2006

Genomic Resources for Asparagales

Michael J. Havey  
*University of Wisconsin-Madison*

Kenneth C. Sink  
*Michigan State University*

Maria Jenderek  
*US Department of Agriculture*

Christopher D. Town  
*The Institute for Genomic Research*

Follow this and additional works at: [https://scholarship.claremont.edu/aliso](https://scholarship.claremont.edu/aliso)

Part of the Botany Commons

**Recommended Citation**  
GENOMIC RESOURCES FOR ASPARAGALES

MICHAEL J. HAVEY,1 KENNETH C. SINK,2 MARIA JENDEREK,3 AND CHRISTOPHER D. TOWN4

1Agricultural Research Service, US Department of Agriculture, Vegetable Crops Unit, Department of Horticulture, 1575 Linden Drive, University of Wisconsin, Madison, Wisconsin 53706, USA; 2Department of Horticulture, Michigan State University, East Lansing, Michigan 48824, USA; 3Agricultural Research Service, US Department of Agriculture, National Arid Land Plant Genetic Resources Unit, 9611 South Riverbend, Parlier, California 93648, USA; 4The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, Maryland 20850, USA

ABSTRACT

Enormous genomic resources have been developed for plants in the monocot order Poales; however, it is not known how useful these resources will be for other economically important monocots. Asparagales are a monophyletic order sister to class Commelinanae that carries Poales, and is the second most economically important monocot order. Development of genomic resources for and their application to Asparagales are challenging because of huge nuclear genomes and the relatively long generation times required to develop segregating families. We synthesized a normalized cDNA library of onion (Allium cepa) and produced 11,008 unique expressed sequence tags (ESTs) for comparative genomic analyses of Asparagales and Poales. Alignments of onion ESTs, Poales ESTs, and genomic sequences from rice were used to design oligonucleotide primers amplifying genomic regions from asparagus, garlic, and onion. Sequence analyses of these genomic regions revealed microsatellites, insertions/deletions, and single nucleotide polymorphisms for comparative mapping of rice and Asparagales vegetables. Initial mapping revealed no obvious synteny at the recombinational level between onion and rice, indicating that genomic resources developed for Poales may not be applicable to the monocots as a whole. Genomic analyses of Asparagales would greatly benefit from EST sequencing and deep-coverage, large-insert genomic libraries of representative small-genome model species within the “higher” and “lower” Asparagales, such as asparagus and orchid, respectively.

Key words: Asparagales, bacterial artificial chromosomes, cDNA, onion.

INTRODUCTION

Class Commelinanae and order Asparagales are two major monophyletic groups within the monocots (Chase et al. 1995; Rudall et al. 1997). Phylogenetic estimates based on chloroplast-gene sequences revealed that the commelinoid monocots are sister to Asparagales, and that these groups together are sister to Liliales (Chase et al. 2000; Fay et al. 2000). The most economically important monocots are in Commelinanae, order Poales. The second most economically important monocot order is Asparagales, which includes such valuable plants as agave, aloe, asparagus, garlic, leek, onion, and vanilla. The “higher” Asparagales form a well-defined clade within Asparagales (Chase et al. 1995, 1996) and include Alliaceae (chive, garlic, leek, and onion), Amar­yllidaceae (various ornamentals and yucca), and Asparagaceae (asparagus).

The development of genomic resources for Asparagales is challenging due to huge nuclear genomes (Fig. 1), relatively long generation times, and few financial resources. Onion, the most economically important member of Asparagales, is a diploid (2n = 2x = 16) with one of the largest nuclear genomes among all plants (Fig. 1) at 15,290 megabase-pairs (Mbp) (Arumuganathan and Earle 1991) of DNA per haploid (1C) nucleus (6, 16, and 107-times greater than maize, tomato, and Arabidopsis thaliana (L.) Heynh., respectively). In fact, diploid onion contains as much DNA as hexaploid wheat and on average each onion chromosome carries an amount of DNA equal to 75% of the 1C content of the maize nuclear genome (Bennett and Smith 1976). Molecular studies have revealed that the GC content of onion DNA is 32%, the lowest known for any angiosperm (Kirk et al. 1970; Stack and Comings 1979; Matsasa et al. 1989). Density-gradient centrifugation revealed no significant satellite DNA bands, except for a 375-bp telomeric sequence (Barnes et al. 1985). Ty1copia-like retrotransposons are present throughout the bulb-onion genome and concentrated in terminal heterochromatic regions (Leeton and Smyth 1993; Pearce et al. 1996). Stack and Comings (1979) used C0 reassociation kinetics to demonstrate that the onion genome consists of middle-repetitive sequences occurring in short-period interspersions among single-copy regions (Stack and Comings 1979). The huge nuclear genome of onion is not likely due to a recent polyploidization event. Jones and Rees (1968) and Ranjekar et al. (1978) proposed that intrachromosomal duplications contributed to increased chromosome sizes in onion. The large nuclear genome of onion has no effect on the number of markers required to produce detailed genetic maps, which have been developed using intra- (King et al. 1998a) and interspecific (van Heusden et al. 2000) crosses.

