The Potential Impact of Genomes for *Allium* Crop Improvement

M.J. Havey  
Agricultural Research Service  
USDA, Dept. of Horticulture  
1575 Linden Drive  
University of Wisconsin  
Madison, WI 53706  
USA

J. Jakse  
Dept. of Horticulture  
1575 Linden Drive  
University of Wisconsin  
Madison, WI 53706  
USA

J. McCallum  
Crop & Food Research  
Private Bag 4704, Christchurch  
New Zealand

C.D. Town  
The Institute for Genomic Research  
9712 Medical Center Drive  
Rockville, MD 20850  
USA

M. Shigyo  
Faculty of Agriculture  
Yamaguchi University  
Yamaguchi 753-8515  
Japan

Keywords: onion, garlic, bacterial artificial library, expressed sequence tags

Abstract

Onion (*Allium cepa* L.) is the most economically important monocot outside of the grasses and is a member of the family Alliaceae in the order Asparagales. The Asparagales are a monophyletic order sister to the Commelininae, which carries the grasses, palms, and bananas. These two important groups of monocots separated at least 130 million years ago and show little synteny at the recombinational and sequence levels. Onion differs from the grasses for other important genomic characteristics, including an enormous diploid genome, lower GC content of coding regions, and unique telomeric sequences. Expressed sequence tags (ESTs) have proven useful for the development of molecular markers and identification of candidate genes in the Alliums. Sequencing from a variety of cDNA libraries will yield a larger sample of expressed regions in the *Allium* genomes. Pilot sequencing of onion BACs revealed low gene densities with long tracts of degenerated transposable elements. Reduced representation sequencing of onion genomic DNA will provide sequences that can be compared with ESTs to monitor coverage of expressed regions. Molecular markers can be developed from these genomic and expressed sequences for comparative mapping among the Alliums, as well as marker-facilitated selection.

INTRODUCTION

Monocot Phylogenies and Values

The monocots are a monophyletic group strongly supported by morphologies and sequencing of nuclear, mitochondrial, and chloroplast genes (Judd et al., 1999). The Poales and Asparagales are monophyletic orders within the monocots, each supported by a plethora of morphological and DNA characters (Chase et al., 1995, 2000). Recent studies demonstrate that the Commelininae [includes the Poales (grasses), Zingiberales (bananas), and Arecales (palms)] is sister to the Asparagales, which together are sister to the Liliales. The most economically important monocots are in the Poaceae (order Poales) that includes barley, maize, pearl millet, rice, sugarcane, and wheat. The order Asparagales carries the Alliaceae, the second most economically important monocot family and includes such important plants as onion (*Allium cepa*), garlic (*A. sativum*), leek (*A. ampeloprasum*), chive (*A. schoenoprasum*), and Japanese bunching onion (*A.
fistulosum). Garlic and onion are consumed daily by a vast majority of the world’s population. Onion is the second most valuable vegetable crop in the world following only tomato (FAO, 2002) and the third most valuable in the USA following only lettuce and tomato (USDA, 2002). The annual farm-gate value of onion in the US routinely exceeds $750 million (greater than peanut or barley), with over $6 billion in value added after processing (USDA, 2002; NOA, 2003). The annual farm-gate value of US garlic and Japanese bunching onion routinely exceeds $170 million (greater than oat) (USDA, 2002).

Power of Comparative Genomics

Extensive genetic-linkage conservation (synteny) has been demonstrated among cultivated plants in the Poaceae (Binelli et al., 1992; Ahn et al., 1993; Dunford et al., 1995), Fabaceae (Weeden et al., 1992; Boutin et al., 1995), and Solanaceae ( Tanksley et al., 1992). Synteny among related species aids in the identification, mapping, and cloning of economically important qualitative and quantitative trait loci from related plants (Maughan et al., 1996; Paterson et al., 2000). Although synteny among the cultivated grasses is widely recognized (Moore, 1995), fine mapping, genomic sequencing, and in silico comparisons across short genomic regions in the grasses revealed gene amplification and movement (Song et al., 2002), duplications and/or deletions (Klein et al., 2003; Ilic et al., 2003), rearrangements (Li and Gill, 2002; Salse et al., 2004), and accumulation of different repetitive DNAs or retrotransposons (Tarchini et al., 2000). Nevertheless, synteny among the grasses supported sequencing of the rice genome as a model for the grasses (Gale and Devos, 1998) and potentially for other monocots (Havukkala, 1996). Research must be completed to determine how representative the rice genome is for other major monocots.

RESULTS AND DISCUSSION

Allium Genomes

We analyzed expressed and genomic regions in onion and made comparisons to rice. Onion has 16,415 megabasepairs (Mbp) of DNA per 1C nucleus (Orhi et al., 1998). The onion genome is 6 and 16-times greater than maize and rice, respectively. Garlic has a nuclear genome about 7% smaller than onion; leek and chive are tetraploid and hexaploid, respectively (Ori et al., 1998). Allium schoenoprasum subsp. sibiricum has the smallest nuclear genome among all Alliums at 7,448 Mbp per 1C. However, this species is not a good genomic model for the Alliums because it is apomictic, exists as a ploidy series from 2× to 8×, and is commonly aneuploid with numerous B chromosomes (Poulsen, 1990).

