Computational and Structural Biotechnology Journal 15 (2017) 185–194

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

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

Deep Assessment of Genomic Diversity in Cassava for Herbicide Tolerance and Starch Biosynthesis

Jorge Duitama [a,d,*], Lina Kafuri [b,c], Daniel Tello [b,c], Ana María Leiva [a], Bernhard Hofinger [b], Sneha Datta [b], Zaida Lentini [c], Ericson Aranzales [a], Bradley Till [b,1], Hernán Ceballos [a,1]

[a] Agrobiodiversity Research Area, International Center for Tropical Agriculture (CIAT), Cali, Colombia
[b] Plant Breeding and Genetics Laboratory, Joint FAO/IAEA Division, International Atomic Energy Agency, Seibersdorf, Austria
[c] Department of Biological Sciences, School of Natural Sciences, Universidad Icesi, Cali, Colombia
[d] Systems and Computing Engineering Department, Universidad de los Andes, Bogotá, Colombia

ARTICLE INFO

Article history:
Received 31 October 2016
Received in revised form 26 December 2016
Accepted 10 January 2017
Available online 14 January 2017

Keywords:
Cassava
Pooled targeted resequencing
Herbicide tolerance
Starch biosynthesis
SNP detection

ABSTRACT

Cassava is one of the most important food security crops in tropical countries, and a competitive resource for the starch, food, feed and ethanol industries. However, genomics research in this crop is much less developed compared to other economically important crops such as rice or maize. The International Center for Tropical Agriculture (CIAT) maintains the largest cassava germplasm collection in the world. Unfortunately, the genetic potential of this diversity for breeding programs remains underexploited due to the difficulties in phenotypic screening and lack of deep genomic information about the different accessions. A chromosome-level assembly of the cassava reference genome was released this year and only a handful of studies have been made, mainly to find quantitative trait loci (QTL) on breeding populations with limited variability. This work presents the results of pooled targeted resequencing of more than 1500 cassava accessions from the CIAT germplasm collection to obtain a dataset of more than 2000 variants within genes related to starch functional properties and herbicide tolerance. Results of twelve bioinformatic pipelines for variant detection in pooled samples were compared to ensure the quality of the variant calling process. Predictions of functional impact were performed using two separate methods to prioritize interesting variation for genotyping and cultivar selection. Targeted resequencing, either by pooled samples or by similar approaches such as Ecotilling or capture, emerges as a cost effective alternative to whole genome sequencing to identify interesting alleles of genes related to relevant traits within large germplasm collections.

1. Introduction

Cassava is one of the most important crops in the tropics, surpassed only by maize and rice [1], and it is usually grown by poor farmers living in marginal and submarginal lands of the tropics [2]. It provides staple food for over 700 million people in Africa (51%), Asia (29%) and South America (20%) [3], being their main source of carbohydrates, in part due to its capacity to produce more energy per hectare than other crops [4,5]. Cassava is also preferred among other crops in these areas because it keeps competitive yields under poor soils, drought, acidic conditions, high air temperatures and evapotranspiration, pests, and diseases [6–8]. In marginal areas where grain crops often fail, cassava can strive, allowing farmers to harvest it when needed [9,10].

In addition to human and animal consumption, cassava has great potential as a source of industrial starch [11]. In fact, cassava is the second most important source of starch worldwide. In the last two decades, cassava production has increased mainly owing to its superior starch quality; which is used primarily in food-processing, paper, glue, textiles, and pharmaceutical industries or occasionally for ethanol production [8]. Therefore one important goal of cassava breeding programs is to develop new varieties with high starch content [12] and with variation in its starch functional properties [13,14]. The biosynthesis of starch involves the production of amylose and amylopectin molecules, which is catalyzed by a series of enzymes (Fig. 1). The synthesis of amylose is catalyzed by the *GBSSI* (Granule bound starch synthase) enzyme [15]. Mutations that knock out this protein are known as *waxy* mutations, because the resulting starches lack amylose [16]. There is a whole complex of enzymes involved in the synthesis of amylopectin: four soluble starch synthases (*SSI*, *SSII*, *SSIII* and *SSIV*), two types of starch branching enzymes (*SBEI* and *SBEII*), the Glucan Water Dikinase (*GWD*), and various debranching enzymes and kinases [17]. The SS and the SBE enzymes contribute glucose units to the main chain, and mediate the cleavage

* Corresponding author at: Cra 1 Este No 19A - 40, Bogotá, Colombia.
E-mail address: ja.duitama@uniandes.edu.co (J. Duitama).
[1] These authors contributed equally to this work and should be considered joint last authors.

