Industrial Crops & Products 109 (2017) 300–309

Contents lists available at ScienceDirect

Industrial Crops & Products

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

Research paper

A century of guayule: Comprehensive genetic characterization of the US national guayule (*Parthenium argentatum* A. Gray) germplasm collection

Daniel C. Ilut[a], Paul L. Sanchez[b], Terry A. Coffelt[b], John M. Dyer[b], Matthew A. Jenks[c], Michael A. Gore[a,*]

[a] *Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, 14853, USA*
[b] *USDA-ARS, U.S. Arid-Land Agricultural Research Center, 21881 North Cardon Lane, Maricopa, AZ, 85138, USA*
[c] *Division of Plant and Soil Sciences, West Virginia University, Morgantown, WV 26506, USA*

ARTICLE INFO

Keywords:
Genotyping-by-sequencing
Guayule
Mariola
Introgression
Parthenium
Polyploidy

ABSTRACT

The fragility of a single-source, geographically concentrated supply of natural rubber, a critical material of the modern economy, has brought guayule (*Parthenium argentatum* A. Gray) to the forefront as an alternative source of natural rubber. The improvement of guayule for commercial-scale production has been limited by the lack of genomic tools and well-characterized genetic resources required for genomics-assisted breeding. To address this issue, we developed nearly 50,000 single-nucleotide polymorphism (SNP) genetic markers and genotyped 69 accessions of guayule and its sister taxa mariola (*Parthenium incanum* Kunth), representing the entire publically available US national germplasm collection. We identified multiple interspecific hybrid accessions previously considered guayule, including five guayule-mariola hybrids and non-mariola interspecific hybrid accessions AZ-2 and AZ-3, two commonly used high-yielding cultivars. We dissected patterns of genetic diversity within the collection to identify a highly diverse subset of guayule accessions, and showed that wild guayule stands in Big Bend National Park, Texas, USA have the potential to provide hitherto untapped guayule genetic diversity. Together, these results provide the most thorough genetic characterization of guayule germplasm to date and lay the foundation for rapid genetic improvement of commercial guayule germplasm.

1. Introduction

Natural rubber is a critical raw material of modern society, essential to a diverse range of industries such as automotive, electronics, clothing, and health care. However, natural rubber is currently a single-source material with geographically concentrated production. The rubber tree [*Hevea brasiliensis* (Willd. Ex A. Juss.) Müll. Arg.] provides over 99.9% of the world supply of natural rubber, and over 75% of it is produced in South-Eastern Asia (Food and Agriculture Organization of the United Nations, Statistics Division; http://faostat3.fao.org). Plantations are generally clonal and vulnerable to the South American leaf blight fungal pathogen *Microcyclus ulei* (Rivano, 1997). In response to this situation, guayule (*Parthenium argentatum* A. Gray), a woody perennial shrub native to the desert regions of northern Mexico and southwestern United States, has been repeatedly assessed and utilized as an alternative source of natural rubber in the United States, and declared a critical agricultural material in times of crisis (Program for development of guayule and other rubber-bearing plants, 7 U.S.C. § 171; Critical Agricultural Materials Act, 7 U.S.C. § 178).

The sporadic nature of research funding, germplasm collection, and stock maintenance (Hammond and Polhamus, 1965; Ilut et al., 2015; Thompson and Ray, 1989), combined with ploidy variation within guayule populations (Gore et al., 2011; Ilut et al., 2015) and the unusual guayule reproductive biology—sporophytic self-incompatibility in diploid plants and facultative apomixis (diplospory type) in polyploid plants—have made it difficult so far to achieve significant increases in rubber yield, a critical step on the path to developing guayule as an alternative commercial source of natural rubber. In order to address this issue with modern genomics-assisted breeding approaches, a comprehensive characterization of ploidy variation and genetic diversity within the USDA-ARS National Arid Land Plant Genetics Resources Unit (NALPGRU) germplasm collection is essential. A previous limited study identified only two primary genotypic groups within the collection (Ilut et al., 2015), stressing the need for greatly expanded sampling of the

