THE DEVELOPMENT OF TOOLS FOR MOLECULAR BREEDING AND GENOMICS RESEARCH IN CULTIVATED SUNFLOWER

Steven J. Knapp, Mary B. Slabaugh, and Shunxue Tang, Department of Crop and Soil Science, Oregon State University, Corvallis, OR 97331-3002, USA
Fax: 541-737-1334; e-mail: steven.j.knapp@orst.edu

SUMMARY

Although significant advances have been made in the development of tools for molecular breeding and genomics research in cultivated sunflower (*Helianthus annuus* L.), coordinated, multi-institutional research alliances are needed to eliminate deficiences and gain ground on other economically important commodity crops. This paper reviews sunflower genomics and molecular breeding work underway in our laboratory and describes deficiencies that need to be addressed to advance genomics research in sunflower. We review two topics: (i) the development of an expressed sequence tag (EST) database for sunflower and (ii) the development of locus-specific, high-throughput DNA markers, specifically simple sequence repeats (SSRs), insertions-deletions (INDELs), and single nucleotide polymorphisms (SNPs) for sunflower.

EST RESOURCES FOR SUNFLOWER

The public DNA sequence databases for sunflower are limited. By last count, GenBank held less than 200 *Helianthus* cDNA and genomic DNA sequences. Furthermore, cDNA and genomic DNA clones corresponding to the sequences are not held in a central despository and are presently either not shared or are distributed on an ad hoc basis by individual investigators. We are in the process of developing public, mass sequence resources for sunflower, specifically expressed sequence tags (ESTs). These resources should accelerate gene discovery through data mining and cDNA microarray analysis and create a resource for DNA marker development in sunflower.

We have begun mass sequencing cDNA clones from a developing seed cDNA library and, in collaboration with David Shintani (University of Nevada, Reno) and Loren Rieseberg (University of Indiana), are working towards producing an initial database of ~4,000 ESTs from developing seed, root, and leaf cDNAs cultivated and wild sunflowers. Our aim is to isolate ~2,500 unique ESTs. The EST database and catalogued cDNA clone collection will be useful resources for identifying and isolating sunflower genes, performing a variety of functional genomic analyses (e.g., cDNA microarray analysis), and developing DNA markers. The EST database will ultimately be publicly released.

Our laboratory, in collaboration with Richard Michelmore (University of California, Davis), Loren Rieseberg, Dave Shintani, and others, are proposing to develop a database comprised of several thousand ESTs isolated from a variety sunflower and lettuce (*Lactuca sativa* L.) cDNA libraries. If funded, the EST database and cDNA clones thus produced will be publicly released.

HIGH-THROUGHPUT DNA MARKER DEVELOPMENT

One of the goals of our laboratory has been to develop several hundred PCR-based, high-throughput, locus-specific, co-dominant DNA markers for cultivated sunflower. Our focus has been on the development of simple sequence repeat (SSR), insertion-deletion (INDEL), and, more

recently, single nucleotide polymorphism (SNP) markers. The lack of freely shared, locus-specific markers has impeded genomics and molecular breeding research in sunflower, particularly in the public sector, and deterred the development of a dense integrated genetic map. The sunflower research community needs to pull together to eliminate this bottleneck.

Although 2,044 genomic and cDNA probes have been screened for RFLPs and 1,141 RFLP loci have been mapped in cultivated sunflower (Berry et al. 1995, 1996, 1997; Gentzbittel et al. 1995, 1999; Jan et al. 1997), only 81 RFLP (cDNA) probes have been released to the public sector (Berry et al. 1997). Sixty-five of the latter have been integrated with 296 AFLP markers, 9 RFLP markers for resistance gene analogs (RGAs) (Gedil et al., 2000b), and 45 RFLP markers from the genetic map of Jan et al. (1997) (Gedil et al., 2000a).

SSR MARKER DEVELOPMENT

Our laboratory has developed three sets of SSR markers. The first set is comprised of 85 SSR markers developed by screening published DNA sequences and a random genomic DNA library for a variety of di- and tri-nucleotide repeats (unpublished data). The primer sequences for this set have been publicly released.

The second set is comprised of 132 SSR markers developed as part of the CARTISOL III research program by screening a genomic DNA library enriched for several di-, tri-, and tetra-nucleotide repeat motifs (unpublished data). The library we used for this work was developed by Keith Edwards and Lucy Thompson (Long Ashton Research Station, UK). Six hundred and thirty-two clones harbored repeats, of which 241 were unique. Because of short flanking sequences, primers could only be developed for 72% of the unique repeats (172/241). The primer sequences for this set are to be publicly released by CARTISOL III once the research is completed.