Garlic, the second most economically important Asparagales, is a diploid (2n = 2x = 16) plant and has a nuclear...
genome about 7% smaller than onion (Ori et al. 1998). Because of its obligate vegetative production, no genetic studies have been reported for garlic. Occasionally, garlic plants can be induced to produce flower stalks with a few fertile flowers. Recently, true seed production has been realized in garlic (Etoh 1983, 1986; Pooler and Simon 1994) and genetically defined families produced (Jenderek and Hannan 2000). These families offer the first opportunity in history to develop a genetic map of garlic, assess synteny with onion and other plants, and open the door for genetic improvement of this important plant.

Asparagus is a diploid (2n = 2x = 20), dioecious plant with a relatively small DNA content at 1800 Mbp per 1C genome (Galli et al. 1993). Much effort has gone into genetic mapping of asparagus using a variety of segregating families (Restivo et al. 1995; Jiang et al. 1997; Spada et al. 1998). A bacterial artificial chromosome (BAC) library of asparagus covering approximately 2.6X of the genome has been synthesized (Nitz et al. 2002). There is variation for genome sizes among asparagus species, due in part to polyploidy (Stajner et al. 2002). Asparagus (A. officinalis) has twice as much nuclear DNA as the South African asparagus fern (A. plumosus Baker), even though both plants have the same chromosome number (2n = 2x = 20). It is presently unclear if the larger asparagus genome is due to accumulation of repetitive DNA or a genome doubling followed by chromosome fusions.

The application of genomic technologies to Asparagales would be greatly enhanced by generation of expressed sequence tags (ESTs), production of high-density genetic maps based on homologous sequences from divergent species, and the assembly of genomic contigs covering large regions of the nuclear DNA. These resources would allow breeders and geneticists to efficiently assign important phenotypes to chromosomes by segregation analyses, identify syntenous genomic regions in small-genome model species, and sequence to reveal candidate genes. Asparagus has one of the smallest nuclear genomes known among the higher Asparagales (Fig. 1); however, this genome is still three times larger than rice and about half that of maize. BAC libraries of asparagus would require 38,000 clones of 130 kb for a 1X coverage at a 99% confidence level, many fewer than for other members of Asparagales (Table 1).

Enormous numbers of expressed sequence tags and deep coverage genomic libraries have been produced for members of Poales, including barley, rice, maize, sorghum, sugarcane, and wheat. Genetic linkage conservation (synteny) among Poales is widely recognized (Moore 1995; Devos and Gale 1997, 2000; Paterson et al. 2000) and supported sequencing...

**Fig. 1.—Relative amounts of nuclear DNA (megabase-pairs per 1C nucleus) for major cultivated monocots (Arumuganathan and Earle 1991).**

