Molecular Phylogenetics and Evolution 51 (2009) 515–530

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Tracking the evolutionary history of polyploidy in *Fragaria* L. (strawberry): New insights from phylogenetic analyses of low-copy nuclear genes

M. Rousseau-Gueutin [a], A. Gaston [a], A. Aïnouche [b], M.L. Aïnouche [b], K. Olbricht [c,1], G. Staudt [d], L. Richard [a], B. Denoyes-Rothan [a,*]

[a] INRA, UR 419, Unité de Recherche sur les Espèces Fruitières, Domaine de la Grande Ferrade, 71 Avenue Edouard Bourlaux – BP 81, F-33883 Villenave d'Ornon, France
[b] CNRS, UMR 6553 Ecobio, Université de Rennes-1, Bât 14A. Campus scientifique de Beaulieu, 35042 Rennes Cedex, France
[c] JKI – Federal Research Centre for Cultivated Plants, Institute for Breeding Research on Horticultural and Fruit Crops, Pillnitzer Platz 3a, D-01326 Dresden, Germany
[d] Bächelhurst 10A, D- 79249 Merzhausen, Germany

ARTICLE INFO

Article history:
Received 17 September 2008
Revised 23 December 2008
Accepted 29 December 2008
Available online 4 January 2009

Keywords:
Fragaria
Polyploidy
Reticulate evolution
Low-copy nuclear genes
GBSSI
DHAR
Duplicated gene

ABSTRACT

Phylogenetic utility of two nuclear genes (*GBSSI*-2 and *DHAR*) was explored in genus *Fragaria* in order to clarify phylogenetic relationships among taxa and to elucidate the origin of the polyploid species. Orthology of the amplified products was assessed by several methods. Our results strongly suggest the loss of one *GBSSI* duplicated copy (*GBSSI*-1) in the *Fragariinae* subtribe. Phylogenetic analyses provided new insights into the evolutionary history of *Fragaria*, such as evidence supporting the presence of three main diploid genomic pools in the genus and demonstrating the occurrence of independent events of polyploidisation. In addition, the data provide evidence supporting an allopolyploid origin of the hexaploid *F. moschata*, and the octoploids *F. chiloensis*, *F. iturupensis* and *F. virginiana*. Accordingly, a new pattern summarizing our present knowledge on the *Fragaria* evolutionary history is proposed. Additionally, sequence analyses also revealed relaxed constraints on homoeologous copies at high ploidy level, as demonstrated by deletion events within *DHAR* coding sequences of some allo-octoploid haplotypes.

1. Introduction

Polyploidy or whole genome duplication has long been recognized to be a major force in evolution (Ohno, 1970). This process has been observed in plants, invertebrates (Adamowicz et al., 2002; Curole and Hedgecock, 2005), lower vertebrates (Leggatt and Iwama, 2003; Holloway et al., 2006) and even humans (Gu et al., 2002; McLysaght et al., 2002; Spring, 2002; Panopoulou et al., 2003). Polyploidy is especially prevalent in plants, where it is known as a major speciation process that has recurrently affected various lineages (Otto and Whitton, 2000; Cui et al., 2006). In addition to the plethora of examples represented by polyploid complexes, all the sequenced plant genomes previously considered as "diploids" (e.g. *Arabidopsis*, rice, poplar or grape vine) have revealed superimposed traces of past genome duplication events (The Arabidopsis Genome Initiative, 2000; Bowers et al., 2003; Paterson et al., 2004; Tuskan et al., 2006; The French-Italian Public Consortium for Grapevine Genome Characterization, 2007).

The success of polyploidy is mainly attributed to genetic and genomic causes involving higher genetic diversity and fixed heterozygosity at duplicated loci (Wendel, 2000; Comai, 2005), structural dynamics (Pontes et al., 2004; Chantret et al., 2005; Grover et al., 2007) and novel patterns of gene regulation and expression (Adams et al., 2003; Osborn et al.; 2003; Salmon et al, 2005; Albertin et al., 2006; Keyte et al., 2006; Tate et al., 2006; Akhunov et al., 2007; Flagel et al., 2008). Polyploidy has important phenotypic consequences (Comai et al., 2000; Schranz and Osborn, 2004; Gaeta et al., 2007) that allow new polyploid species responding to natural (Brochmann et al., 2004; Soltis et al., 2004; Aïnouche et al., in press) and artificial (Paterson, 2005; Udall and Wendel, 2006) selection challenges.

