
S. N. Raina, S. Sharma, T. Sasakuma, M. Kishii, and S. Vaishnavi

From the Kihara Institute for Biological Research and Graduate School of Integrated Sciences, Yokohama City University, 641-12 Maioka-cho, Totsuka-ku, Yokohama 244-0813, Japan (Raina, Sasakuma, and Kishii); and Laboratory of Cellular and Molecular Cytogenetics, Department of Botany, University of Delhi, Delhi 110007, India (Sharma and Vaishnavi).

Address correspondence to S. N. Raina, Laboratory of Cellular and Molecular Cytogenetics, Department of Botany, University of Delhi, Delhi 110007, India, or e-mail: snraina@satyam.net.in or soomr@yahoo.com.

Two novel repetitive DNA sequences, pCtKpnI-1 and pCtKpnI-2, were isolated from Carthamus tinctorius (2n = 2x = 24) and cloned. Both represent tandemly repeated sequences. The pCtKpnI-1 and pCtKpnI-2 clones constitute repeat units of 343–345 bp and 367 bp, respectively, with 63% sequence heterogeneity between the two. Fluorescence in situ hybridization (FISH) was employed on metaphase chromosomes of C. tinctorius using, simultaneously, pCtKpnI-1 and pCtKpnI-2 repeated sequences. The pCtKpnI-1 sequence was found to be exclusively localized at subtelomeric regions on most of the chromosomes. On the other hand, sequence of the pCtKpnI-2 clone was distributed on two nucleolar and one nonnucleolar chromosome pairs. The satellite, and the intervening chromosome segment between the primary and secondary constrictions, in the two nucleolar chromosome pairs were wholly constituted by pCtKpnI-2 repeated sequence. The pCtKpnI-2 repeated sequence, showing partial homology to intergenic spacer (IGS) of 18S-25S ribosomal RNA genes of an Asteraceae taxon (Centaurea stoebe), and the 18S-25S rRNA gene clusters were located at independent, but juxtaposed sites in the nucleolar chromosomes. Variability in the number, size, and location of the two repeated sequences provided identification of most of the chromosomes in the otherwise not too distinctive homologues within the complement. This article reports the start of a molecular cytogenetics program targeting the genome of safflower, a major world oil crop about whose genetics very little is known.

The genus Carthamus L. of the tribe Cynareae, subfamily Tubuliflorae, and family Asteraceae (Compositae) comprises approximately 25 species and subspecies that are widely distributed from Spain and North Africa across the Middle East and Iran to northwestern India (Estilai 1977; Kumar 1991). Carthamus tinctorius (2n = 2x = 24), commonly known as safflower, is the only domesticated species. It is widely cultivated in various agricultural production systems in Asia, Europe, Australia, and the Americas as a source of high-quality vegetable and industrial oil, and as a feed for livestock (Knowles 1980).

The taxonomy of Carthamus is inconsistent. For example, on the basis of cytogenetical and crossability data, Carthamus ozyantha, Carthamus palustinus, and Carthamus flavescens, closely related to C. tinctorius, are considered by Ashri and Knowles (1960) to be true wild species and not forms of C. tinctorius that have escaped from cultivation. Baker (1970), on the other hand, regards them as wild form conspecies (“biological species”) of C. tinctorius. In comparison to other oil-yielding crop plants, safflower has been the focus of relatively very little research in assessing intra- or interspecies relationships based on various molecular tools. The information available so far is restricted to only the chloroplast genome in safflower (Ma and Smith 1985a,b; Smith and Ma 1985; Tippett et al. 1991). Thus the genomic composition of C. tinctorius remains unresolved and new approaches are required. Lack of a clear understanding of patterns of genomic constitution in C. tinctorius vis-à-vis the above wild taxa, possessing agronomically useful characters of importance, limits the effective utilization of wild germplasm resources for safflower breeding (Kumar 1991).

