De novo Short Read Assembly and Functional Annotation of Eleocharis vivipara, a C₃/C₄ Interconvertible Sedge Plant

Daijiro Harada¹, Katsuyuki T. Yamato², Katsura Izui² and Motomu Akita¹,²

¹Department of Biotechnological Science, Kindai University, Kinokawa, Wakayama 649-6493, Japan
²Institute of Advanced Technology, Kindai University, Kainan, Wakayama 642-0017, Japan

(Received November 22, 2017; Accepted January 10, 2018)

Eleocharis vivipara is an amphibious sedge that displays C₄ traits under terrestrial environments and C₃ traits in submerged environments. This plant is thus potentially advantageous for screening genes indispensable to the development of C₄ photosynthesis. In this study, we performed de novo transcriptome analysis of E. vivipara using its terrestrial- and submerged-type plants. By next-generation sequencing (NGS), approximately 90 and 89 million reads were yielded for the terrestrial and submerged types, respectively, and were assembled into 27,249 unigenes. Of these de novo consensus sequences, 94.5% showed similarities to database-registered sequences, and 69.4% were assigned with Gene Ontology terms. Our de novo assembled sequence data should provide a foundation for genetic analysis of the C₄ photosynthetic system.

Keywords: C₄ photosynthesis genes, Eleocharis vivipara, short read assembly, transcriptome, unigenes

INTRODUCTION

More than 90% of terrestrial plant species possess the C₃ photosynthetic system, while only about 3% undergo C₄ photosynthesis. Nevertheless, C₃ plants are responsible for 25% of the annual total terrestrial plant biomass production (Langdale, 2011; Sage and Zhu, 2011). Compared with C₄ plants, C₃ plants generally exhibit higher CO₂ assimilation abilities (about 1.5-2-fold higher in individual blades), higher yields per growing area (approximately 2-3-fold) and higher water and nitrogen use efficiencies (Black, 1979).

The productivity of major C₃ crops, such as rice, wheat and soybean, could be greatly increased if the CO₂ concentrating mechanism of C₃ photosynthesis could be introduced into these C₃ plants through genetic engineering. Such attempts have been made by various investigators (Häusler et al., 2002; Miyao et al., 2011), but no promising results have been obtained thus far. For example, several genes encoding C₄ photosynthesis-related enzymes were successfully overexpressed individually or in combination in mesophyll cells of rice plants, but the transgenic plants did not perform C₄-like photosynthesis and their growth rate was not accelerated appreciably (Taniguchi et al., 2008). These observations imply the introduction of genes for CO₂-concentrating proteins is not sufficient to confer the C₄ photosynthetic framework to C₃ plants.

In fact, two functionally differentiated cell types—mesophyll cells and bundle-sheath cells—and their structural arrangement called Kranz anatomy are required to concentrate CO₂ and deliver it to ribulose-1,5-bisphosphate carboxylase/oxygenase. This CO₂ concentration system has been thought to be essential for C₄ photosynthesis. To convert C₃ plants into C₄ ones, all genes that are related to cellular and functional C₄ differentiation, including those associated with C₄ metabolism, transport of C₄-related compounds and development of Kranz anatomy, must be identified and introduced into C₃ plants (Covshoff and Hibberd, 2012). To identify the genes required for C₄ photosynthesis, a straightforward approach was comparative analysis of C₃ and C₄ plants. However, identification of the key C₃ photosynthetic genes is difficult because many of them are species-specific even among C₄ model plants (e.g., Zea mays and Sorghum bicolor). Although the genera Flaveria and Cleome include both C₃ and C₄ species and have attracted much attention as alternative systems (Brown et al., 2005; Külahoglu et al., 2014), there are still species-specific genes, which makes it difficult to identify the core functional set of C₃ photosynthetic genes (Gowik et al., 2011).

