Molecular Phylogenetics and Evolution 66 (2013) 17–29

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

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

Insights into phylogeny, sex function and age of *Fragaria* based on whole chloroplast genome sequencing

Wambui Njuguna [a,*], Aaron Liston [b], Richard Cronn [c], Tia-Lynn Ashman [d], Nahla Bassil [e]

[a] Department of Horticulture, Oregon State University, ALS 4017, Corvallis, OR 97331, USA
[b] Department of Botany and Plant Pathology, Oregon State University, Cordley Hall 2082, Corvallis, OR 97331, USA
[c] Pacific Northwest Research Station, United States Forest Service, 3200 SW Jefferson Way, Corvallis, OR 97331, USA
[d] Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260, USA
[e] United States Department of Agriculture, Agricultural Research Service, National Clonal Germplasm Repository, Corvallis, OR 97333, USA

ARTICLE INFO

Article history:
Received 13 February 2012
Revised 20 August 2012
Accepted 30 August 2012
Available online 14 September 2012

Keywords:
Plastome
Polyploidy
Self-compatibility

ABSTRACT

The cultivated strawberry is one of the youngest domesticated plants, developed in France in the 1700s from chance hybridization between two western hemisphere octoploid species. However, little is known about the evolution of the species that gave rise to this important fruit crop. Phylogenetic analysis of chloroplast genome sequences of 21 *Fragaria* species and subspecies resolves the western North American diploid *F. vesca* subsp. *bracteata* as sister to the clade of octoploid/decaploid species. No extant tetraploids or hexaploids are directly involved in the maternal ancestry of the octoploids.

There is strong geographic segregation of chloroplast haplotypes in subsp. *bracteata*, and the gynodioecious Pacific Coast populations are implicated as both the maternal lineage and the source of male-sterility in the octoploid strawberries. Analysis of sexual system evolution in *Fragaria* provides evidence that the loss of male and female function can follow polyploidization, but does not seem to be associated with loss of self-incompatibility following genome doubling. Character-state mapping provided insight into sexual system evolution and its association with loss of self-incompatibility and genome doubling/merger. *Fragaria* attained its circumboreal and amphitropical distribution within the past one to four million years and the rise of the octoploid clade is dated at 0.372–2.05 million years ago.

1. Introduction

The domestication of strawberry (*Fragaria* ×*ananassa* subsp. *ananassa*) is well documented historically. Female plants of *F. chiloensis* from Chile were brought to France in 1716 by a French army officer, Amédée Frézier. He gave one of the plants to Antoine de Jussieu, the director of the King's garden in Paris, where clones of the eastern North American dioecious *F. virginiana* imported from eastern N. America in the early to mid 17th century were growing. Cultivated individuals derived from the interspecific hybrids of these two species (*F. chiloensis*, *F. virginiana*) in the early 18th century created the modern strawberry grown today (Hancock, 1999). Beyond these recent historical events, however, there is far less known about the evolution of the two species that contributed to this important fruit crop.

As is common in domesticated plants, strawberry is a polyploid (octoploid; $2n = 8x = 56$), and the genus *Fragaria* contains extensive natural ploidy variation with species known at five ploidy levels (diploid through decaploid). In addition to differences in ploidy, *Fragaria* contains a wide range of diversity in sexual systems (Staudt, 1989). Two-thirds of the 27 recognized taxa (species and subspecies) are hermaphroditic and either self-compatible (SC) (10 species) or self-incompatible (SI) (seven species). Ten species show some degree of sexual polymorphism, including: gynodioecy (females and hermaphrodites) in *F. vesca* subsp. *bracteata*; subdioecy (females, hermaphrodites and males) in *F. chiloensis*, *F. virginiana*, and their naturally occurring hybrid *F.* ×*ananassa* subsp. *cuneifolia*; and dioecy (females and males) in *F. corymbosa*, *F. gracilis*, *F. moschata*, *F. moupinensis*, *F. orientalis and F. tibetica* (Staudt, 2009). The diversity in chromosome number and mating systems makes *Fragaria* an exceptional system for understanding sexual system evolution, especially in light of recent genetic mapping that uncovered proto-sex chromosomes and sex chromosome turnover in the genus (Goldberg et al., 2010; Spigler et al., 2008).