Reconstructing the evolutionary history of plant polyploid species using molecular data has to deal with the presence and the evolutionary fate of multiple genes copies resulting from both individual gene duplication within a genome (*i.e.* paralogous) and whole-genome duplication in nuclear genomes (*i.e.* homoeologous). Identification of homoeologues in polyploids is critical for reliable phylogeny reconstruction, identification of parental lineages and mode of polyploid formation (auto versus allopolyploid) (Doyle et al., 2004; Popp and Oxelman, 2007; Fortune et al., 2008). In this context of polyploidy, the use of low-copy nuclear genes is

* Corresponding author. Fax: +33 5 57 12 24 39.
E-mail address: denoyes@bordeaux.inra.fr (B. Denoyes-Rothan).
[1] Present address: Hansabred Radeburger Landstrasse 12, D-01108 Dresden, Germany.

1055-7903/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2008.12.024

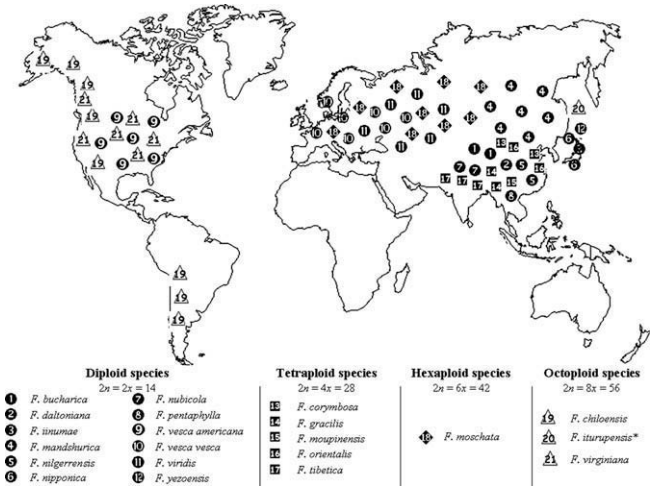

Fig. 1. Geographic repartition of *Fragaria* species and sub-species used in this study. The asterisk symbol indicates that two ploidy levels (octoploid and decaploid) have been reported for *F. iturupensis*.

of particular interest for reconstructing reticulate evolution since they are biparentally inherited (Small et al., 2004). Assuming retention of the duplicated copies, the sequences duplicated by polyploidy should be each other's closest relatives in an autopolyploid species or be distributed in different clades in an allopolyploid species (Wendel and Doyle, 1998; Fortune et al., 2007). Another significant advantage of low-copy nuclear genes is their elevated rate of sequence evolution relative to organellar genes (Small et al. 2004), which improve the resolution of relationships among low taxonomic levels (Álvarez and Wendel, 2003; Small et al., 1998, 2004; Ghiselli et al., 2007; Steele et al., 2008). In addition, these low-copy nuclear genes are less prone to concerted evolution than ribosomal DNA (rDNA) and are less homoplasious (Sang, 2002; Senchina et al., 2003; Lihová et al., 2006; Popp and Oxelman, 2007). However, these markers are still poorly exploited because of the difficulties in isolating and characterizing orthologous loci (Doyle et al., 2003; Small et al., 2004).