Eukaryotic genomes, including those of higher plants, contain highly repeated DNA sequences. They constitute
mostly the noncoding sequences of plant genomes and may in some cases account for more than 90% of a genome (Flavell 1980; Vershinin et al. 2001). These sequences can evolve rapidly in concerted fashion, so they are sometimes present throughout the genome and are often species specific (Flavell 1980; Linares et al. 1998; Zhang et al. 2002). The repetitive DNA sequences may influence chromosome structure and recombination events, and may be active in the process of genome differentiation and evolution (Calderini et al. 1997; Uozu et al. 1997; Yan et al. 2002). In view of the above, detailed information about repetitive DNA sequences, including their abundance and sequence divergence and their distribution on the chromosomes, is critical to gaining a complete understanding of any genome organization (Yan et al. 2002). These sequences include dispersed repetitive DNA elements and tandem repeats. Tandemly repeated DNA sequences are generally localized in certain specific regions of all or a few chromosomes, while dispersed repetitive DNA sequences are found scattered on most of the chromosomes (Kubis et al. 1998). The tandemly arranged repetitive DNA elements in particular may provide information as to the evolution of constituent genomes (Badaeva et al. 2002; Pich et al. 1996; Vershinin et al. 2001). Such sequences are also useful for molecular karyotyping of chromosomes, which otherwise cannot be identified by simple morphological features alone (Fuchs and Schubert 1998).

In an attempt to elucidate the genomic composition of safflower, we report here the isolation, cloning, and sequencing of novel repetitive DNA elements of *C. tinctorius* and their distribution on the chromosomes.

### Materials and Methods

#### Plant Material and DNA Isolation

The seeds of *C. tinctorius* L. were obtained from the U.S. Department of Agriculture (USDA), Beltsville, MD. The leaves were collected from 2-month-old plants grown under controlled conditions and stored at $-80^\circ$C. Total genomic DNA was isolated from leaves by modified CTAB method (Porchebski et al. 1997).

#### Isolation, Sequencing, and Sequence Analysis of Repeated DNA Elements

Total DNA, separately digested with 21 restriction enzymes (Amersham Pharmacia Biotech, Buckinghamshire, UK) for 16–18 h at 37°C, was fractionated by 1.0% agarose gel electrophoresis. The restriction profile of *Kpn*I digestion showed a ladder of bands corresponding to multimeric DNA fragments. Subsequently 60 μg of DNA was digested with *Kpn*I in order to clone these multimeric fragments. The basic and dimeric *Kpn*I repeat sequences were eluted, purified with GENECLEAN II (Qbiogene, Irvine, CA), and cloned into pUC19 vector in GENECLEAN II (Qbiogene, Irvine, CA), and cloned into *Escherichia coli* strain DH5α. Nucleotide sequences of six clones (2, 3, 4, 6, 8, and 12) containing the inserts were determined with an ABI PRISM 377 automated DNA sequencer (Applied Biosystems, Foster City, CA). The six nucleotide sequences submitted to the European Molecular Biology Laboratory (EMBL) Nucleotide Sequence Database bear the accession numbers AJ515526–AJ515531. The sequences were analyzed using BLAST and the software tools Clustal W and BOXSHADE.

### Southern Hybridization

For Southern blots, 5 μg of *Kpn*I-digested genomic DNA was fractionated on a 0.85% agarose gel and transferred to nylon membrane (Hybond-N⁺, Amersham Pharmacia Biotech). The monomeric *Kpn*I fragments, separated from plasmid vectors and gel purified by the spin-down method (Heery et al. 1990), were labeled with fluorescein-11-dUTP (Amersham Pharmacia Biotech). The genomic DNA fragments that hybridized to the probes were detected enzymatically according to manufacturer’s protocols.

### Fluorescence in situ Hybridization

The method used for fluorescence in situ hybridization (FISH) was carried out as described previously (Raina and Mukai 1999). Cloned fragments 12 (pCt*Kpn*I-1) and 2 (pCt*Kpn*I-2) and pTa71 (containing an 8.95 kb full-length repeat unit of the 18S-5.8S-26S ribosomal RNA gene of *Triticum aestivum*) used as probes were labeled with digoxigenin-11-dUTP (Roche Diagnostics GmbH, Mannheim, Germany), biotin-11-dUTP (Gibco BRL, Gaithersburg, MD), and a 1:2 mix of digoxigenin-11-dUTP and biotin-11-dUTP, respectively, by nick translation. The chromosomes were counterstained with 4’, 6-diamidino-2-phenylindole (DAPI). The slides were examined with a fluorescence microscope (Olympus). Photographs were taken with a cooled charged-coupled device (CCD) camera (Photometrics, Huntington Beach, CA) and analyzed with IPLab Spectrum computer software (Signal Analytics, Vienna, VA). In the karyogram constructed, chromosome pairs designated from I to XII were ordered from longest to shortest, based on relative length of each pair.

