Gene 600 (2017) 36–43

Contents lists available at ScienceDirect

Gene

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

Research paper

Complete plastid genome sequence of goosegrass (*Eleusine indica*) and comparison with other Poaceae

Hui Zhang [a], Nathan Hall [b], J. Scott. McElroy [a,*], Elijah K. Lowe [c,d], Leslie R. Goertzen [b]

[a] Department of Crop, Soil and Environmental Science, Auburn University, AL 36849, USA
[b] Department of Biological Sciences, Auburn University, AL 36849, USA
[c] Department of Biology and Evolution of Marine Organisms, Stazione Zoologica Anton Dohrn, Villa Comunale, 80121 Napoli, Italy
[d] BEACON Center for the Study of Evolution in Action, Michigan State University, East Lansing, MI, USA

ARTICLE INFO

Article history:
Received 7 July 2016
Received in revised form 12 November 2016
Accepted 24 November 2016
Available online 27 November 2016

Keywords:
Chloroplast genome
Next-generation sequencing
Goosegrass
Simple sequence repeat
Plastomes comparation

ABSTRACT

Eleusine indica, also known as goosegrass, is a serious weed in at least 42 countries. In this paper we report the complete plastid genome sequence of goosegrass obtained by de novo assembly of paired-end and mate-paired reads generated by Illumina sequencing of total genomic DNA. The goosegrass plastome is a circular molecule of 135,151 bp in length, consisting of two single-copy regions separated by a pair of inverted repeats (IRs) of 20,919 bases. The large (LSC) and the small (SSC) single-copy regions span 80,667 bases and 12,646 bases, respectively. The plastome of goosegrass has 38.19% GC content and includes 108 unique genes, of which 76 are protein-coding, 28 are transfer RNA, and 4 are ribosomal RNA. The goosegrass plastome sequence was compared to eight other species of Poaceae. Although generally conserved with respect to Poaceae, this genomic resource will be useful for evolutionary studies within this weed species and the genus *Eleusine*.

1. Introduction

The chloroplast plays a significant role in numerous plant cell functions, including photosynthesis, the manufacture of certain amino acids and lipids, starch, pigment production, and some key aspects of nitrogen and sulfur metabolism (Cui, 2006). It is considered to have originated from cyanobacteria through endosymbiosis (Raven and Allen, 2003). Chloroplast genomes most commonly exist as a single large circular DNA molecule, typically range in size from 120 to 170 kilobase pairs (kb) (Shaw et al., 2007). In angiosperms, chloroplast genomes have a quadripartite organization, composed of two copies of inverted repeat (IR), one large single copy (LSC), and one small single copy (SSC) (Jansen et al., 2005; Zhao et al., 2015). During plant evolution, some chloroplast genes (i.e., *infA*, *rps16*, *ycf1*, *ycf2*, and *ycf4*) were lost through gene transfer to the nucleus or were lost from the cell entirely (Millen et al., 2001). However, the small size and highly conserved of chloroplast genome still makes it suitable and invaluable for complete sequencing and phylogenetic analysis (Cho et al., 2015).

Traditional methods to sequence the chloroplast genome or partial chloroplast genes rely on costly and time-consuming plastid isolation, PCR and amplicon sequencing. With the advent of next-generation sequencing (NGS) technology, new approaches for chloroplast genome sequencing have been gradually proposed due to their high-throughput, time-saving and low-cost (Cronn et al., 2008). The number of available complete chloroplast genomes has increased rapidly due to high-throughput sequencing technology.

Eleusine indica (2n = 2x = 18), commonly known as goosegrass, is an annual, diploid, self-pollinating grass. A single plant can produce > 50.000 small seeds easily dispersed by water and wind (Waterhouse, 1993). Goosegrass is primarily listed as an agricultural and environmental weed (Randall, 2012) and is considered a "serious weed" in at least 42 countries due to its high reproductive capacity and wide tolerance to various environments (Holm et al., 1977). Once goosegrass is established it has a notoriously tough root system making it necessary to use physical or mechanical control. In addition, the resistance of goosegrass to a wide range of herbicides has greatly compounded the difficulties with its control. According to the International Survey of Herbicide Resistant Weeds, a total of 25 resistance cases have been reported in goosegrass (Chen et al., 2015). In addition, there is also strong evidence that *E. indica* is the maternal genome donor of the domesticated crop species finger millet (*E. coracana*) (Bisht and Mukai, 2001a, 2001b; Hilu, 1988 and Hiremath and Salimath, 1992).

Despite the rapidly developing technology, there are relatively few weed genomes and transcriptome available. Transcriptome assemblies

Abbreviations: NCBI, National Center for Biotechnology Information; tRNA, transfer RNA; rRNA, ribosomal RNA; LSC, large single-copy; SSC, small single-copy; IR, inverted repeat; nt, nucleotide.