Eleocharis vivipara (Cyperaceae), first investigated by Ueno et al. (1988) is an amphibious leafless sedge. This plant develops Kranz anatomy and shows C₄ biochemical traits under terrestrial conditions, and performs NAD-dependent malic enzyme (NAD-ME)-type C₄ photosynthesis. Interestingly, under submerged conditions, it grows without Kranz anatomy and exhibits C₃ biochemical traits. When the submerged plants are exposed to air, they develop new shoots with C₄ traits within about a week. This species is thus suitable for screening genes indispensable to the biochemical, cellular and structural transition from the C₃ to the C₄ photosynthetic system; however, its relevant genomic and transcriptomic information is currently lim-
ied. Although C4 photosynthesis-related enzymes such as phosphoenolpyruvate carboxylase (PEPC) and pyruvate phosphate dikinase (PPDK) have been identified and characterized in *E. vivipara* (Agarie et al., 1997), the remaining genes involved in C4 photosynthesis are yet to be identified. Chen et al. (2014) reported the transcriptome analysis of *E. baldwinii*, but this congener of *E. vivipara* displays interconversion between C3–C4 intermediate and C4 traits, which makes *E. vivipara* more suitable for identification of genes indispensable to C4 development, as a distinct interconversion is observed between the C4 and C3 traits in this species (Ueno, 2001).

In this study, we obtained transcriptome data of the terrestrial and submerged types of *E. vivipara* using next-generation sequencing (NGS) and performed *de novo* assembly of the short-read sequences to provide a transcriptomic profile of this unique plant. To our knowledge, this is the first report of the *de novo* transcriptome analysis of *E. vivipara*. We expect that the catalogue of *E. vivipara* transcripts we generated should serve further analyses of the C4 genetic makeup.

**MATERIALS AND METHODS**

**Plant materials and total RNA isolation**

*E. vivipara* was kindly supplied by Prof. O. Ueno, Faculty of Agriculture, Kyushu University, Japan. Plants were maintained in a greenhouse controlled at 25°C with natural light. For preparing the plant materials, plants with about 15 cm in height were transplanted to 7.5 (i.d.) × 10 (h) cm pots containing regular soil for gardening and grown for 10 d in the same controlled greenhouse for rooting. Then each plant was moved and continued cultivation in a greenhouse with roof and window ventilators and natural light. For submersion, potted plants were set in a water filled acrylic tank (60 × 60 × 60 cm) and kept then under water. The tank was continuously supplied with tap water of which temperature was equilibrated with that of the greenhouse at the rate of approximately 1 L/min. Each plant was cultivated for one month and harvested on a fine day in October, 2011. Entire culms including meristem tissues of greenhouse-grown plants were harvested during the day and pulverized in a Mixer Mill MM301 (Retsch GmbH, Haan, Germany). Total RNA was extracted using an Agilent Plant RNA Isolation mini kit (Agilent, Santa Clara, CA, USA).

**NGS sequencing and preprocessing of NGS short reads**

cDNA libraries were prepared using a SMART cDNA synthesis system (Takara Bio USA, Mountain View, CA, USA) and single-end sequenced on a HiSeq2000 instrument (Illumina, San Diego, CA, USA) by Hokkaido System Science (Sapporo, Japan). To remove low-quality reads and primer adaptor contaminants, the short reads were filtered using CASAVA ver. 1.8.1 (http://support.illumina.com/sequencing/sequencing_software/casava.html) supplied by Illumina and then preprocessed using the DDBJ Read Annotation Pipeline (Nagasaki et al., 2013) with default settings. The transcriptome datasets are available under accession numbers DRA003124 and DRA003123 for terrestrial-type (C4) and submerged-type (C3) reads, respectively, in the public databases.

**De novo short read assembly**

The short reads were assembled using four commonly used *de novo* short read assemblers, SOAPdenovo (Luo et al., 2012), Velvet (Zerbino and Birney, 2008), ABySS (Simpson et al., 2009) and Trinity (Grabherr et al., 2011), available at the DDBJ Read Annotation Pipeline. Since these assemblers rely on de Bruijn graphs, it is important to optimize the k-mer size. In addition to the default of each assembler, an optimal k-mer size was also estimated by KmerGenie (Chikhi and Medvedev, 2013) and used for the assemblers (Supplementary Figure S1) except Trinity, which has the fixed value of 25 for k-mer size.