The diploid ancestry of the octoploid parents of the cultivated strawberry, *F.* ×*ananassa* subsp. *ananassa,* has been studied through the use of interspecific hybridization and cytogenetic methods (Bringhurst, 1990; Federova, 1946; Senanayake and Bringhurst, 1967). These studies primarily implicated the widespread species *F. vesca* and its close relatives as potential diploid ancestors to these

* Corresponding author. Present address: Eurofins Lancaster Laboratories, 2425 New Holland Pike, Lancaster, PA 17605, USA. Fax: +1 541 738 4205.
E-mail addresses: WYoung@lancasterlabs.com (W. Njuguna), listona@science.oregonstate.edu (A. Liston), rcronn@fs.fed.us (R. Cronn), tia1@pitt.edu (T.-L. Ashman), nahla.bassil@ars.usda.gov (N. Bassil).

1055-7903/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ympev.2012.08.026

lineages. Molecular phylogenetic studies based on chloroplast DNA and nuclear ribosomal ITS sequences (Harrison et al., 1997; Potter et al., 2000) identified a clade of *F. vesca* and related species as the likely maternal ancestral lineage contributing to the octoploid cytoplasm. The first published study to use low-copy nuclear loci (Rousseau-Gueutin et al., 2009) also supported the *F. vesca* clade as a diploid ancestor, and further identified a Japanese species, *F. iinumae*, as contributing to the genome composition of the octoploids. Despite these advances, several important questions remain unresolved: (1) was there a single origin of the octoploids?; (2) can a single extant member of the *F. vesca* clade be identified as a direct ancestor of the octoploids?; (3) what is the likely geographic location of the octoploid origin? and (4) how long ago did this divergence take place?

Furthermore, given the diversity of sexual systems in strawberry, phylogenetic analysis can test conflicting hypotheses for the relationship between sexual system and whole-genome duplication (Baker, 1984; Jennings, 1976; Brunet and Liston, 2001; Osborn et al., 2003; Miller and Venable, 2000). For instance, we can determine whether loss of self-incompatibility with increase in ploidy facilitates the evolution of separate sexes (dioecy) (Miller and Venable, 2000) or if other mechanisms like the acquisition of male sterility or female sterility mutations (Brunet and Liston, 2001) are involved.

Phylogenetic analysis of closely related taxa using nearly-complete chloroplast genomes (Parks et al., 2009; Straub et al., 2012) can provide unprecedented insights into phylogenetic relationships and biogeographic history, as well as allow the first explicit tests of hypotheses for sexual system evolution in the genus *Fragaria*. We therefore sequenced nearly complete chloroplast genomes from 21 *Fragaria* species to assess the phylogeny, biogeography and sexual system evolution.

2. Materials and methods

2.1. Plant material

Twenty-five accessions representing 21wild *Fragaria* species, subspecies and hybrids, and one accession of *Potentilla*, a close relative of *Fragaria* in Rosaceae (Eriksson et al., 2003; Lundberg et al., 2009; Dobeš and Paule, 2010), were included in the study (Table 1).

2.2. DNA extraction and PCR

DNA was extracted from actively-growing leaves using a protocol based on the PUREGENE® kit (Gentra Systems Inc., Minneapolis, MN). Preparations of chloroplast DNA for sequencing were obtained from genomic and chloroplast PCR fragment pools (Table 1). To generate PCR fragment pools, 203 chloroplast primers (108 forward, 95 reverse) were screened in various combinations in four species (*F. iinumae*, *F. nipponica*, *F. orientalis*, and *F. virginiana*) to identify pairs that amplified fragments that were at least 2.5 kb in size, and to provide maximum coverage of the chloroplast genome. Where possible, primer pairs that amplified single bands in most or all of the species were chosen. Of these 203 primers, 141 were previously reported to amplify the *Cucumis sativus* L. chloroplast genome (Chung et al., 2007); 25 were designed from the genome sequence of *Morus indica* M. alba 'K2' (Ravi et al., 2006) and 36 primers were designed in this study from *F. vesca* 'Hawaii 4' (Shulaev et al., 2010). Sixty-three primer pairs (Njuguna, 2010) were finally chosen to amplify the entire chloroplast genomes of 17 accessions (Table 1). We used Phusion™ High-Fidelity DNA polymerase (New England Biolabs, Ipswich, MA) for long-range PCR. Amplifications were performed in 10 μl reactions containing 1× Phusion GC buffer, 2.5 mM of each dNTP, 10 μM of each primer, 5U of Phusion DNA polymerase, 0.05 μl of 3% DMSO and 5 ng of DNA template. PCR product quantification was carried out using the Quant-iT™ PicoGreen® dsDNA quantification protocol (Molecular Probes, Inc., Eugene, OR) following the manufacturer's specifications. Equimolar amounts of PCR products were pooled for each species to generate 1–5 μg of chloroplast DNA for Illumina sample preparations.

