Available online at www.sciencedirect.com
SCIENCE DIRECT
Gene 350 (2005) 117–128
GENE
SECTION
EVOLUTIONARY GENOMICS
www.elsevier.com/locate/gene

The first complete chloroplast genome sequence of a lycophyte, *Huperzia lucidula* (Lycopodiaceae)

Paul G. Wolf[a,*], Kenneth G. Karol[b], Dina F. Mandoli[b,c], Jennifer Kuehl[d], K. Arumuganathan[e], Mark W. Ellis[a], Brent D. Mishler[f,g], Dean G. Kelch[f,g], Richard G. Olmstead[b], Jeffrey L. Boore[d,f]

[a]*Department of Biology, Utah State University, 5305 Old Main Hill, Logan, UT 84322-5305, USA*
[b]*Department of Biology, University of Washington, Seattle, WA 98195-5325, USA*
[c]*Center for Developmental Biology, University of Washington, Seattle, WA 98195-5325, USA*
[d]*DOE Joint Genome Institute and Lawrence Berkeley National Laboratory, Walnut Creek, CA 94598, USA*
[e]*Benaroya Research Institute at Virginia Mason, Seattle, WA 98101, USA*
[f]*Department of Integrative Biology, University of California, Berkeley, CA 94720, USA*
[g]*University Herbarium, Jepson Herbarium, University of California, Berkeley, CA 94720, USA*

Received 11 August 2004; received in revised form 29 November 2004; accepted 24 January 2005
Available online 19 March 2005
Received by A. Roger

Abstract

We used a unique combination of techniques to sequence the first complete chloroplast genome of a lycophyte, *Huperzia lucidula*. This plant belongs to a significant clade hypothesized to represent the sister group to all other vascular plants. We used fluorescence-activated cell sorting (FACS) to isolate the organelles, rolling circle amplification (RCA) to amplify the genome, and shotgun sequencing to 8× depth coverage to obtain the complete chloroplast genome sequence. The genome is 154,373 bp, containing inverted repeats of 15,314 bp each, a large single-copy region of 104,088 bp, and a small single-copy region of 19,657 bp. Gene order is more similar to those of mosses, liverworts, and hornworts than to gene order for other vascular plants. For example, the *Huperzia* chloroplast genome possesses the bryophyte gene order for a previously characterized 30 kb inversion, thus supporting the hypothesis that lycophytes are sister to all other extant vascular plants. The lycophyte chloroplast genome data also enable a better reconstruction of the basal tracheophyte genome, which is useful for inferring relationships among bryophyte lineages. Several unique characters are observed in *Huperzia*, such as movement of the gene *ndhF* from the small single copy region into the inverted repeat. We present several analyses of evolutionary relationships among land plants by using nucleotide data, inferred amino acid sequences, and by comparing gene arrangements from chloroplast genomes. The results, while still tentative pending the large number of chloroplast genomes from other key lineages that are soon to be sequenced, are intriguing in themselves, and contribute to a growing comparative database of genomic and morphological data across the green plants.

Keywords: Organelle; Evolution; Gene order

1. Introduction

Green plants are an old group dating back about 1 billion years (Mishler, 2000). There are about half a million extant species (Mishler, 2000), including the main primary energy producers in both terrestrial and aquatic ecosystems. Reconstructing the pattern and processes of the evolution of this large and diverse group is imperative, yet challenging. Arguably, the fastest growing front in these efforts is the rapid growth in genome sequencing, which has ignited the

Abbreviations: LSC, large single-copy region; SSC, small single-copy region; IR, inverted repeat; RCA, rolling circle amplification; FACS, fluorescence activated cell sorting; BS, bootstrap support; Γ, gamma-distributed rates; ML, maximum likelihood; MP, maximum parsimony.

* Corresponding author. Tel.: +1 435 797 4034; fax: +1 435 797 1575.
E-mail address: wolf@biology.usu.edu (P.G. Wolf).

0378-1119/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2005.01.018

fields of comparative and evolutionary genomics (Normile, 2001). Although large scale phylogenetic analyses of complete eukaryotic nuclear genomes are just beginning, many phylogenomic studies of the much smaller organellar genomes are complete or underway. Most of this work has been on animal mitochondrial genomes (Boore, 1999), of which over 400 species are currently represented in public databases. More recently, chloroplast genomes have been sequenced from several clades of green plants and these genomes have been found to contain considerable amounts of phylogenetically useful data (Lemieux et al., 2000).

