Plant Science 246 (2016) 62–69

Contents lists available at ScienceDirect

Plant Science

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

Unusual RNA plant virus integration in the soybean genome leads to the production of small RNAs

Guilherme Cordenonsi da Fonseca[a], Luiz Felipe Valter de Oliveira[b], Guilherme Loss de Morais[c], Ricardo Vilela Abdelnor[d], Alexandre Lima Nepomuceno[d], Peter M. Waterhouse[e], Laurent Farinelli[f], Rogerio Margis[a,b,*]

[a] Centro de Biotecnologia, Programa de Pós-Graduação em Biologia Celular e Molecular, Universidade Federal do Rio Grande do Sul, RS, Brazil
[b] Departamento de Genética, Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul, RS, Brazil
[c] Laboratório Nacional de Computação Científica, LNCC, Petropolis, RJ, Brazil
[d] Empresa Brasileira de Pesquisa Agropecuária, EMBRAPA, Londrina, PR, Brazil
[e] Centre for Tropical Crops and Biocommodities, Queensland University of Technology, Brisbane, Queensland, Australia
[f] Fasteris SA, Plan-les-Ouates, Switzerland

ARTICLE INFO

Article history:
Received 14 October 2015
Received in revised form
15 December 2015
Accepted 29 January 2016
Available online 11 February 2016

Keywords:
Horizontal gene transfer
RNA plant virus
Cucumber mosaic virus
Soybean
Transposon

ABSTRACT

Horizontal gene transfer (HGT) is known to be a major force in genome evolution. The acquisition of genes from viruses by eukaryotic genomes is a well-studied example of HGT, including rare cases of non-retroviral RNA virus integration. The present study describes the integration of cucumber mosaic virus RNA-1 into soybean genome. After an initial metatranscriptomic analysis of small RNAs derived from soybean, the *de novo* assembly resulted a 3029-nt contig homologous to RNA-1. The integration of this sequence in the soybean genome was confirmed by DNA deep sequencing. The locus where the integration occurred harbors the full RNA-1 sequence followed by the partial sequence of an endogenous mRNA and another sequence of RNA-1 as an inverted repeat and allowing the formation of a hairpin structure. This region recombined into a retrotransposon located inside an exon of a soybean gene. The nucleotide similarity of the integrated sequence compared to other *Cucumber mosaic virus* sequences indicates that the integration event occurred recently. We described a rare event of non-retroviral RNA virus integration in soybean that leads to the production of a double-stranded RNA in a similar fashion to virus resistance RNAi plants.

1. Introduction

Horizontal gene transfer (HGT) is characterized by the exchange of DNA between unrelated species by means other than sexual reproduction. For many years, HGT events were considered to be rare and restricted to a few types of organisms. More recently, however, the increased availability of genomic data has revealed this phenomenon to be an important force in prokaryotic evolution [1,2]. Through HGT, different species of bacteria are able to rapidly adapt to new environments by acquiring novel metabolic features, antibiotic sistance and pathogenicity [3–6]. On the other hand, HGT has long been assumed to play a limited role in the evolution of eukaryotic genomes. One of the reasons for this underestimation was the lack of sequence data from eukaryotes, and a misinterpretation of the number of genes transferred from bacteria to the human genome also helped to undermine the importance of this process in eukaryotic organisms [7–9]. Nevertheless, with the increasing number of genomes sequenced in the past few years, many cases of HGT have been reported in eukaryotes [10,11]; even the nearly complete genome of *Wolbachia* was found in the chromosome of its host insect [12]. As with bacteria, HGT events in eukaryotes can provide important adaptive traits such as virulence factors, new metabolic pathways, resistance to abiotic stress and the capability to process new types of nutrients [13–20]. Interestingly, in the asexual metazoa bdelloid rotifer, the high ratio of genes acquired from bacteria and fungi may have partially compensated for the lack of genetic variation that sexual reproduction provides [21,22].

The transfer of genetic material between viruses and eukaryotes has also been documented in past years. Retroviral sequences were found in the chromosomes of vertebrates, and although their functions remain unclear, these endogenous retroviruses comprise

* Corresponding author.
E-mail addresses: guicf13@gmail.com (G.C. da Fonseca), rogerio.margis@ufrgs.br (R. Margis).

http://dx.doi.org/10.1016/j.plantsci.2016.01.011
0168-9452/

a significant fraction of vertebrate genomes [23–25]. For a long time, plant viruses were considered to be unable of integrating into the host genome, as they did not possess integrase activity; however, sequences from pararetroviruses and DNA viruses has been encountered in plant genomes in recent years [26–28], including soybean [29]. Surprisingly, even the transfer of nonretroviral RNA viruses to plants, mammals and fungi genomes were reported recently [30–33]. The function of these nonretroviral integrated RNA viruses (NIRVs) remains elusive; however, some authors have suggested a possible role in antiviral immunity [34,35].

Cucumber mosaic virus (CMV) is one of the most common among the RNA viruses of substantial economic importance, as it can infect more than 1000 different species of plants [36]. The CMV genome is composed of three single-stranded, positive-sense RNAs. RNA 1 encodes a protein that has RNA helicase and methyltranferase activities and is required for viral replication, RNA 2 encodes the viral RNA-dependent RNA polymerase and a protein that is able to suppress host post-transcriptional gene silencing (PTGS) mechanisms and RNA 3 encodes the coat protein and a protein involved in viral movement [37]. Recent studies have reported the sequencing and identification of viruses by metatrascriptome analyses of deep sequence small RNA (sRNA) libraries [38,39]. Using a similar approach, our group identified several pathogenic, symbiotic and free-living organisms in soybean sRNA libraries [40]. Interestingly, we also identified sRNAs corresponding to RNA 1 of the CMV in all of the libraries surveyed but not for RNA 2 or RNA 3. In this work, we demonstrated that these sRNAs originate from a CMV RNA 1 sequence that previously integrated into the soybean genome.