1. Analyses of Expressed Regions of the Onion Genome. We synthesized a normalized cDNA library of onion and completed 20,000 single-pass sequencing reactions from the 5′ ends of cDNAs (Kuhl et al., 2004), yielding 18,388 sequences (Genbank accessions CF434396 to CF452784) of which 11,008 were unique. Codon biases and GC content of the onion ESTs were much more similar to Arabidopsis and the eudicots than the grasses. These differences affect comparative-genomic analyses, such as gene prediction (Yang and Nielsen, 2002), identification of CpG islands (Bernardi, 2000), and codon usages (Campbell and Gowri, 1990).

2. Onion Genome is Dominated by Middle-Repetitive DNAs. The GC content of onion DNA is 32%, among the lowest known for angiosperms (Kirk et al., 1970). Cot reassociation kinetics revealed that the onion genome consists of middle-repetitive sequences occurring in short-period interspersions among single-copy regions (Stack and Comings, 1979). This structure of the onion DNA was supported by FISH analyses of random onion BACs. An onion BAC library of 48,000 clones (0.3× coverage of the onion nuclear genome) has been constructed (Suzuki et al., 2001). FISH analysis revealed that 80% carried common repetitive DNAs and hybridized to entire chromosomes, 15% hybridized to centromeric or telomeric regions, and only 5% of BACs hybridized to
specific regions on chromosomes (Suzuki et al., 2001). We sequenced two onion BACs (Genbank accessions DQ273270 and DQ273272) and only 4.5% of the sequences were similar to genes in Arabidopsis or rice! The rest of the sequence was similar to retro-elements and transposons, was devoid of any open-reading frames, and carried sequences highly repeated throughout the onion genome. This pilot sequencing revealed very low gene densities and high frequencies of repetitive DNAs in the onion genome, which will greatly complicate contig building for cloning of economically important loci.

3. There is No Synteny between Onion and Rice. We selected 225 onion ESTs showing high similarities to single-copy expressed regions of the rice genome (Kuhl et al., 2004) and mapped indels and single nucleotide polymorphisms (SNPs) in onion. This study revealed little to no syntenic on the recombinational between onion and rice (Martin et al., 2005), demonstrating that the smaller genome of rice is not syntenic with the enormous nuclear genome of onion. Although the order Asparagales (carries the Alliaceae) and the commelinid monocots (carries the Poaceae) are sister monophyletic groups, they separated at least 130 million years ago (Jasson and Bremer, 2004) and the grass genomes are not representative of all monocots.

**Allium Genomics**

Our research has demonstrated that genomic resources must be independently developed for major plants in the Asparagales, such as the Alliaceae. We recommend that a coordinated program be organized to minimize duplication of effort in the construction of genomic resources for the Alliums. Specific aspects of this international, collaborative effort could include:

1. Expanded EST Sequencing and Generation of an Onion Unigene Set. Our previously developed have been useful to researchers around the world as expressed sequences outside of the commelinoid monocots. The *Allium* research community should work together to significantly increase the numbers of ESTs by completing >100,000 sequencing reactions on a variety of cDNA libraries. These libraries should be normalized to remove high-copy transcripts and maximized for full-length cDNAs. Prompt deposition of all sequences into public databases is necessary and desirable.

2. Assignment of Onion Unigenes to Chromosomes. The *Allium* community is very fortunate to have a complete set of alien addition lines of JBO carrying single onion chromosomes (Shigyo et al., 1996). These alien addition lines can be used to efficiently assign thousands onion unigenes to chromosomes in a high throughput, PCR format. Onion unigenes showing significant similarities to low-copy sequences in other plants can be selected and aligned with the most similar rice EST and genomic sequences (Kuhl et al., 2004). We previously showed that 83% of introns are shared between expressed regions of onion and rice genomes (Martin et al., 2005). Fluorescently labeled primers can be designed using conserved sequences flanking the rice introns and amplicons resolved on capillary sequencing machines. An onion EST can be assigned to a chromosome when amplicons from *A. fistulosum* (when present) and onion are of different sizes and the onion amplicon is present in only one of the alien addition lines.

3. Development of Comparative Maps in *Allium*. Synteny among the Alliums would be of great benefit for the identification and tagging of orthologous qualitative and quantitative traits. We produced the most detailed genetic map of onion based primarily on RFLP, SNP, and SSR markers (King et al., 1998; Martin et al., 2005) and observed that the onion ESTs are an excellent source of codominant SSR markers. Primer sets designed for 85 ESTs, of which 46 amplified single fragments carrying polymorphisms between the parental inbreds of our onion mapping family and 35 (42%) segregated (Martin et al., 2005). This was a much higher frequency of segregating markers than we revealed with RFLPs (12%) or SNPs (18%). The newly generated onion ESTs can be screened for SSRs or SNPs and primer pairs designed from the EST sequences flanking these motifs. Amplicons can be resolved on capillary sequencing machines and segregations established using the following segregating families:
**Onion.**

- Three intraspecific segregating families of onion have been developed. The first consists of $F_2$ progenies from the cross of inbreds BYG15-23 and AC43. This family has been used to map over 250 molecular markers (King et al., 1998; Martin et al., 2005), as well as male-fertility restoration (Gokce et al., 2001) and major QTLs affecting health-enhancing attributes of onion (Galmarini et al., 2001; Havey et al., 2004). A second family consists of $F_2$ progenies from the cross of white and yellow inbreds and segregates for the major bulb- and seed-color loci (Havey, 1996). A third family was developed in New Zealand from a cross between Colossal and P12 (McCallum et al., 2006).