### Results and Discussion

The genomic DNA of *C. tinctorius* digested with *Kpn*I showed clear electrophoretic bands at 345, 690, 1035, and 1380 bp (Figure 1A), indicating the presence of tandemly arranged repetitive elements in the genome. The 345 and 690 bp fragments were cloned. Clones 3, 4, 6, and 8 contained the monomer element, and clone 12 contained the dimer element with the same repeat unit length of 343–345 bp, having about 48% GC content. The homology among these clones was 96–98%. This repeat sequence was called pCt*Kpn*I-1. Clone 2, however, turned out to be different. It was 367 bp in length with 55% GC content. It showed only 37% sequence homology with the five clones of pCt*Kpn*I-1. This repeat sequence was called pCt*Kpn*I-2. Hybridization of pCt*Kpn*I-1 and pCt*Kpn*I-2 repeated sequences to *Kpn*I-digested genomic DNA of *C. tinctorius* gave a ladder-like pattern of DNA fragments typical of tandem repetitive sequences (Figure 1B,C).
The polymeric bands, even after complete digestion, may originate by modification of the restriction enzyme site due to sequence divergence and/or methylation. In *T. aestivum*, for example, one cloned *EcoR* I element (p551) contained seven repeated units of 320 bp each, but only two internal recognizable *EcoR* I sites (Ueng et al. 2000).

No sequence homologous to pCt*Kpn*I-1 could be retrieved from the databases available online. Interestingly, pCt*Kpn*I-2 (from nucleotide position 246 to 367) showed 88% homology to the partially cloned intergenic spacer (IGS) region of the 18S-25S rRNA gene family of *Centaurea stoebe* (Asteraceae), including a putative transcription initiation site (TIS), TATATATAGGGGG (Figure 2). The IGS regions from several plant taxa have been shown to carry promoter or potential TIS sequences (Macas et al. 2003; Stupar et al. 2002). The homology of the pCt*Kpn*I-2 satellite repeat with the complete IGS sequence of *C. stoebe* or IGS sequence of *C. tinctorius* could not be determined because of the lack of sequence data from these species.

Double-target in situ hybridization using the pCt*Kpn*I-1 and pCt*Kpn*I-2 satellite repeats as probes revealed that pCt*Kpn*I-1 signals of very minute to very high intensity were always distributed at subtelomeric sites on most of the *C. tinctorius* chromosomes (Figure 3A,C). Up to 38 signals at one or both subtelomeric ends in 22 of 24 chromosomes could be seen in most metaphases. Several types of tandemly repeated DNA families have been found to be mainly located at the subtelomeric regions, sometimes corresponding to the C-band regions of the chromosomes (Kishii et al. 1999; Pich et al. 1996; Vershinin et al. 1996; Zhang et al. 2002). pCt*Kpn*I-1 is yet another type of tandemly repeated DNA element, elucidative of the novel molecular nature of safflower chromosomes in the subtelomeric regions. It will be interesting to determine whether this sequence is universal in Asteraceae or specific for *Carthamus* genomes.

In comparison, only six chromosomes, including the two nucleolar chromosome pairs, contained six unique intercalary and distal hybridization sites for pCt*Kpn*I-2 repeated sequence (Figure 3A). Four large signals identifying the entire intervening segment between the primary and secondary constriction in one and the satellite in the other pair of nucleolar chromosomes was the most common and characteristic observation in all the metaphases (Figure 3A,C). Extended minute signals corresponding to the nucleolus-organizing regions (NORs) were also observed in a few metaphases (Figure 3C). The two remaining minute signals were observed to occur at the centromeric region of one of the nonnucleolar chromosome pairs (Figure 3A,B). From the hybridization pattern of the signals obtained for the two repetitive sequences, several chromosome pairs could be clearly distinguished from each other. The nucleolar chromosome pairs I and II, for example, could be distinguished

---

**Figure 1.** (A) Multimeric electrophoretic bands of *C. tinctorius* that appeared after digestion with *Kpn*I. (B) Southern hybridization of genomic DNA in *C. tinctorius* digested with *Kpn*I and hybridized with pCt*Kpn*I-1, and (C) pCt*Kpn*I-2 repetitive sequences.