The genus *Fragaria* (Rosaceae) is comprised of a limited number of wild species (about 22, Staudt, 2008) ranging from diploids ($2n = 2x = 14$) to decaploids ($2n = 10x = 70$). An additional taxon is the main cultivated strawberry crop, *F.* x *ananassa* ($8x$), which originated following accidental hybridization between two New World octoploid strawberry species, *F. chiloensis* and *F. virginiana*. The wild species of *Fragaria* are distributed throughout the Holarctic zone with a few endemic zones into the tropics. All species are restricted to single continents or specific areas (Fig. 1), except the diploid species *F. vesca*, which is found in both Eurasia and America (Staudt, 1962, 1989). Ten diploid species are distributed in Eastern, South Eastern and Central Asia and the three others in Europe and Western Siberia (Staudt, 1962, 1989, 1999, 2003, 2005, 2006, 2008; Staudt and Dickoré, 2001; Staudt and Olbricht, in press). Five tetraploid species ($2n = 4x = 28$) are restricted to Eastern and South Eastern Asia (Staudt and Dickoré, 2001; Staudt, 2003), while the hexaploid species ($2n = 6x = 42$) occurs in Europe and Western Siberia (Staudt, 1962, 1989, 2008). Concerning the three octoploid species, one, *F. iturupensis*, is endemic to the Far East Asia (Southern Kurisls) (Staudt, 1973, 1989; Staudt and Olbricht, in press) ($2n = 8x = 56$), and the two others, *F. chiloensis and F. virginiana*, are present in America with different distributions in South and North America (Staudt, 1962, 1989). Recently, a decaploid number ($2n = 10x = 70$) has been reported for three accessions of *F. iturupensis* originating from Mount Atsunupuri (Hummer et al., in press). Only two phylogenetic analyses have been performed to date in *Fragaria* based on chloroplast (cp) DNA RFLPs (Harrison et al., 1997a) or combined sequence data sets from the cpDNA *trnL-trnF* region and the Internal Transcribed Spacer (ITS) region of the nuclear ribosomal DNA (rDNA) (Potter et al., 2000). However, due to a limited taxon sampling, and to a low level of sequence variation of these regions, few insights were provided with regard to the phylogenetic relationships among most *Fragaria* species.

In this paper, we explore the utility of two low-copy nuclear genes, *GBSSI*-2 (or *Waxy*) and *DHAR* (DeHydro Ascorbate Reductase), to elucidate the evolutionary history of genus *Fragaria* (Rosaceae) and particularly the putative parental lineages of the polyploid species. The *GBSSI* gene has provided high phylogenetic resolution in several plant systems including polyploids (Ingram and Doyle, 2003; Smedmark et al., 2003; Mason-Gamer, 2001, 2004; Winkworth and Donoghue, 2004; Fortune et al., 2007, 2008). In the Rosaceae family, the *GBSSI* gene is present in two copies, *GBSSI*-1 and *GBSSI*-2, which can be readily distinguished by specific indels (Evans et al., 2000; Evans and Campbell, 2002). The *DHAR* gene is known to be present in a few copies in flowering plants: one copy in *Oryza sativa* (Urano et al., 2000) and spinach (Shimaoka et al., 2000), two copies in tomato (Zou et al., 2006), and three active copies and a pseudogene in *Arabidopsis thaliana* (Dixon et al., 2002). In this latter species, the three active copies include two cytosolic copies, which are similar in sequence and structure, but differ from the third one (chloroplastic copy) by the number of introns. This gene is studied here for the first time in the Rosaceae family. The use of two independent nuclear loci and nearly all the wild *Fragaria* species increased the relevance and the confidence of our phylogenetic inferences and provided new insights on polyploid origins.

2. Material and methods

2.1. Plant material

Twenty wild species belonging to the *Fragaria* genus are represented in our sampling (Table 1). It includes 11 diploid species ($2n = 2x = 14$), five tetraploid species ($2n = 4x = 28$), one hexaploid species ($2n = 6x = 42$) and three octoploid species ($2n = 8x = 56$). For the widely distributed *F. vesca*, two subspecies (from North America and Asia) were included (Fig. 1). Additionally, most of the *Fragaria* diploid species were each represented by two individuals. In total, 30 samples belonging to twenty wild *Fragaria* species were used in this study. Outgroup species were chosen based on their close relationships with the *Fragaria* genus (Harrison et al. 1997a; Potter et al., 2000, 2002, 2007; Eriksson et al., 2003). They included *Potentilla andicola*, *P. megalantha, P. nepalensis* and *P. recta*.

2.2. DNA isolation

Genomic DNA was isolated from 90 mg of fresh leaves from each individual using the DNeasy Plant Mini kit (Qiagen), following instructions provided by the manufacturer. When no amplification was observed after PCR, an additional step of purification using Wizard® SV Gel and PCR Clean-Up System (Promega) was performed.

2.3. Primer pairs design and PCR amplification

The oligonucleotide primers were developed based on selected *Fragaria* EST (Expressed Sequence Tag) contigs with high homology to *GBSSI* and *DHAR* proteins available in GenBank. For each gene, two oligonucleotides were designed in the coding regions surrounding introns (Primer3 program), which were inferred firstly after alignment of the contigs with the corresponding genes of *A.*