The third set is comprised of SSR markers developed from genomic DNA libraries enriched for GA and CA repeats. We have sequenced a total of 972 clones thus far from the libraries (492 from the CA-enriched and 480 from the GA-enriched library). Repeats were found in 758 clones, of which 569 were unique. Because of short flanking sequences, primers could only be developed for 92% of the unique repeats (524/569). Based on our initial results, 70% of the primers (367/524) should yield functional SSR markers. The primer sequences for this set will be publicly released once the initial screening and mapping work are completed.

The genotyping qualities of SSR markers vary in sunflower. We have tested over 500 SSR primer pairs thus far and found that some amplify a single locus with no null alleles, some amplify a single locus but fail to amplify alleles in some lines (produce null alleles), some amplify two or more seqeuences (either duplicate loci or non-target artifacts), and some fail to amplify the target sequence. Our genotyping experience has mostly come from assays performed on an ABI377 automated DNA sequencer using fluorescently labelled amplicons. The genotypes for six SSR markers (two three-plex panels of 16 inbred lines) labelled with different fluorophores (6FAM, TET, and HEX) are shown in Fig. 1.

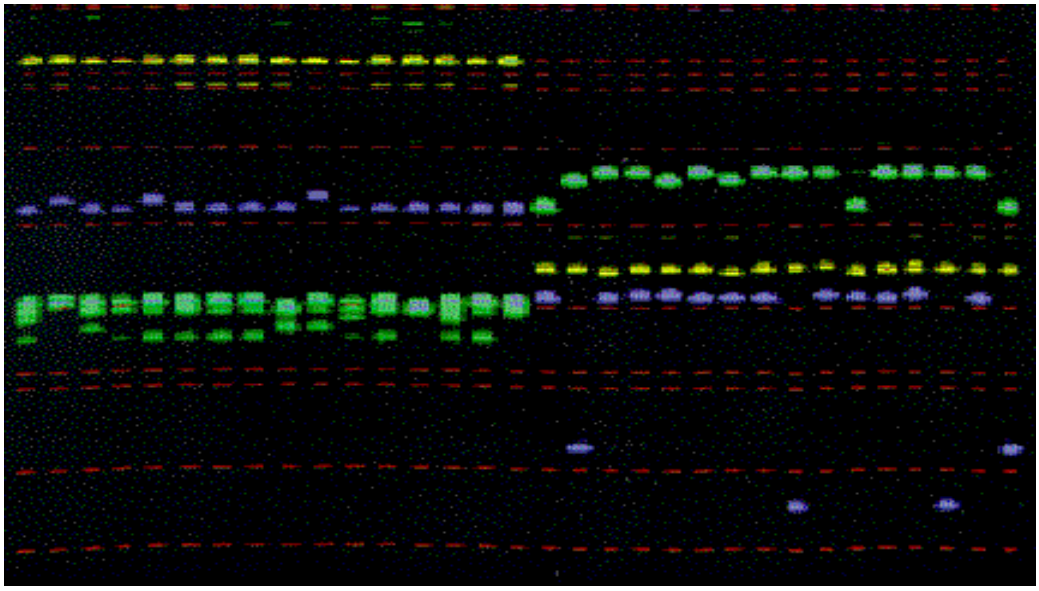

Figure 1. Genotypes for six SSR markers (three in lanes 1-16 and three in lanes 17-32) assayed on 16 inbred lines of sunflower. The arrows show the positions of each marker. Genotyping assays were performed on an ABI377 using fluorescently end-labelled forward primers (6FAM, TET, or HEX).

We typically gel-multiplex three to six SSRs by separately amplifying individual loci and pooling the amplicons. PCR-multiplexes, where primer pairs are pooled and amplicons are simultaneously produced in single tube, are difficult to implement and have two serious drawbacks. First, many primer pairs fail to PCR-multiplex or do not PCR- multiplex well. Second, PCR-multiplexing is inefficient when the sample of SSR markers needed for different analyses changes. Gel-multiplexing has lower throughput than PCR- multiplexing but permits the experimenter to change the sample of SSR markers from experiment to experiment. Throughput is increased with gel-multiplexing performed with different SSR amplicon lengths and multiple fluorophores (Fig. 1), but amplifying individual loci is still a bottleneck. Our laboratory is exploring ways to reduce this bottleneck by using a liquid handling robot to prepare PCR samples.

SEQUENCE TAGGED SITE (STS) MARKER DEVELOPMENT

We have been employing and testing a variety of strategies and methods for identifying DNA polymorphisms in sunflower genes and using the sequence knowledge to develop sequence-tagged-site (STS) markers. The process entails (i) developing locus specific oligonucleotide primers (LSOPs) for the target sequence, (ii) amplifying the target sequence from two or more genotypes, (iii) assessing amplicon complexity, and (iv) assaying amplicons for single nucleotide polymorphisms (SNPs), insertions-deletions (INDELs), and simple sequence repeats (SSRs).