**Figure 2.** Alignment of the pCt*Kpn*I-2 satellite sequence of *C. tinctorius* with the partial rDNA IGS sequence of *C. stoebe*. Identical nucleotides and a potential transcription initiation site are indicated by gray and black backgrounds, respectively.
Figure 3. (A) Fluorescence in situ hybridization of *C. tinctorius* (*2n* = 24) metaphase chromosomes with pCt*Kpn*I-1 (red fluorescence) and pCt*Kpn*I-2 (green fluorescence) probes. (B) Metaphase chromosomes after FISH with pCt*Kpn*I-1, pCt*Kpn*I-2, and 18S-26S rDNA (yellowish green fluorescence) probes. (C) The chromosomes were counterstained with DAPI. Karyograms of metaphase chromosomes after DAPI staining (upper row) and FISH with pCt*Kpn*I-1 and pCt*Kpn*I-2 repeats (lower row). The bar represents 10 μm.
from each other by the labeling pattern of pCtKpnI-2 and absence of the site for pCtKpnI-1 in chromosome pair 1 (Figure 3C).

The partial homology of the pCtKpnI-2 sequence with IGS sequence of C. stoebe and the occurrence of very strong signals of this sequence in the vicinity of NORs, harboring the 18S-25S rRNA genes, in the two nucleolar chromosome pairs of C. tinctorius prompted us to obtain information with regard to the location of the pCtKpnI-2 clone vis-à-vis rRNA gene clusters. In situ hybridization with heterologous 18S-26S wheat rDNA and pCtKpnI-2 probes distinguished independent, but characteristic juxtaposed locations for these repeated sequences at the expanded NOR, and the intervening segment between the primary and secondary constrictions in one and the satellite in the other nucleolar chromosome pair, respectively (Figure 3B). The juxtaposed signals for both probes were much larger than the signals obtained with either pCtKpnI-2 (Figure 3A,B) or pTa71 clone singly. The pCtKpnI-1 (approximately 37% homology with pCtKpnI-2) sites remained unchanged (Figure 3B).

Highly amplified satellite repeats with sequence homology to IGS subrepeats (termed rDNA orphans) have been mapped at nonspecific chromosome sites independent of rRNA gene clusters in several legume species (see Macas et al. 2003), while in potato, a tandem repeat (2D8) that is highly homologous to IGS sequence of the same species has been found to exclusively occur in its pericentromeric heterochromatin (Stupar et al. 2002). Here we report another type of rDNA orphan (pCtKpnI-2) mainly located in the nucleolar chromosomes, specifically in the vicinity of rRNA gene clusters. In the absence of sequence data of the IGS region and additional related pCtKpnI-2 sequences of safflower, it is difficult to interpret the present results in resolving the controversy with regard to the origin of rDNA orphans. Whether these are derived from the IGS repeats or vice versa, or both the repeat elements are derived from an independent common ancestor remains to be determined (see Macas et al. 2003; Stagninnus et al. 1999; Stupar et al. 2002).

Our investigation is the first to not only isolate repeated DNA sequences from any member of the third largest angiosperm family (Asteraceae), but is the first to apply the FISH technique to Carthamus as well. The existence of such sequences in other members of Asteraceae, including other Carthamus species, is not known, but the preliminary results obtained so far (unpublished data) suggest that such investigations will be of particular use in the study of phyllogenetic and evolutionary pathways not only in Carthamus taxa, but family Asteraceae as well.

Acknowledgments

Grateful thanks are given to the Japan Society for Promotion of Science (JSPS) and the Indian National Science Academy (INSA) for financial assistance (to S.N.R.). This work was supported, in part, by the Department of Biotechnology, Ministry of Science and Technology, India. The authors thank the U.S. Department of Agriculture for supplying the seed sample. We thank Prof. H. Tsujimoto, Faculty of Agriculture, Tottori University, Tottori, Japan, for providing helpful suggestions.

References


Received February 10, 2004
Accepted September 20, 2004

Corresponding Editor: Prem Jauhar