2. Material and methods

2.1. Plant material and nucleic acid isolation

The small RNA sequences used in this work were obtained from different soybean cultivars used for the generation of deep sequencing libraries related to projects of the GenoSoja consortium [41].

Leaf and root samples were collected from *Conquista* cultivar plants grown in a hydroponic system under an aerated 6.6 pH-balanced nutrient solution. First, seeds were pre-germinated on moistened filter paper in a greenhouse with a temperature of $24 \pm 2\,^{\circ}C$ in the dark for one day and then under light with a 12:12 h photoperiod for 2 days. The plantlets were placed in polystyrene supports, and the roots were immersed in nutrient solution for 15 days until the first trifoliate leaf was fully developed (V2 developmental stage). The leaves and roots were extracted, frozen in liquid nitrogen, and stored at $-80\,^{\circ}C$ for subsequent RNA extraction.

Seeds of the *Conquista* cultivar were grown in moistened filter paper for 0, 3, 5 and 7 days in a greenhouse under a temperature regime of $24 \pm 2\,^{\circ}C$. At the end of the respective time periods, the samples were frozen in liquid nitrogen and stored at $-80\,^{\circ}C$.

Pods of the same cultivar (developmental stage R3–R5) were collected from field plants grown at the Federal University of Rio Grande do Sul (UFRGS) in Porto Alegre, Brazil.

Flowers samples (developmental stage R2–R3) from the cultivar *Urano* were collected from plants grown at the experimental field of the University of Passo Fundo (UPF), Brazil. Whole flowers, stamens, carpels and petals samples were immediately powdered in Trizol (Invitrogen, CA, USA), separately, and stored at $-80\,^{\circ}C$ for subsequent RNA extraction.

The frozen samples were ground to a fine powder in liquid nitrogen, and total RNA was isolated using Trizol (Invitrogen, CA, USA), following the manufacturer's instructions. RNA quality was evaluated by 1.0% agarose gel electrophoresis, and the amount of RNA was assessed using a Quibit fluorometer and the Quanti-iT RNA assay kit (Invitrogen, CA, USA), following the manufacturer's instructions.

Seeds from 12 different soybean cultivars (Conquista, Urano, BR 16, Embrapa 48, Williams 82, Davis, Hills, Paraná, BR 4, IAS 4, Força and Potência) were incubated in moistened filter paper for a week at $24 \pm 2\,^{\circ}C$ in a growth chamber. Leaves from the plantlets were collected and ground in liquid nitrogen. DNA was extracted using the PureLink ™ (Invitrogen, CA, USA) kit, according to the manufacturer's instructions. The DNA samples were used in posterior PCR analysis.

2.2. DNA, RNA and small RNA sequencing

Approximately 10 μg of total RNA from each sample was sent to Fasteris Life Sciences SA (Plan-les-Ouates, Switzerland) for processing and sequencing using Solexa technology. Each library was constructed with a pool of 4–5 RNA samples from different plants. The "germinating seed" library was constructed from a pool of RNA samples from seeds at 3, 5 and 7 days post incubation in moistened filter paper under $24 \pm 2\,^{\circ}C$, as described above. The construction of small RNA libraries from leaf and root samples of the cultivars *BR-16*, *Embrapa-48* and *PI 561356* was described previously [42]. The sequencing of the small RNA samples from seeds, germinating seeds, pods, flowers, carpels, petals and stamens was performed on the Illumina Genome Analyzer GAII, and for the sequencing of the small RNAs, mRNA and DNA from *Conquista* leaves, the Illumina HiSeq 2000 Sequencing System was used.

Briefly, the process for constructing the small RNA libraries consisted of the following: acrylamide gel purification of the RNA bands corresponding to the size range from 20 to 30 nt, the ligation of 3′ and 5′ adapters to the RNA in two separate subsequent steps, cDNA synthesis followed by acrylamide gel purification and PCR amplification to generate a DNA colony template library for sequencing. After the trimming of the sequences, reads from 19 to 24 nt were used for further analysis.

The mRNAseq library was construct after the Poly(A) purification of the total RNA extract followed by cDNA synthesis using Poli(T) primer. The paired-end 50 bases sequencing was perfomed in inserts of 500 nt.

2.3. De novo *assembly of small RNAs*

The 19- to 24-nt reads of the *Conquista* leaves small RNA library were assembled into contigs using the Velvet 0.7.3.1 *de novo* assembly algorithm [43] with the following parameters: a harsh length of 17–23, a coverage cut-off of 50, an expected coverage of 1000 and a minimum scaffold length of 100. Only the contigs that matched the CMV RNA 1 sequence were used to assemble the entire sequence in the Molecular Evolutionary Genetics Analysis (MEGA version 5.0) [44] software using the RNA 1 sequence of CMV isolate CM95 [GenBank: AB188234.1] as a reference.

2.4. Mapping analysis

All mapping analyses in this study were performed using the Bowtie 2 tool [45]. The small RNA reads were mapped against the reference sequence using the standard parameters, except that no mismatches were allowed (-n 0). For the RNA and DNA reads analysis, mapping was performed using the standard paired-end options, adjusting the minimum insert length to 0 (-I 0) and the maximum insert length to 1000 (-X 1000).

2.5. Determination of the entire integrated sequence

To identify the flanking regions of the inserted sequence, the *Conquista* DNA sequencing reads were first mapped, single-end and