**Descriptive annotation**

Trinity *de novo* contigs were assembled into unigenes using CD-HIT ver. 4.5.4 (Huang et al., 2010). Putative open reading frames (ORF) of at least 200 amino acid residues (option –m 200) were identified by TransDecoder. LongOrfs (https://transdecoder.github.io). To remove mis-assembled contigs, the reads from terrestrial and submerged plants were back-mapped to the unigenes of the TransDecoder group by Bowtie with default options (Langmead et al., 2009). The sequences of the filtered unigenes are available at the public databases under accession numbers IAAE01000001–IAAE01027249. Functional annotation was performed by BLASTx with an e-value cutoff of 1 × 10−5 against the Swiss-Prot database, InterProScan, and the MapMan annotation pipeline (Lohse et al., 2014). BLASTx and InterProScan results were used for functional annotation with Gene Ontology (GO) terms (Hu et al., 2008). The unigenes were classified by orthologous group using OrthoMCL (Li et al., 2003).

**RESULTS AND DISCUSSION**

**De novo transcriptome assembly**

From the terrestrial- and submerged-type plants, 90,200,673 (approximately 9.1 Gbp) and 89,210,579 (approximately 9.0 Gbp) reads were obtained, respectively (Table 1). Among the four assemblers used in this study, Trinity generated an assembly of the highest N50 value (1,220 bp) and with the fewest number of contigs (213,423) (Table 2), and thus its assembly was used for further analysis. To reduce the redundancy, the *de novo* contigs by

| Table 1 Summary of filtered Illumina HiSeq2000 reads obtained from *Eleocharis vivipara* cDNA libraries |
|--------------------------------------------------|
| cDNA libraries | Terrestrial | Submerged |
| Total No. | 90,200,673 | 89,210,579 |
| Total Length (M bases) | 9,110 | 9,010 |
| % of ≥ QS 30 | 94 | 94 |
| Mean QS | 37 | 37 |

1. After filtering with CASAVA.
2. Percentage of filtered bases that were high quality (Quality score [QS] ≥ 30).
3. QS: Quality score.
Trinity were clustered into 199,664 unigenes by CD-HIT, 31,846 of which were predicted to contain ORFs with more than 200 amino acid (aa) residues by TransDecoder. LongOrfs. Then the terrestrial and submerged NGS reads were back-mapped to the processed group of the unigenes using Bowtie to exclude assembly artifacts, resulting in 27,249 de novo unigenes (IAAE01000001-TypeOne0 IAAE01027249). The numbers of the de novo unigenes in different ranges of length were 669 for 600–699 bp, 740 for 700–799 bp, 989 for 800–899 bp, 1,311 for 900–999 bp and 23,540 for 1,000 bp or more (Supplementary Table S1).

Compared to the *E. baldwinii* transcriptome, where the number of the unigenes longer than 600 bp was 12,841 (Chen et al., 2014), our *E. vivipara* unigene set represents longer transcripts (Fig. 1).

### Annotation
The unigene sequences were BLAST analyzed against the UniRef100, Swiss-Prot, NR and nucleotide sequence (NT) databases and 25,921 (95.1%), 22,187 (81.4%), 25,917 (95.1%) and 26,324 (96.6%), respectively, and 26,403 (96.9%) as a whole of the unigenes showed significant similarities to known sequences (Supplementary Table S2). The remaining 3.1% of the unigenes may account for species-specific genes or non-coding RNAs. InterProScan detected conserved domains in 26,514 (97.3%) of the unigenes (Supplementary Table S3).

