Current Plant Biology xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Current Plant Biology

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

The repetitive component of the sunflower genome

T. Giordani, A. Cavallini, L. Natali *

Department of Agriculture, Food, and Environment, University of Pisa, Via del Borghetto 80, I-56124 Pisa, Italy

ARTICLE INFO

Article history:
Received 1 February 2014
Received in revised form 19 May 2014
Accepted 24 May 2014

Keywords:
Genome composition
Helianthus
LTR-retrotransposons
Repetitive DNA
Sunflower

ABSTRACT

The sunflower (*Helianthus annuus*) and species belonging to the genus *Helianthus* are emerging as a model species and genus for a number of studies on genome evolution. In this review, we report on the repetitive component of the *H. annuus* genome at the biochemical, molecular, cytological, and genomic levels. Recent work on sunflower genome composition is described, with emphasis on different types of repeat sequences, especially LTR-retrotransposons, of which we report on isolation, characterisation, cytological localisation, transcription, dynamics of proliferation, and comparative analyses within the genus *Helianthus*.

Contents

1. Introduction

1.1. The repetitive component of plant genomes

Eukaryotes show large variation in genome size, especially in higher plants. Angiosperm genome size (1C) ranges from 63 Mbp in *Genlisea margaretae* to 150 Gbp in *Paris japonica* [1,2]. Such differences arise from two main processes: polyploidy and amplification of transposons and related sequences. Transposon amplification has resulted in the accumulation of many repeated sequences (i.e., sequences that are identical or similar to sequences elsewhere in the genome but whose copy number is much larger than that possibly achieved through polyploidisation). Some repeats are considered to be non-functional, whereas others have played key roles in the evolution of species [3]. For example, the mutagenic action of transposons provides substantial increases in genetic variability [4]. Transposons also create novel functions by fine-tuning gene activity, resulting in phenotypic variation [5,6]. Although changes in the repetitive component have played major roles in the evolution of plant genomes, large datasets of repetitive DNA are available for such monocots as maize, rice, and barley, and a comprehensive analysis

* Corresponding author. Tel.: +39 0502216665; fax: +39 0502216661.
E-mail addresses: lucia.natali@unipi.it, lnatali@agr.unipi.it (L. Natali).

http://dx.doi.org/10.1016/j.cpb.2014.05.001

of the repetitive component is still lacking in many genera of dicots.

Among dicotyledonous species, sunflowers (*Helianthus* spp.) are emerging as a model system for genomic research on adaptation and speciation [7,8]. Given the extensive characterisation of their ecology and genomes, sunflowers are very suitable for establishing the occurrence of different types of repeats, their timing of amplification or reduction, and the role(s) of different repetitive sequences in the maintenance of a unique genome. Exploring the diversity and evolutionary dynamics of transposable elements in the sunflower (*H. annuus* L.) is of considerable importance for understanding the evolutionary history of this species, given its large genome size (3500 Mbp) [9] and the well-documented cases of retrotransposon amplification within the genus [10,11].

1.2. The Helianthus genus

The Asteraceae family is the largest plant family on Earth, with more than 24,000 described species, corresponding roughly to 10% of all angiosperms [12]. Asteraceae species live in a number of different environments, including forests, grasslands, deserts, wetlands, mountain tops, salt marshes, lawns, and agricultural fields [13]. They include economically important crops, wildflowers, valuable medicinals, invasive plants, and rangeland weeds [14]. The most important crop is the cultivated sunflower (*H. annuus* L.), which ranked 14th in 2012 among the world's food crops in terms of area harvested (http://www.fao.org/).

Helianthus includes 49 species, which are widespread in the continental United States [15], although other ecotypes are assuming the status of the species [16]. These species differ in many phenotypic traits, including reproductive timing, branching patterns, height [17,18], and especially habitat preferences. In some cases, species coexist in the same environment, and interbreeding between species is very common [19], despite large-scale karyotypic differences [20,21] and high levels of pollen and seed non-viability in the hybrids [22].

The described sunflower species, native to diverse environments throughout North America, include examples of allo- and autopolyploids [23], ecologically isolated sympatric and allopatric species [15], karyotypically divergent species [20,21,24], allopatric species with weak barriers to gene flow [15], and several homoploid hybrids [25]. Such diversity of speciation mechanisms and barriers to gene flow make sunflower an ideal model for understanding speciation and species divergence [26,27].