http://dx.doi.org/10.1016/j.csbj.2017.01.002

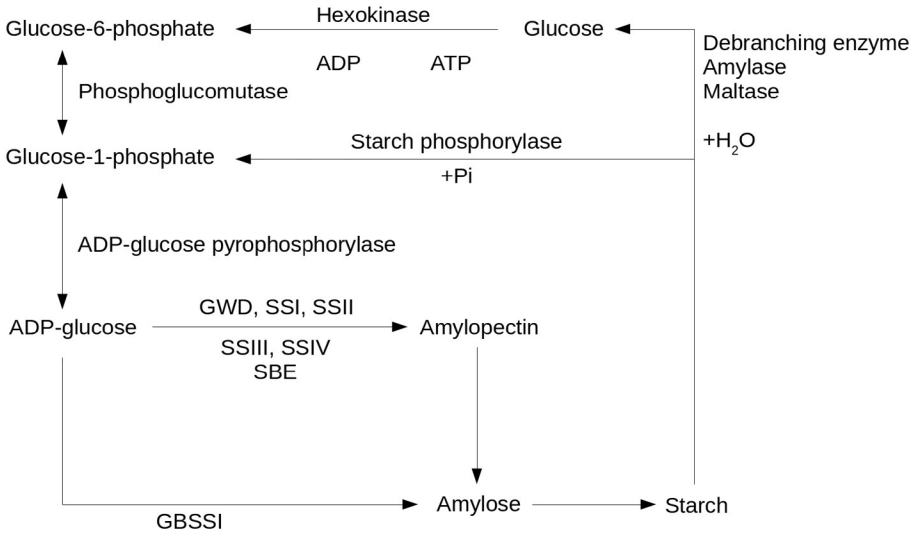

Fig. 1. Metabolic reactions related to starch biosynthesis. Arrows indicate reactions catalyzed by the enzymes listed close to the corresponding arrow.

and branch formation of the amylopectin units [18]. Alteration in SBE activity affects the number of and size distribution of amylopectin branches [17]. It is hard to determine the exact role of each isoform of the soluble starch synthases in this process due to their different gene expression, which depends on both genotypic and environmental variations [18]. GWD controls the overall rate of starch breakdown with a central rate limiting role in starch breakdown machinery and downstream starch synthesis [19]. Plants lacking this protein accumulate abnormally high levels of starch [20].

Another central goal in cassava breeding is the development of herbicide-tolerant cultivars, because the use of herbicides is an effective mechanism to control weeds, reducing labor and alleviating problems of soil erosion associated with mechanical weeding [21]. Studies on the impact of introducing herbicide resistance cassava in Colombia estimated production cost savings between 15% and 25% [22]. Additionally, the positive environmental effects which reduce tillage would bring for increased sustainability of the crop on marginal lands [23].

Resistance to two types of herbicides, inhibiting amino acid biosynthesis, has been commercially exploited in different crops and was targeted in this study. The first group of herbicides (imidazolinones, sulfonylureas, triazolopyrimidine, pyrimidinyl-thiobenzoates, and sulphonyl-aminocarbonyl-triazolinone), interact with the enzymes Acetohydroxyacid synthase (AHAS) and acetolactate synthase (ALS) [24,25]. AHAS has an important role during the synthesis of branched chain amino acids such as valine, leucine, and isoleucine, which are important for the synthesis of several proteins [24]. However, variations in just one amino acid in the binding site of AHAS enzymes can lead to a change in their quaternary structure, blocking herbicide binding and conferring tolerance in the plant. At least five naturally occurring mutations in AHAS, leading to resistance, have been reported in different plant species [24]. The second class of herbicides also affecting amino acid synthesis is the PPT (L-phosphinothricin), also known as glufosinate, and act on the glutamine synthase enzyme (GS). GS synthesizes glutamine and is very important in the regulation of the nitrogen metabolism [26,27]. With the development of transgenic technology, studies established a protocol of using somatic cotyledons as explants for the transformation of cassava [28] successfully transformed a herbicide-resistance gene into the cotyledons of cassava Per 183 by the Agrobacterium mediated method [21]. However, the development of transgenic herbicide-resistant cassava faces regulatory problems that have restricted the adoption of the technology in Africa (with the exception of South Africa).

CIAT holds in trust the largest global germplasm collection of cassava and other *Manihot* species (more than 6000 accessions). The *in vitro* collection at CIAT was initiated in 1978 soon after the technology for slow growth *in vitro* became available [29]. The germplasm collection is a valuable asset and the main repository of genetic variability of cassava. Advanced materials developed from it were the sources of amylose free starch mutations [14]. Although these discoveries provided important proof of the value of the collection, it also highlighted the limited exploration and exploitation of its genetic variability. This work also highlighted how time consuming and inefficient it is to expose useful recessive traits by conventional self-pollination methods. A recent partial screening of the collection allowed discovering varieties carrying two mutations responsible for improved starch quality traits [30]. These findings are encouraging to explore cost-effective alternatives to screen the germplasm collection in search for useful mutations for agronomically relevant traits.

In recent years, the development of high throughput sequencing technologies led to major progress in the understanding of genomic variation in plants, increasing the number of sequenced genomes [31]. However, despite the economic importance of cassava, studies of its genomic diversity are much less complete, compared to other crops such as rice, wheat or maize. Up-to-date the largest study of genomic variability in cassava, which includes 1280 accessions, is based on 402 single nucleotide polymorphisms (SNPs) scattered across the genome [32]. Although a draft cassava genome was assembled and made available in 2012 [33], a chromosome-level assembly was only achieved in 2016 [34]. In the meantime, genotyping by sequencing (GBS) has been a commonly used alternative to obtain dense datasets of genome-wide SNP markers [35]. These SNPs have been used to develop saturated genetic maps for breeding populations, genetic mapping of traits [36–38], and markers for fingerprinting [39]. More recently they have been used to perform a Genome-wide Association Study (GWAS) to identify loci related to resistance to the Cassava mosaic disease [40]. Although GBS is an efficient technique to screen markers and gather information across the genome, it does not allow the study and discovery of variability within specific genes. Sequencing of RNA has also been used as an alternative to identify expressed variation across thousands of genes [41]. However, the cost per sample of this technique is still prohibitive for large numbers of samples. For this reason, targeted resequencing remains an alternative approach to study genetic variability in specific loci.

In this study, we performed pooled targeted resequencing of DNA from 1667 cassava accessions to detect rare SNPs in specific genes associated with the starch biosynthesis pathway and with herbicide resistance. Selected accessions represent about one fourth of the entire collection and include landraces from the most important regions of