### Clusters of Orthologous Group (COG) database
The unigene sequences were then assigned to COG. Among 14,601 (53.6%) unigenes with COG annotation (Fig. 2 and Supplementary Table S4), 3,562 (24.4%) and 614 (4.2%) were classified to poorly characterized groups, “general function prediction only” and “function unknown”, respectively. The others were classified to “signal transduction mechanisms” (1,535, 10.5%), “carbohydrate transport and metabolism” (1,140, 7.8%), “translation, ribosomal structure and biogenesis” (1,010, 6.9%), “posttranslational modification, protein turnover, chaperones” (959, 6.6%), and other 18 functions. No unigene was assigned to the groups, “extracellular structures” and “nuclear structure”.

### MapMan and Gene Ontology (GO) analysis
26,403 annotated unigenes were then assigned to MapMan bins (Supplementary Table S5) and accordingly to GO plant slim terms (Supplementary Table S6). In total, 16,276 (60.0%) of the unigenes were assigned to 35 MapMan bins (Fig. 3A and Supplementary Table S6), and 22,683 (83.2%) were assigned to GO plant slim terms. Of the 22,683 GO-assigned unigenes, 17,060 were assigned to GO terms associated with “biological process”, 19,900 to “molecular function” and 15,211 to “cellular component” (Fig. 3B–D and Supplementary Tables S7–S9).

### Comparison to known Eleocharis sequences
Among the 785 amino acid sequences from *E. vivipara* and *E. baldwinii* registered in NR, 572 showed similarity (E<10−5) to 130 of our *E. vivipara* unigenes by BLASTX.

Two of the *E. vivipara* unigenes, IAAE01002648 and IAAE01002648, showed putative amino acid sequence identities of 98–100% to the *Eleocharis* PEPC and PPDK.
sequences which have been registered in the public databases, respectively (Table 3). Three *E. vivipara* full-length sequences are predicted to code for PEPC (BAV81364.1, BAC19851.1, and BAV81374.1) and two for PPDK (BAA21654.1 and BAA21653.1). Among them, the numbers of amino acid residues corresponding to IAAE01002648, predicted PEPC, were the same as BAV81364.1, BAC19851.1 and BAV81374.1 (968 aa), and, especially, the sequence was completely identical to BAV81364.1. However, the PPDK sequence of IAAE01007373, 904 aa, were different from BAA21654.1 and BAA21653.1, 884 and 947 amino acids, respectively. Without genomic sequences, it is generally difficult to reconstruct full-length gene transcripts from short NGS reads, however the *E. vivipara* PPDK, 884/947 amino acids, is expected to be sufficiently longer than that from *E. baldwinii*, 209–400 amino acids (Chen et al., 2014). To the best of our present knowledge, our transcriptome is the most comprehensive data set available for the genus *Eleocharis* (see also Fig. 1).

Fig. 2 COG classification of the *E. vivipara* unigenes. The unigenes were classified into the functional categories of COG. The poorly characterized categories of “General function prediction only” and “Function unknown” are also included.

Fig. 3 Functional annotation of the transcriptome of *E. vivipara*. MapMan annotation (A). Level 2 GO term assignment and distribution in categories of Biological Process (B), Molecular Function (C), and Cellular Component (D).
**Identification of unigenes coding for C₄ enzymes**

All enzymes that are known to be characteristic of the NAD-ME type C₄ photosynthesis were identified in the annotated unigenes. The copy numbers of the characteristic C₄ enzyme genes in *E. vivipara*, *Z. mays* and *S. bicolor* were similar to those in rice (Table 4), which is consistent with the previous observation (Xu et al., 2013), suggesting that increase of the copy number of the C₄ related enzyme genes may not be a prerequisite for C₄ development.