* Corresponding author at: Department of Crop, Soil and Environmental Science, Auburn University, AL, USA.
E-mail address: jsm0010@auburn.edu (J.S. McElroy).

http://dx.doi.org/10.1016/j.gene.2016.11.038
0378-1119/

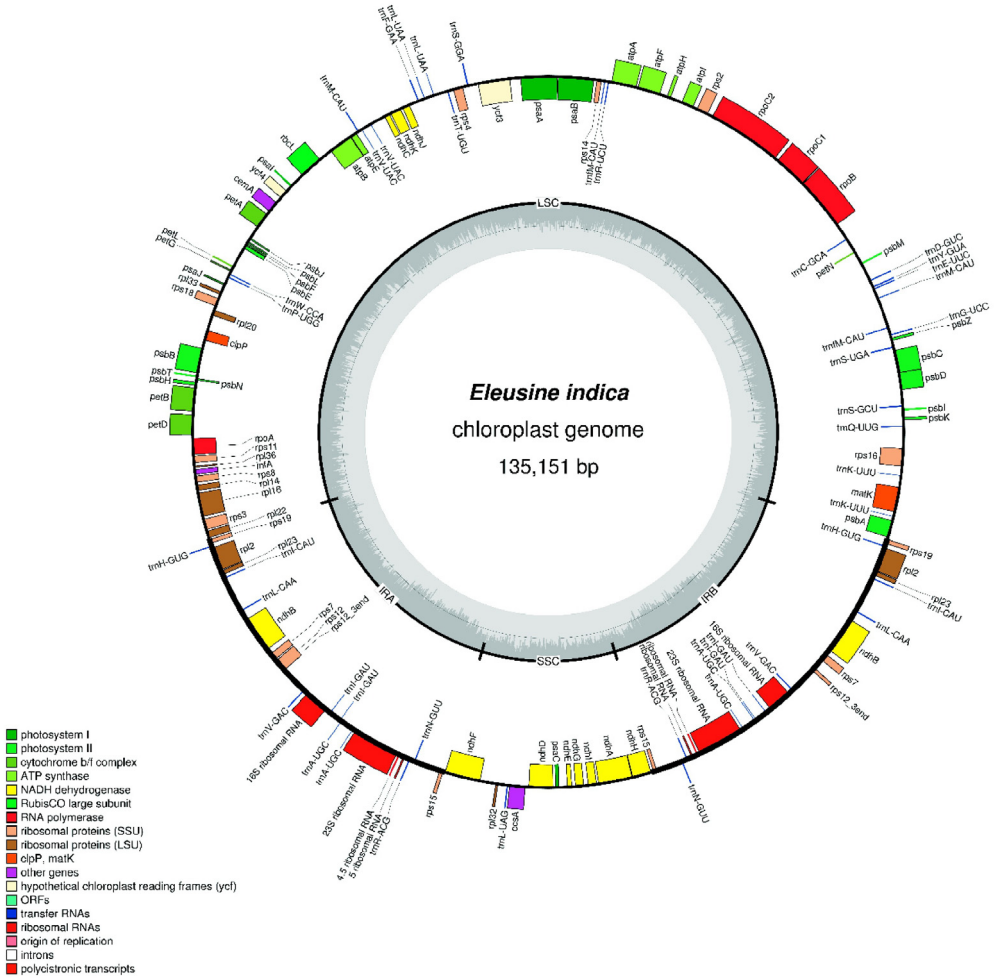

Fig. 1. Gene map of the *Eleusine indica* plastid genome sequence. Genes shown outside the outer circle are transcribed counterclockwise, and those inside are transcribed clockwise. Genes belonging to different functional groups are color coded. The innermost darker gray corresponds to GC while the lighter gray corresponds to AT content.

have been produced for goosegrass (An et al., 2014; Chen et al., 2015), and no complete chloroplast genome has been reported for any *Eleusine* species. Recently, 78 plastid protein coding loci were sequenced for *E. coracana* (Givnish et al., 2010). Here, we present the complete chloroplast genome sequence of goosegrass based on a high-throughput sequencing approach and perform comparative analyses of the plastid genomes of goosegrass and other Poaceae.

2. Materials and methods

2.1. Plant materials and DNA extraction

The goosegrass population used in this research has been previously utilized for transcriptomic research (Chen et al., 2015). We refer to this goosegrass biotype as PBU, as it was collected from the E.V. Smith Research and Education Center-Plant Breeding Unit (PBU) of Auburn University. We consider this biotype as a holotype because it possesses typical characteristics of goosegrass grown in row-crop agricultural settings. Previously collected seeds were grown in potting medium (Miracle-Gro Potting Soil, Scotts Miracle-Gro Products, and Marysville, OH) to allow uniform germination. Four weeks later, seedlings were transplanted to plastic pots (10 cm × 10 cm × 8.5 cm) containing a native Wickham sandy loam soil with pH 6.3 and 0.5% organic matter. All plants were seeded and grown in Auburn, AL (32.35°N, 85.29°W) in a glasshouse at 23 ± 2 °C, and 71% average relative humidity. Total genomic DNA was extracted from fresh leaves using the DNeasy Plant Mini Kit (Qiagen, CA, USA).

2.2. Next generation sequencing

2.2.1. DNA library preparation

At GENEWIZ, DNA samples were quantified using Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA) and the DNA integrity was checked with 0.6% agarose gel with 50–60 ng sample loaded in each well. DNA library preparations and sequencing reactions were conducted at GENEWIZ, Inc. (South Plainfield, NJ, USA). NEB NextUltra DNA Library Preparation kit was used following the manufacturer's recommendations (Illumina, San Diego, CA, USA). Briefly, the genomic DNA was fragmented by acoustic shearing with a Covaris S220 instrument. The DNA was then end repaired and adenylated. Adapters were ligated after adenylation of the 3′ ends. Adapter-ligated DNA were indexed and enriched by limited cycle PCR. DNA libraries were validated using a DNA Chip on the Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA), and were quantified using Qubit 2.0 Fluorometer. At the Auburn University Genomic and Sequencing Lab (http://www.ag.auburn.edu/enpl/gsl/seq.php), DNA libraries were constructed using Illumina's TruSeq Stranded DNA Sample Preparation Kit (Illumina, San Diego, CA USA) and library quantification was performed using the Kapa