Abbreviations: ERP, Emergency Rubber Project; GBS, genotyping-by-sequencing; GRIN, Germplasm Resources Information Network; SNP, single-nucleotide polymorphism; NCBI, National Center for Biotechnology Information; NALPGRU, USDA-ARS National Arid Land Plant Genetics Resources Unit; NPGS, USDA-ARS National Plant Germplasm System

* Corresponding author.
E-mail addresses: dci1@cornell.edu (D.C. Ilut), plsanchez@crrf.org (P.L. Sanchez), terry.coffelt.arc@gmail.com (T.A. Coffelt), John.Dyer@ars.usda.gov (J.M. Dyer), majenks@mail.wvu.edu (M.A. Jenks), mag87@cornell.edu (M.A. Gore).

http://dx.doi.org/10.1016/j.indcrop.2017.08.029
Received 9 June 2017; Received in revised form 9 August 2017; Accepted 16 August 2017

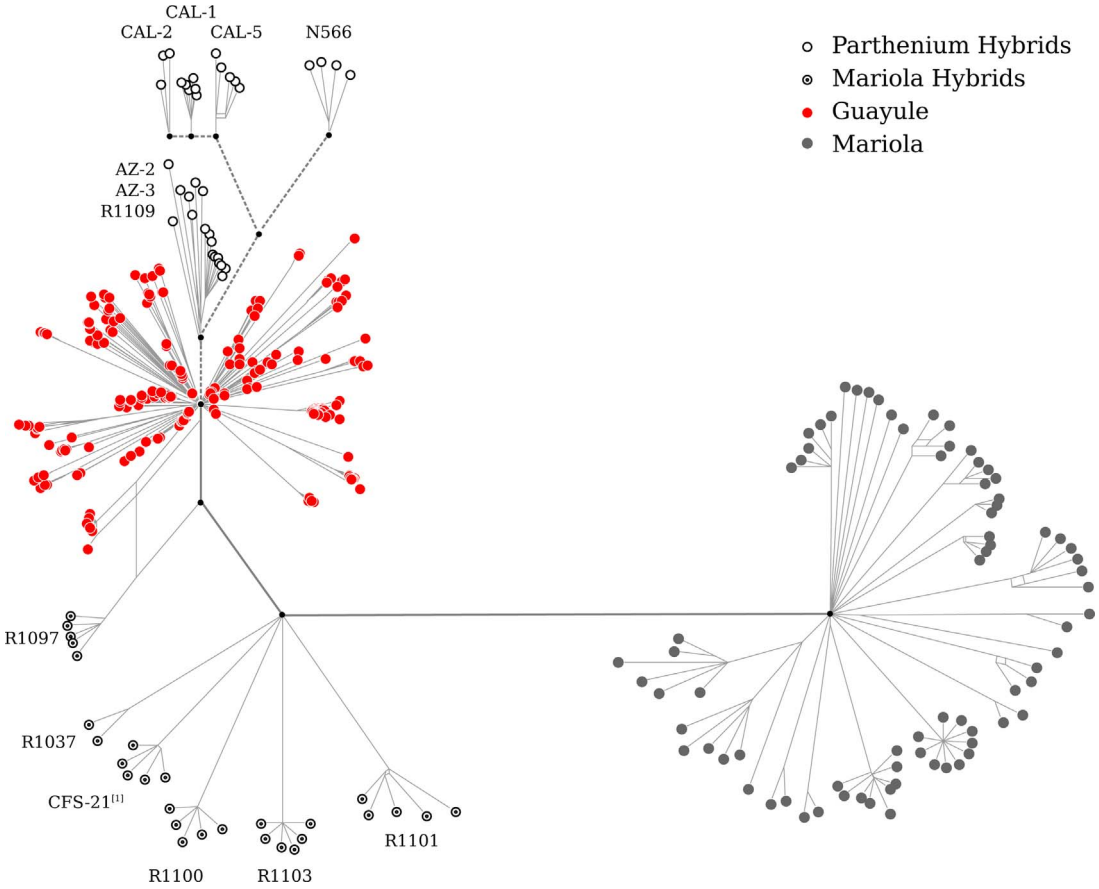

Fig. 1. Phylogenetic relationships of accessions in the NALPGRU guayule and mariola germplasm collection.
Terminal node colors and symbols indicate guayule accessions (red filled circles, center left), mariola accessions (gray filled circles, lower right), guayule x mariola hybrids (black open circle and dot, lower left), and guayule x non-mariola *Parthenium* hybrids (black open circle, top left). Hybrid accessions are identified with their respective accession names. Main branches connecting hybrids of guayule and unknown *Parthenium* species to the guayule clade are marked with dotted lines. Accessions containing a mixture of guayule, mariola, and guayule x mariola hybrids are denoted with the superscript [1].

germplasm collection in order to genetically enrich the guayule germplasm pool.

In this study, we used genotyping-by-sequencing (GBS) to simultaneously identify and score nearly 50,000 single nucleotide polymorphism (SNP) markers on the complete inventory of guayule, guayule hybrid, and mariola (*Parthenium incanum* Kunth) accessions available from the NALPGRU germplasm collection. This newly generated genomic data set combined with ploidy level information was used to (i) provide a comprehensive survey of the genetic diversity captured in the collection, (ii) identify cryptic hybrid accessions, (iii) clarify genetic relationships between accessions, and (iv) select a diverse subset of genotypes within the collection.

2. Materials and methods

2.1. Plant material, growth conditions, and ploidy analysis