**Expanded gene families in C₄ plants**

It has been suggested that gene duplication may be a key event for the C₄ evolution. For example, it was suggested that an increase in the number of paralogs may be associated with the C₄ tissue differentiation in *Z. mays* (Chang et al., 2013). We compared the numbers of predicted paralogs among *E. vivipara*, *Z. mays*, *S. bicolor*, and *O. sativa* by OrthoMCL. The 26,493 of the 27,249 unigenes of *E. vivipara* were classified into 8,928 orthogroups (Supplementary Table S10). Comparing with those of *O. sativa* (Sakai et al., 2013), *Z. mays* (Andorf et al., 2016) and *S. bicolor* (Monaco et al., 2014), 7,839 orthogroups were shared by all four plants (Supplementary Table S11). Interestingly, 486 of the 7,839 orthogroups of the C₄ plants contain 2 times or more genes than those of *O. sativa* (Fig. 4). In *E. vivipara*, 1,767 unigenes were assigned to the 486 orthogroups (Supplementary Table S12). Comparison of MapMan annotations of the 1,767 unigenes between the expanded 486 and the total 8,928 orthogroups (Fig. 5A and B) revealed that unigenes belonging to 13 MapMan bins might have been enriched during the C₄-associated evolution (Supplementary Table S12). Among the enriched unigenes, those assigned to the “Cell wall”, “RNA”, and “Development” classes, which have been pointed out to contribute to the differentiation and development of C₄-specific mesophyll and bundle-sheath cells (Chang et al., 2013; Yu et al., 2015), were found. For example, 4 of the SCARECROW genes and 2 of SHORT ROOT genes both of which are suggested to be involved in the development of bundle sheath cells (Cui et al., 2014; Gao et al., 2014) and 2 of the nodulin (*Nodule inception 1*) genes which are reported to be involved in vascular strand formation in gramineous plants (Gamas et al., 1996) and Kranz structure development (Chang et al., 2012) were found in the “Development” class.

**CONCLUSIONS**

In this study, we generated the first comprehensive transcriptome dataset for *E. vivipara*, a C₃–C₄
photosynthesis interconvertible plant. Total of 27,249 non-redundant unigenes were obtained. The information presented in this study will be valuable in future studies to elucidate the evolution of the C₄ photosynthetic system and to identify those genes that confer the C₄ traits to C₃ plants for the improvement of crop productivity.

**SUPPLEMENTARY DATA**

Supplementary data are available on the J-Stage.

**FUNDING**

This work was supported by the Project Research Fund of the School of Biology-Oriented Science and Technology, Kindai University [grant number 10-I-3].

**DISCLOSURES**

No conflicts of interest are declared.

**ACKNOWLEDGMENTS**

We thank Prof. O. Ueno for *E. vivipara* material and collaboration, and the HPC Computing Center in Kindai University for giving us the computing resources.

**REFERENCES**

Agarie, S., Kai, M., Takatsuji, H., Ueno, O. 1997. Expression of C₃ and C₄ photosynthetic characteristics in the amphiobious plant *Eleocharis vivipara*: structure and analysis of the expressions of isogenes for pyruvate, orthophosphate dikinase. Plant Mol. Biol. 34: 363–369.

Andorf, C. M., Cannon, E. K., Portwood, J. L., Gardiner, J. M., Harper, L. C., Schaeffer, M. L., Braun, B. L., Campbell, D. A., Vinnakota, A. G., Sribalusu, V. V., Huerta, M., Cho, K. T., Wimalanathan, K., Richter, J. D., Mauch, E. D., Rao, B. S., Birkett, S. M., Sen, T. Z., Lawrence-Dill, C. J. 2016. MaizeGDB update: new tools, data and interface for the maize model organism database. Nucleic Acids Res. 44: D1195–201.

Black, Jr. C. C. 1979. Photosynthetic carbon fixation in relation to net CO₂ uptake. Annu. Rev. Plant Physiol. 24: 253–286.

Brown, N. J., Parsley, K., Hibberd, J. M. 2005. The future of C₃ research—maize, *Flaveria* or *Cleome?* Trends Plant Sci. 10: 215–221.