2. The *Helianthus* genome structure

Sunflower genomic DNA was first studied by thermal denaturation and analysis of reassociation kinetics [28,29]. Fig. 1A shows the melting profile and the first derivative curve of the DNA extracted from roots of seedlings of a selfed sunflower line. The T_m value indicates a GC content around 40%. Clear-cut shoulders were observed both on the light and the heavy sides of the derivative curves, indicating that minor specific DNA repeat families occur in the genome [28,29].

Analysis of the reassociation curves (Fig. 1B) revealed that the genome is organised into three main fractions according to their redundancy. Highly repeated (HR) sequences account for 5% of the genome, medium repeated (MR) sequences for around 60%, and the so-called unique sequences comprise the remaining 35% of the genome [28,29]. The same analyses, performed on different sunflower genotypes, showed differences in the repetitive fractions (either HR or MR), reflecting variations in the genome size [28,29].

Biochemical analyses did not consider DNA sequence but only denaturation and reassociation kinetics of DNA. Therefore, the sequence composition of the isolated fraction was not studied. Moreover, those analyses could not evaluate the occurrence in the "unique" fraction of the genome of rare forms of repeats, such as retrotransposon remnants, that were excluded from the repetitive component.

The repetitive fraction of the sunflower genome was characterised at the molecular level using a Sanger-sequenced small insert library [30]. That library provided a first set of sequences (1638, for a total of 954,517 bp) that were used, in combination with slot-blot hybridisation and fluorescent *in situ* hybridisation (FISH), to analyse the composition of the genome in terms of repeat types and

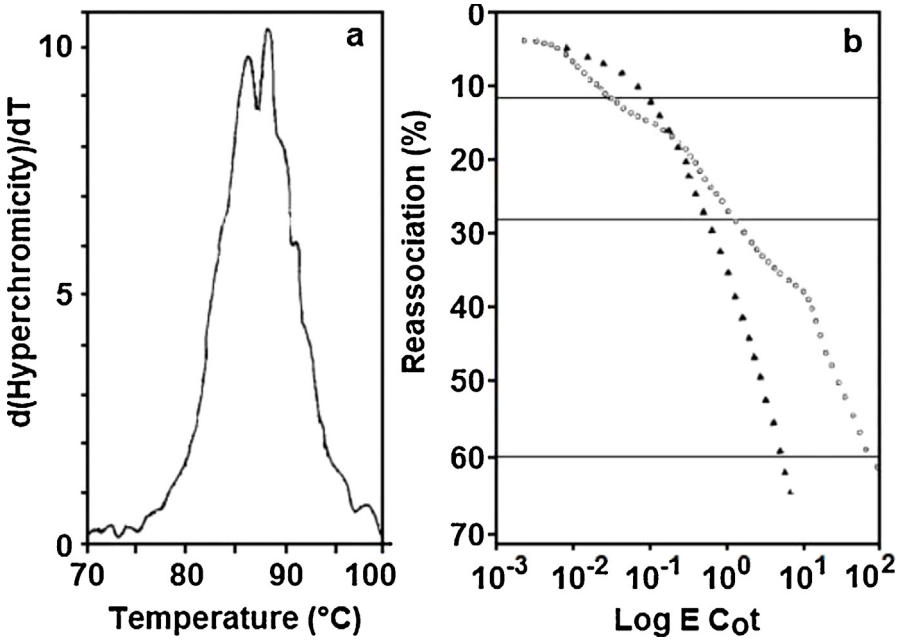

Fig. 1. (a) First derivative curves of the melting profile of sunflower genomic DNA. Small shoulders indicate specific A-T or G-C rich families of repeats. (b) Reassociation kinetics of the same DNA (circles) and of DNA of *E. coli* (triangles). Horizontal lines separate groups of sunflower sequences according to their redundancy. Redrawn from [29].

Please cite this article in press as: T. Giordani, et al., The repetitive component of the sunflower genome, Curr. Plant Biol. (2014), http://dx.doi.org/10.1016/j.cpb.2014.05.001