

Capturing the Biofuel Wellhead and Powerhouse: The Chloroplast and Mitochondrial Genomes of the Leguminous Feedstock Tree *Pongamia pinnata*

Stephen H. Kazakoff¹, Michael Imelfort², David Edwards³, Jasper Koehorst¹, Bandana Biswas¹, Jacqueline Batley¹, Paul T. Scott¹, Peter M. Gresshoff^{1*}

1 Australian Research Council Centre of Excellence for Integrative Legume Research, The University of Queensland, Brisbane, Queensland, Australia, **2** Advanced Water Management Centre, The University of Queensland, Brisbane, Queensland, Australia, **3** Australian Centre for Plant Functional Genomics, The University of Queensland, Brisbane, Queensland, Australia

Abstract

Pongamia pinnata (syn. *Millettia pinnata*) is a novel, fast-growing arboreal legume that bears prolific quantities of oil-rich seeds suitable for the production of biodiesel and aviation biofuel. Here, we have used Illumina® 'Second Generation DNA Sequencing (2GS)' and a new short-read *de novo* assembler, SaSSy, to assemble and annotate the *Pongamia* chloroplast (152,968 bp; cpDNA) and mitochondrial (425,718 bp; mtDNA) genomes. We also show that SaSSy can be used to accurately assemble 2GS data, by re-assembling the *Lotus japonicus* cpDNA and in the process assemble its mtDNA (380,861 bp). The *Pongamia* cpDNA contains 77 unique protein-coding genes and is almost 60% gene-dense. It contains a 50 kb inversion common to other legumes, as well as a novel 6.5 kb inversion that is responsible for the non-disruptive, re-orientation of five protein-coding genes. Additionally, two copies of an inverted repeat firmly place the species outside the subclade of the Fabaceae lacking the inverted repeat. The *Pongamia* and *L. japonicus* mtDNA contain just 33 and 31 unique protein-coding genes, respectively, and like other angiosperm mtDNA, have expanded intergenic and multiple repeat regions. Through comparative analysis with *Vigna radiata* we measured the average synonymous and non-synonymous divergence of all three legume mitochondrial (1.59% and 2.40%, respectively) and chloroplast (8.37% and 8.99%, respectively) protein-coding genes. Finally, we explored the relatedness of *Pongamia* within the Fabaceae and showed the utility of the organellar genome sequences by mapping transcriptomic data to identify up- and down-regulated stress-responsive gene candidates and confirm *in silico* predicted RNA editing sites.

Citation: Kazakoff SH, Imelfort M, Edwards D, Koehorst J, Biswas B, et al. (2012) Capturing the Biofuel Wellhead and Powerhouse: The Chloroplast and Mitochondrial Genomes of the Leguminous Feedstock Tree *Pongamia pinnata*. PLoS ONE 7(12): e51687. doi:10.1371/journal.pone.0051687

Editor: Jonathan H. Badger, J. Craig Venter Institute, United States of America

Received: August 14, 2012; **Accepted:** November 5, 2012; **Published:** December 14, 2012

Copyright: © 2012 Kazakoff et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The authors thank BioEnergy Plantation Australia for research support in *Pongamia*, BES Pty Ltd. for university liaison. Research was also funded by an Australian Research Council Centre of Excellence grant to PMG. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: This research was funded in part by BioEnergy Plantation Australia and BES Pty Ltd. BioEnergy Plantation Australia is developing critical technology platforms for the genetic, genomic and biotechnology application to the legume tree *Pongamia pinnata*. There are no patents, products in development or marketed products to declare. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: p.gresshoff@uq.edu.au

Introduction

In view of diminishing liquid fossil fuel reserves, the large-scale supply of plant-derived oil for sustainable biofuel production is an important challenge for translational biotechnology. *Pongamia pinnata* (also called *Millettia pinnata*), is an undomesticated arboreal legume that can provide biofuel feedstock through high productivity of oil-rich seeds [1]. Its drought- and salinity-tolerant capabilities, free it from concerns regarding the use of land, labour and water [2]. However, it is the low nitrogen input requirements that really give *Pongamia* an 'edge' over other biofuel candidates. As a legume, *Pongamia* is capable of forming specialized organs, called root nodules, through symbiotic interactions with nitrogen-fixing soil bacteria collectively called Rhizobia. Biological nitrogen fixation, resulting from a functional symbiosis, negates or reduces the need for nitrogen fertilizer application, which is critical for environmentally sustainable plant oil production. With respect to greenhouse gas emissions, all steps in the production, transport

and application of nitrogen fertilizers are costly, can cause significant environmental pollution through run-off and seepage, and are highly dependent on fossil fuels. For *Pongamia* to become an industrial-scale feedstock for the emerging biodiesel and aviation biofuel industries, biotechnologies will be needed to advance selection and breeding ahead of the impending peak oil crisis.

Central to the domestication of *Pongamia* is an understanding of its molecular genetics, and this includes aspects of gene content and structural organization in relation to other plant genomes [3]. Here, we have taken a crucial step forward in the development pathway of this biofuel tree by using Illumina® Second Generation Sequencing (2GS) and SaSSy bioinformatics software [4] to sequence and assemble its chloroplast (cpDNA) and mitochondrial (mtDNA) genomes. At a fraction of the cost of whole nuclear genome sequencing, the *Pongamia* organellar genomes not only provide valuable insight into the evolution of a largely unexplored species, but will also greatly assist and accelerate the process of