Chang, Y. M., Chang, C. L., Li, W. H., Shih, A. C. 2013. Historical profiling of maize duplicate genes sheds light on the evolution of C₄ photosynthesis in grasses. Mol. Phylogenet. Evol. 66: 453–462.

Chang, Y. M., Liu, W. Y., Shih, A. C., Shen, M. N., Lu, C. H., Lu, M. Y., Yang, H. W., Wang, T. Y., Chen, S. C., Chen, S. M., Li, W. H., Ku, M. S. 2012. Characterizing regulatory and functional differentiation between maize mesophyll and bundle sheath cells by transcriptomic analysis. Plant Physiol. 160: 165–177.

Chen, T., Zhu, X. G., Lin, Y. 2014. Major alterations in transcript profiles between C₃-C₄ and C₄ photosynthesis of an amphiobious species *Eleocharis baldwinii*. Plant Mol. Biol. 86: 93–110.

Chikhi, R., Medvedev, P. 2013. Informed and automated k-mer size selection for genome assembly. Bioinformatics 30: 31–37.

Covshoff, S., Hibberd, J. M. 2012. Integrating C photosynthesis into C₃ crops to increase yield potential. Curr. Opin. Biotechnol. 23: 209–214.

Cui, H., Kong, D., Liu, X., Hao, Y. 2014. SCARECROW, SCR-LIKE 23 and SHORT-ROOT control bundle sheath cell fate and function in *Arabidopsis thaliana*. Plant J. 78: 319–327.

Gamas, P., Niebel, F. C., Lescure, N. C. J. 1996. Use of a subtractive hybridization approach to identify new *Medicago truncatula* genes induced during root nodule development. Mol. Plant Microbe Interact. 9: 233–242.

Gao, X., Wang, C., Cui, H. 2014. Identification of bundle sheath cell fate factors provides new tools for C₃-to-C₄ engineering. Plant Signal Behav. 1: 9.

Gowik, U., Bräutigam, A., Weber, K. L., Weber, A. P. M., Westhoff, P. 2011. Evolution of C₄ photosynthesis in the genus *Flaveria*: how many and which genes does it take to make C₄? Plant Cell 23: 2087–2105.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Ginrite, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Rogey, A. 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat. Biotechnol. 29: 644–652.

Hu, Z. L., Bao, J., Reecy, J. M. 2008. CateGOrizer: a web-based program to batch analyze gene ontology classification categories. Online J. Bioinform. 9: 108–112.

Huang, Y., Niu, B., Gao, Y., Fu, L., Li, W. 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics 26: 680–682.

Häusler, R. E., Hirsch, H. J., Kreuzaler, F., Peterhansel, C. 2002. Overexpression of C₄-cycle enzymes in transgenic C₃ plants: a biotechnological approach to improve C₃-photosynthesis. J. Exp. Bot. 53: 591–607.

Külahołgu, C., Denton, A. K., Sommer, M., Maß, J., Schliesky, S., Wrobel, T. J., Berckmans, B., Góngora-Castillo, E., Ebuel, C. R., Simon, R., De Veylder, L., Bräutigam, A., Weber, A. P. 2014. Comparative transcriptome atlases reveal altered gene expression modules between two Cleomaceae C₃ and C₄ plant species. Plant Cell 26: 3243–3260.

Langdale, J. A. 2011. C Cycles: past, present, and future research on C photosynthesis. Plant Cell 23: 1250–1258.

Langmead, B., Trapnell, C., Pop, M., Salzberg, S. L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10: R25.

Li, L., Christian, J., Stoeckert, Jr., David, S. R. 2003. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Res. 13: 2178–2189.

Lohse, M., Nagel, A., Herter, T., May, P., Schroda, M., Zrenner, R., Tohge, T., Fernie, A. R., Sittl, M., Usadel, B. 2014. Mercator: a fast and simple web server for genome scale functional annotation of plant sequence data. Plant Cell Environ. 37: 1250–1258.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D. W., Yiu, S. M., Peng, S., Xiaoqian, Z., Liu, G., Xiao, J., Li, Y., Yang, H., Wang, J., Lam, T. W., Wang, J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Giga Sci. 1: 18.