<table>
<thead>
<tr>
<th>Plant</th>
<th>Genome size (Mbp/1C)</th>
<th>No. BAC clones</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asparagus (Asparagus officinalis L.)</td>
<td>1323</td>
<td>40,615</td>
</tr>
<tr>
<td>Yucca (Yucca kaibabensis McKelvey)</td>
<td>2499</td>
<td>76,720</td>
</tr>
<tr>
<td>Agave (Agave tequilana Web.)</td>
<td>4312</td>
<td>132,381</td>
</tr>
<tr>
<td>Vanilla (Vanilla phaenantha Rchb. f.)</td>
<td>7448</td>
<td>228,660</td>
</tr>
<tr>
<td>Iris (Iris tubergeniana Foster)</td>
<td>9197</td>
<td>282,356</td>
</tr>
<tr>
<td>Diploid (2n = 2x = 16)</td>
<td>14,330</td>
<td>439,945</td>
</tr>
<tr>
<td>Leek (Allium ampeloprasum L.)</td>
<td>14,749</td>
<td>452,809</td>
</tr>
<tr>
<td>Amaryllis (Amaryllis belladonna L.)</td>
<td>15,901</td>
<td>488,176</td>
</tr>
<tr>
<td>Garlic (Allium sativum L.)</td>
<td>16,072</td>
<td>493,426</td>
</tr>
<tr>
<td>Aloe (Aloe vera L.)</td>
<td>16,415</td>
<td>503,957</td>
</tr>
<tr>
<td>Onion (Allium cepa L.)</td>
<td>16,415</td>
<td>503,957</td>
</tr>
</tbody>
</table>
of the rice nuclear genome as model system for Poaceae (Gale and Devos 1998). It is not clear how representative Poales as a group, and rice as an individual species, are for the monocots as a whole or how widely genomic resources developed for Poales will be applicable to other monocots with large, complex genomes. To address these questions, we generated 11,008 unique onion ESTs and are completing comparative mapping of asparagus, garlic, onion, and rice.

MATERIALS AND METHODS

Synthesis of and sequencing from a normalized complementary (c) DNA library of onion has been previously described by Kuhl et al. (2004) and its characteristics are repeated here for convenience only. Tissue was harvested from immature bulbs of onion cultivar ‘Red Creole,’ the callus of an unknown onion cultivar, and roots of cultivars ‘Ebano’ and ‘Texas Legend.’ All tissues were immediately frozen in liquid nitrogen after harvest and stored at −80°C until RNA extraction by the Trizol/Messagemaker system (Invitrogen, Carlsbad, California, USA) and oligo-dT chromatography (Sambrook et al. 1989). Equal molar amounts of polyA + messenger (m) RNA from the three tissues were combined and cDNAs synthesized after priming with oligo-dT. cDNAs were size selected to enrich for molecules >1.0 kb and subjected to a proprietary normalization process (Invitrogen). cDNAs were directionally cloned into the pCMVSport6.1-ccdb (Invitrogen) vector. A sample of cDNAs from the normalized and non-normalized libraries were plated, transferred to colony lift membranes, and hybridized with a β-tubulin clone from onion to assess the efficacy of the normalization step.

A total of 20,000 random clones were subjected to single-pass sequencing reactions from the 5′-end. Base calling, vector trimming, and the removal of low-quality bases was performed (Chuo and Holmes 2001). ESTs were assembled into tentative consensus (TC) groups using clustering tools as described by Pertea et al. (2003). The onion TCs and singletons were searched against the rice gene index at The Institute for Genomic Research (TIGR) using BLAST (Yuan et al. 2000) and requiring >70% identity extending to within 30 bp of the ends. Matching rice ESTs were then searched against rice BAC sequences requiring matches >95% identity over 80% of EST lengths. The positions of these accesses on the rice genetic map were identified based on the alignments of rice marker and BAC sequences. Onion ESTs were selected that showed high similarities (e < −10) to single positions on rice chromosomes. The selected onion ESTs were aligned on corresponding rice BACs and external and nested primers were designed based on sequence conservation between onion and rice. Onion genomic regions were amplified from inbred lines Alisa Craig (AC) 43 and Brigham Yellow Globe (BYG) 15–23. Single amplicons were excised from agarose gels, cloned into a plasmid vector, and sequenced (Lilly and Havey 2001). Single nucleotide polymorphisms, insertions/deletions (indels), or polymorphic restriction-enzyme sites were revealed after sequence alignments and scored using amplicons from the segregating onion family from BYG15-23 by AC43 (King et al. 1998a).

RESULTS AND DISCUSSION

The normalized onion cDNA library consisted of 3.4 × 10⁷ recombinants with an average insert size of 1.6 kb. The normalization process was successful and reduced 70-fold the frequency of β-tubulin cDNAs. We completed 20,000 single-pass sequencing reactions on random clones from this library to yield 18,484 high-quality ESTs of at least 200 bp
in size. These 18,484 sequences assembled into 3690 tentative consensus (TC) sequences and 7318 singletons, yielding 11,008 unique transcripts. These onion sequences represent the first large set of publicly available ESTs from a monocot outside of Poales. Putative functions were assigned to the 11,008 unique onion sequences by searching against manually annotated gene ontology (GO) proteins from model species. A total of 2608 sequences could be assigned to GO functional annotations (Fig. 2). The most common annotation class among onion ESTs was metabolism (19%), as previously observed in tomato (Lycopersicon esculentum L.; Van der Hoeven et al. 2002) and rice (Kikuchi et al. 2003). These ESTs were used to develop the onion gene index (http://www.tigr.org/tigr-scripts/tgi/T_index.cgi?species=onion), the first for a monocot outside of class Commelinaceae. This is a searchable gene index with assignment of onion ESTs to likely gene ontologies and metabolic pathways.