Seeds (achenes) for 69 accessions were obtained from the USDA-ARS National Plant Germplasm System (NPGS; www.ars-grin.gov/npgs) through the NALPGRU in Parlier, CA. Plants were greenhouse grown and leaf tissue samples were collected and prepared for ploidy analysis as previously described (Ilut et al., 2015, sec. 2.3; Sanchez et al., 2014), with an average of two technical replicates for each plant. For each plant, we calculated its ploidy level as described in Sanchez et al. (2014), then assigned an integer-valued ploidy level if the nearest integer value was within two standard deviations of the mean across technical replicates, and a fractional value (denoting likely aneuploid samples) otherwise. Supplementary Table A1 contains detailed source information for all the plants genotyped in this study, including NALPGRU seed inventory identifiers and detailed ploidy measurements.

2.2. DNA extraction, sequencing, and genotyping

On average, we selected six plants (samples) per accession and followed previously described protocols for tissue sampling, DNA library construction, and sequencing (Ilut et al., 2015, sec. 2.6), read filtering (Ilut et al., 2015, sec. 2.7), and de-novo GBS reference sequence construction (Ilut et al., 2015, sec. 2.8). We selected samples 72_1_1-l5, 72_1_1-l3, and 72_1 (see Supplementary Table A1), which represent three technical replicates of the same tissue source, for the construction of a de-novo GBS reference sequence.

The trimmed and filtered reads for each sequencing sample were aligned against the de-novo GBS reference sequence using bwa-mem (version 0.7.12-r1039; Li, 2013), and GATK (version 3.5; McKenna et al., 2010) was used to perform indel realignment. The HaplotypeCaller and CombineGVCFs modules of GATK were used to generate sample-level genotype calls and produce combined genotyping results for all samples.

2.3. Estimation of heterozygosity levels

Heterozygosity was estimated independently for each sample. For a given sample, we extracted all nucleotides in the de-novo reference which had at least 10 reads aligned to them with a mapping quality of 30 or better using samtools (version 1.2-25; Li et al., 2009) and calculated the proportion of these nucleotides with heterozygous genotype