Miyao, M., Masumoto, C., Miyazawa, S., Fukaya, H. 2011. Lessons from engineering a single-cell C photosynthetic pathway into rice. J. Exp. Bot. 62: 3021–3029.

Monaco, M. K., Stein, J., Naithani, S., Wei, S., Dharmawardhana, P., Kumari, S., Amarasinghe, V., Youens-Clark, K.,
Thomason, J., Preece, J., Pasternak, S., Olson, A., Jiao, Y., Lu, Z., Bolser, D., Kerhornou, A., Staines, D., Walts, B., Wu, G., D’Eustachio, P., Haw, R., Croft, D., Kersey, P. J., Stein, L., Jaiswal, P., Ware, D. 2014. Gramene 2013: comparative plant genomics resources. Nucleic Acids Res. 42: D1193–D1199.

Nagasaki, H., Mochizuki, T., Kodama, Y., Saruhashi, S., Morizaki, S., Sugawara, H., Ohyanagi, H., Kurata, N., Okubo, K., Takagi, T., Kaminuma, E., Nakamura, Y. 2013. DDBJ read annotation pipeline: a cloud computing-based pipeline for high-throughput analysis of next-generation sequencing data. DNA Res. 20: 383–390.

Sage, R. F., Zhu, X. G. 2011. Exploiting the engine of C4 photosynthesis. J. Exp. Bot. 62: 2989–3000.

Sakai, H., Lee, S. S., Tanaka, T., Numa, H., Kim, J., Kawahara, Y., Wakimoto, H., Yang, C. C., Iwamoto, M., Abe, T., Yamada, Y., Muto, A., Inokuchi, H., Ikemura, T., Matsumoto, T., Sasaki, T., Itoh, T. 2013. Rice Annotation Project Database (RAP-DB): An integrative and interactive database for rice genomics. Plant Cell Physiol. 54: 2×e6.

Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J. M., Birol, I. 2009. ABySS: A parallel assembler for short read sequence data. Genome Res. 19: 1117–1123.

Taniguchi, Y., Ohkawa, H., Masumoto, C., Fukuda, T., Tamai, T., Lee, K., Sudoh, S., Tsuchida, H., Sasaki, H., Fukayama, H., Miyao, M. 2008. Overproduction of C4 photosynthetic enzymes in transgenic rice plants: an approach to introduce the C4-like photosynthetic pathway into rice. J. Exp. Bot. 59: 3021–3029.

Ueno, O. 2001. Environmental regulation of C3 and C4 differentiation in the amphibious sedge Eleocharis vivipara. Plant Physiol. 127: 1524–1532.

Ueno, O., Samejima, M., Muto, S., Miyachi, S. 1988. Photosynthetic characteristics of an amphibious plant, Eleocharis vivipara: Expression of C3 and C4 modes in contrasting environments. Proc. Natl. Acad. Sci. U.S.A. 85: 6733–6737.

Xu, J., Li, Y., Ma, X., Ding, J., Wang, K., Wang, S., Tian, Y., Zhang, H., Zhu, X. G. 2013. Whole transcriptome analysis using next-generation sequencing of model species Setaria viridis to support C photosynthesis research. Plant Mol Biol. 83: 77–87.

Yu, C. P., Chen, S. C., Chang, Y. M., Liu, W. Y., Lin, H. H., Lin, J. J., Chen, H. J., Lu, Y. J., Wu, Y. H., Lu, M. Y., Lu, C. H., Shih, A. C., Ku, M. S., Shiu, S. H., Wu, S. H., Li, W. H. 2015. Transcriptome dynamics of developing maize leaves and genomewide prediction of cis elements and their cognate transcription factors. Proc. Natl. Acad. Sci. U.S.A. 112: E2477–E2486.

Zerbino, D. R., Birney, E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18: 821–829.