2.3. Illumina library preparation

For 11 accessions, we directly assembled the plastome from genomic sequencing (reviewed in Straub et al., 2011). For both genomic DNA and that obtained from chloroplast PCR fragment pools, sequencing libraries (Table 1) were prepared for sequencing using the sample preparation kit from Illumina (Illumina Inc., San Diego, CA) and as described by Cronn et al. (2008). Briefly, this approach utilizes Illumina sequencing technology to sequence multiple barcoded PCR amplified chloroplast genomes in one lane of a flow cell. Multiplexing of small organellar genomes in a single lane utilizes the high sequencing capacity of this platform (>40 million clusters per flowcell during this study). For details on the Illumina sequencing runs performed, refer to Table 1.

2.4. Sanger sequencing of SNPs

DNA sequences encompassing loci containing three parsimony-informative octoploid and decaploid specific SNPs were PCR amplified and sequenced using Sanger sequencing methods. Primers that can amplify three genes (*ndh*F, *ccs*A, and *rpo*C2) containing parsimony-informative SNPs between *F. vesca* subsp. *bracteata* and the octoploid clade were designed with Primer 3 (Rozen and Skaletsky, 2000). They included ndhF-715F/ndhF-715R (5′- GTAAAAGGTTTATGGACGGAGTT-3′, 5′-GCATTGTTGTTTTTAGGATCTGG-3′); ccsA-731F/ccsA-731R (5′-CCTTTGGTGAGATTCAATACGTG-3′, 5′-GACAAGGCCGAAGCTATTCTATC-3′), and rpoC2F/rpoC2R (5′-GGAATTCGAAATTCTCCCGTTT-3′, 5′- AGGGATAATCTAGAGCTTCGAGTTG-3′), respectively. PCR was carried out as described above, followed by Exonuclease-Shrimp Alkaline Phosphatase (ExoSAP) cleanup. This procedure involved mixing 4 μl of PCR product with an 8.1 μl mixture of 2 μl shrimp alkaline phosphatase (SAP; 1 unit/μl), 0.1 μl exonuclease I (Exo I; 20 units/μl) and 6 μl of water. The mixture was incubated at 37 °C for 60 min, followed by 72 °C for 15 min. The samples were submitted for sequencing at the Center for Genome Research and Biocomputing (CGRB) at Oregon State University in Corvallis, Oregon.

2.5. Data analysis

After the sequencing run, raw image data for each sequencing cycle was processed into base calls and alignment files through the Illumina Pipeline (version 0.2.2.6). Binning was carried out using the three nucleotide barcodes. After sorting microreads (36, 40 or 60 bp) into sample-specific bins, the barcodes and adapter tags were removed, and resulting microreads (32, 36 or 56 bp) were used for subsequent analysis. The 155,691 kb *F. vesca* 'Hawaii 4' annotated chloroplast genome (Genbank accession JF345175; Shulaev et al., 2010) was used for reference-guided microread assembly and was also included in the phylogenetic analysis. Microreads were assembled into contigs using YASRA (Ratan, 2009), and Mulan (Ovcharenko et al., 2005) was used to assemble and align contigs. Bioedit (Hall, 1999) was used to manually check and correct mis-alignments, remove primer sequences and score indels. Assembly errors resulting in high sequence divergence and insertions were recognized by low sequencing depth (<5*x*) data from the YASRA output, and were masked in the final assemblies. Calculation of variable and parsimony informative sites in