The chloroplasts of green plants are descendents of cyanobacteria that established an endosymbiotic relationship with a primitive eukaryote. Although many proteins necessary for chloroplast functioning are imported from the cytoplasm, chloroplasts have retained their own, now diminished, genome (Stoebe et al., 1999), along with systems for expressing these genes. Across green plants, there is a high degree of consistency in chloroplast genome structure and in gene content and arrangement (Palmer and Stein, 1986). However, these features vary sufficiently among lineages to provide useful characters for phylogenetic reconstruction. Such genome-level characters have proven to be especially robust indicators of evolutionary relatedness due to their complexity and low frequency of reversal (Helfenbein and Boore, 2004).

Comparing complete chloroplast genome sequences also enables a reconstruction of events, such as gene transfers between intracellular compartments (i.e., nucleus, chloroplast, mitochondrion), and a better understanding of the evolutionary processes that account for the features of today's chloroplast genomes. Unfortunately, as of the beginning of 2004, there are still only 25 complete chloroplast genomes published and many critical clades remain unrepresented. Here we describe the first of a series of complete chloroplast genome sequences selected to fill important phylogenetic gaps, initially focusing on land plants. Currently, complete chloroplast genomes are available from each of the three main bryophyte lineages (a hornwort, a moss, and a liverwort), 2 ferns, 2 gymnosperms, and 13 angiosperms. These taxa represent the bulk of phylogenetic diversity, but no chloroplast genome sequence has been published for any lycophyte. This is somewhat surprising because the best evidence that the lycophytes are sister to remaining extant vascular plants comes from the observation of a 30 kb inversion in the chloroplast genome, detected by restriction-site mapping studies (Raubeson and Jansen, 1992). Here we describe (1) the first complete chloroplast genome sequence of a lycophyte (*Huperzia lucidula* (Michx.) Trevis.); (2) a novel method of providing chloroplast genome-enhanced material from which to obtain the sequence; and (3) the unique aspects of the genome. We also present phylogenetic analyses based on amino acid sequences and DNA sequences extracted from published land plant chloroplast genomes plus that of *H. lucidula*. Furthermore, we explore the use of genome structure to infer land plant phylogeny.

2. Materials and methods

2.1. Preparation and DNA sequencing

Vegetative material of *H. lucidula* was collected from Balsam Gap Overlook, NC (USA). A voucher specimen (Renzaglia #3200) is deposited at the University of California Herbarium at Berkeley (UC). Purified fractions of intact chloroplasts of *H. lucidula* were collected by fluorescence-activated cell sorting (FACS). One hundred milligrams of fresh leaf tissue was placed on ice in a sterile plastic Petri dish containing 1.0 mL of an organelle isolation solution containing 0.33 M sorbitol, 50 mM HEPES at pH 7.6, 2 mM EDTA, 1 mM $MgCl_2$, 0.1% BSA, 1% PVP-40, and 5 mM β-Mercaptoethanol, and the tissue was sliced into 0.25–1 mm segments. Suspended organelles (chloroplasts, mitochondria, and nuclei) were withdrawn using a pipette, filtered through 30 μm nylon mesh, and stained with 2 μg/mL DAPI (Sigma-Aldrich, St. Louis, MO, USA) and 100 nM Mitotracker Green (Molecular Probes, Eugene, OR, USA). The organelle suspension was incubated on ice for 15 min, then analyzed on a FACS DiVa using sterile phosphate buffered solution (Invitrogen, Carlsbad, CA, USA) as sheath fluid. We used a Coherent INNOVA Enterprise Ion laser (Coherent, Santa Paula, CA, USA) emitting a 488 nm beam at 275 mW to excite chlorophyll and Mitotracker Green, and a UV beam at 30 mW to excite DAPI. Red fluorescence from chlorophyll was passed through 675±20 nm filter, held within the FL3 photomultiplier tube (PMT), and green fluorescence from Mitotracker Green was passed through a 530±30 nm filter held within the FL1 PMT. DAPI fluorescence from DNA was passed through a 424±44 nm pass filter held within the FL4 PMT. Organelles were collected into separate sterile 15 ml centrifuge tubes by flow cytometric sorting based on the respective sorting gates (Fig. 1). Sorted organelles were pelleted and shipped frozen for DNA isolation and amplification.

The DNA preparation was then processed for sequencing by the Production Genomics Facility of the DOE Joint Genome Institute. Template was first amplified through rolling circle amplification (RCA) with random hexamers (Dean et al., 2001). The DNA was then mechanically sheared into random fragments of about 3 kb by repeated passage through a narrow aperture using a Hydroshear device (Genemachines, San Carlos, CA, USA). These fragments were then enzymatically repaired to ensure blunt ends, purified by gel electrophoresis to select for a narrow distribution of fragment sizes, ligated into dephosphorylated pUC18 vector, and transformed into *E. coli* to create plasmid libraries. Automated colony pickers were used to select and transfer colonies into 384-well plates containing LB media and glycerol. After overnight incubation, a small