 Primers used to perform qPCR of related genes.