Onion genomic amplicons from BYG15-23 and AC43 revealed primarily single nucleotide polymorphisms (SNPs) with relatively few microsatellites or indels. Transitions were the most common class of SNPs at approximately 30% for TC and AG polymorphisms; transversions were less common at approximately 10% each for GT, AC, CG, and AT polymorphisms. Because of the residual heterozygosity in onion inbreds (King et al. 1998b), some SNPs between parental inbreds do not segregate in our onion families.

A useful application of Poales genomic resources would be syntenic chromosome regions carrying unique, economically important genes in Asparagales. A case in point is cytoplasmic-genic male sterility (CMS) used to produce hybrid-onion seed. The onion has perfect flowers and the production of hybrid seed requires systems of male sterility. For the most widely used source of CMS in onion, male sterility is conditioned by the interaction of sterile (S) cytoplasm and the homozygous recessive genotype at a single nuclear male-fertility restoration locus (Ms) (Jones and Clarke 1943). Maintainer lines used to seed propagate male-sterile lines possess normal fertile (N) and are homozygous recessive at the Ms locus (Jones and Davis 1944). Because onion is a biennial, four to eight years are required to establish if maintainer lines are to be extracted from a population or family (Gokce et al. 2002). This requires great expenditures of time and money and hybrid-onion development would significantly benefit from molecular markers that establish the nuclear genotype of onion plants, significantly reducing the number of testcrosses required to identify maintainer plants. We previously identified molecular markers flanking Ms at 0.9 and 8.6 centiMorgans (cM) (Fig. 3; Gokce et al. 2002); however, these molecular markers were in linkage equilibrium with Ms in open-pollinated onion populations (Gokce and Havey 2002). We require more closely linked markers flanking Ms to make practical marker-facilitated selection of maintainer lines from open-pollinated populations. Instead of relying on a random approach to identify closely linked markers, we would prefer to exploit the genomic resources developed for Poales, using rice as the model. The onion cDNA AOB272 reveals a molecular marker tightly linked to Ms at 0.9 cM (Fig. 3). AOB272 shows a single, highly significant \( (e < -97) \) hit to one position in the rice genome on chromosome #3 at 139.8 cM. Sequence comparisons among molecular markers across a 10 cM region close to the Ms locus of onion revealed no
obvious synteny on the recombinational scale with rice (Fig. 3). In order to test if microsynteny existed between onion and rice, we are mapping onion ESTs showing highly significant similarities to single positions in the rice genome physically linked to position 139.8 on chromosome #3. If microsynteny existed, these onion ESTs should show tight linkage to the chromosome region carrying AOB272 and Ms. Allele-specific markers could then be developed, allowing the seed industry to establish genotypes at the Ms locus without labor-intensive and time-consuming testcrosses.

In conclusion, initial comparative mapping of rice and onion revealed no obvious synteny between Asparagales and Poales at the recombinational level (Fig. 3). It is possible that shorter regions of synteny may exist that could be exploited for identification and cloning of candidate genes in Asparagales. However, our initial results indicate that Poales may not be representative of all monocots requiring that genomic resources be independently developed for major monophyletic lineages within the monocots. Specific small-genome species within the higher and lower Asparagales, such as asparagus and orchid, respectively, represent obvious model plants for the development of deep coverage genomic libraries and comparative mapping. Genetic and genomic analyses of these model Asparagales, such as comparative mapping and BAC-sequence analyses of Asparagales to estimate the level of synteny on the recombinational and sequence levels, will allow for direct comparisons among Asparagales and Poales and will reveal the most appropriate genomic model for important monophyletic lineages outside of Poales. Comparative genomic analyses among Asparagales, Poales, and other monocot orders will provide important insights about genome evolution and diversification for these important plants.

ACKNOWLEDGMENTS

We gratefully acknowledge the funding support of USDA-IFAFS grant 2001-52100-11344.

LITERATURE CITED