- An $F_2$ family from an interspecific hybrid between onion and *A. roylei* has been used to map AFLPs and EST-SSRs (Heusden et al., 2000; McCallum et al., 2006).

**Japanese Bunching Onion.** An $F_2$ population is presently being used for map development by Dr. H. Tsukazaki in Japan (Song et al., 2004).

**Garlic.** The production of true seed of garlic has opened the door for genetic analyses of this important *Allium* (Simon and Jenderek, 2003). Two segregating families have been developed. The first was used to map AFLPs (Ipek et al., 2003) and the second was used to produce the first genetic map of expressed sequences in garlic (Zewdie et al., 2005).

4. **Development of a Large-Insert Library for an Allium Genome.** A BAC library of onion would require an enormous number of clones; for example a 6×-coverage library would need 757,616 clones of average size 130 kb. Although this library would be extremely useful for isolation of promoter regions for ESTs and revealing the structure of the onion genome, a huge number of the BAC clones would likely not carry any coding regions. Pilot sequencing of onion BACs and BAC ends from the 0.3×-coverage library (Suzuki et al., 2001) revealed very low gene densities, indicating that there may be on average less than one gene for every three onion BACs. It may be a more judicious use of resources to synthesize the first BAC library of an *Allium* with well defined euchromatic regions. Japanese bunching onion (*A. fistulosum*) has a nuclear genome of 12,275 Mbp per 1C, 28% smaller than onion, and would require 569,584 random BAC clones of 130 kb for 6× coverage. CoT analyses demonstrated that the larger onion genome has a significant middle-repetitive fraction not present in JBO (Stack and Comings, 1979). Narayan (1988) documented that each Japanese bunching onion chromosome is 28% smaller than the corresponding onion chromosome, indicating that the middle repetitive DNAs in onion are uniformly distributed across chromosomes. Genomic in-situ hybridizations (GISH) clearly distinguish chromosomes of Japanese bunching onion and onion (Peffley and deVries, 1993; Krustaleva and Kik, 1998). Both C-banding and GISH analyses reveal that the terminal half of JBO chromosomes are heterochromatic and that clearly defined euchromatic regions are adjacent to the centromeres. Chiasmata are localized in the euchromatin near the Japanese bunching onion centromeres (Emmsweller and Jones, 1935; Krustaleva et al., 2005). Therefore nuclear genome of Japanese bunching onion may contain gene-rich euchromatic regions proximal to the centromeres and yield BACs with higher gene densities than other Alliums with larger genomes.

5. **Reduced Representation Sequencing of an Allium Genome.** Reduced-representation sequencing of maize using methyl or Cot-filtered libraries increased the proportion of random shot-gun reads showing significant similarities to expressed sequences (Peterson et al., 2002; Palmer et al., 2003). We completed pilot sequencing from whole-genome shot-gun (WGS) and methyl-filtered genomic libraries created from a double-haploid population of onion. Our preliminary results indicate that methyl-filtration of onion DNA was effective in reducing the proportion of both identifiable transposons and anonymous sequences, as well as increasing non-organellar protein hits. Therefore, sequencing of methyl-filtered genomic clones should complement EST sequencing as an efficient approach to enrich for genic regions in onion.
Translational genomics refers to the exploitation of genomic technologies, such as extensive EST resources or deep coverage genomic libraries, for the genetic improvement of economically important plants. If synteny existed among the Alliums, enormous genomic resources developed for one *Allium* can be applied to other species in this genus, and greatly aid in the identification of candidate genes conditioning economically important traits. Candidate genes can then be mapped in relation to target traits, molecular markers developed for high throughput analyses, and marker-facilitates selection used to produce new high-value vegetable products. International collaboration bringing together sequencing resources with scientists working on genetic improvement of the Alliums will avoid duplication of effort by concentrating resources on the development of comparative maps and genomic resources. Close working relationships will apply these genomic resources to the genetic improvement of important *Allium* vegetables.

ACKNOWLEDGEMENTS

This work was supported by U.S. Department of Agriculture, Initiative for Future Agriculture and Food Systems Grant 2001-04434; Grant-in-Aid for Young Scientists (No. 40314827) from the Ministry of Education, Culture, Sports, Science and Technology, Japan; and the Fulbright-Hayes Post-doctoral Fellowship Program. Names are necessary to report factually on available data; however, the U.S. Department of Agriculture (USDA) neither guarantees nor warrants the standard of the product, and the use of the name by USDA implies no approval of the product to the exclusion of others that may also be suitable.

Literature Cited