The methods we employ to screen amplicons for DNA polymorphisms are (i) direct allele sequencing, (ii) single stranded conformational polymorphism (SSCP) analysis, and (iii) length polymorphism analysis on agarose or polyacrylamide gels. The latter are performed on automated DNA sequencers using fluorescent assays. The same chemistry can be used for INDELs and SSRs. The only difference between the two is that genetic markers for the former do not produce stutter bands. The electropherogram shown in Fig. 2 was produced using the SAD17-3 LSOPs described by Hongtrakul et al. (1998) for a Δ9 stearoyl-ACP desaturase gene. The genotyping was performed on an ABI377 automated DNA sequencer (Celera Genomics, Foster City, California) using a fluorescently end- labelled forward primer and unlabelled reverse primer. Ten alleles were observed among 24 elite oilseed inbred lines and two confectionery inbred lines. Allele lengths ranged from 430 to 474 bp in lanes 1-25. The allele in lane 26 was 760 bp long (not shown). The 760 bp allele was present in a confectionery inbred lines not shown. was found in . The forward and reverse primers (SAD17-3) flank intron length variants.

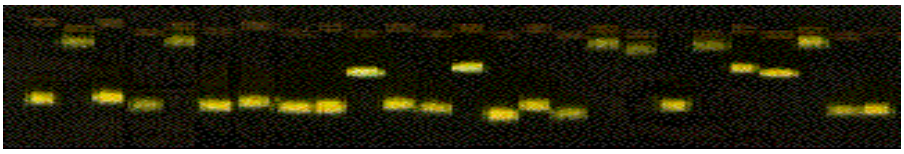

Figure 2. DNA marker genotypes (allele length variants) for a Δ9 stearoyl-ACP desaturase gene assayed among 26 inbred lines of sunflower on a ABI377 automated DNA sequencer using amplicons produced by the SAD17-3 primer pair (Hongtrakul et al., 1998). The forward primer was end-labelled with FAM.

Direct allele sequencing is the `gold standard' for identifying polymorphisms, but is more costly than the other methods and is limited to locus specific amplicons. When a primer pair amplifies fragments of the same length from two or more loci belonging to a multigene family, methods other than DNA sequencing must be used to screen for polymorphisms.

SSCP analysis (Orita et al. 1989) has broad utility for DNA polymorphism screening and genotyping. Although SSCP analysis does not identify every DNA sequence polymorphism, the methodology is straightforward and identifies a large fraction of the polymorphisms. SSCPs are produced by differential folding of single-stranded DNA harboring mutations (Orita et al. 1989). The conformation of the folded DNA molecule is produd by intra-molecular interactions and is thus a function of the DNA sequence. SSCP analysis is performed using heat-denatured DNA on non-denaturing DNA sequencing gels. Special gels (e.g., mutation detection enhancement gels) have been developed to increase sensitivity. SSCP analysis is particularly useful for assessing amplicon complexity and electrophoretically separating individual members of a gene family for subcloning and sequencing (Slabaugh et al., 1997; unpublished data).

The development of locus specific assays for genes belonging to multigene families can be challenging. The most complex gene families we have worked with are candidate diseases resistance genes or resistance gene analogs (RGAs), e.g., nucleotide-binding-site, leucine-rich-region (NBS-LRR) and protein kinase (PK) gene families (Gentzbittel et al., 1998; Meyers et al., 1998; Shen et al., 1998; Gedil et al., 2000), such as the NBS-LRR cluster linked to downy mildew resistance genes in sunflower (Mouzeyar et. al., 1995; Roeckel-Drevet et. al., 1996; Vear et al., 1997). RFLP analysis of RGAs, while tractable, is tedious and requires large, clean, DNA samples. Furthermore, most RGA probes hybridize, with varying degrees of stringency, to multiple DNA fragments and can be difficult to genotype (Gedil et al., 2000). Because most NBS-LRR genes belong to tandemly repeated clusters (Meyers et al., 1998; Shen et al., 1998), developing LSOPs to amplify one or two members of a gene family can be difficult, particularly when sequences are lacking for one or more members of the family (Gedil et al., 2000). SSCP analysis has been a particularly powerful method for assessing the complexity of amplicons, estimating the number of loci amplified by an LSOP pair, and isolating, cloning, and sequencing individual members of RGA gene families (Gedil et al., 2000; unpublished data). We are presently testing direct staining methods to simplify and accelerate SSCP analysis.

The other DNA polymorphism screening methods worth exploring are cleavase fragment length polymorphism (CFLP) analysis (Sreevatsan et al., 1998; Oldenburg and Siebert, 2000) and heteroduplex analysis (Cho et al., 1999; Escary et al., 1999; Wang et al., 1999). CFLP analysis has been shown to be more sensitive than SSCP analysis and can be used to locate SNPs and other polymorphisms in long amplicons (1000 to 3000 bp). Because long fragments can be scanned, CFLP analysis may be an efficient tool for identifying SNPs and preselecting alleles for sequencing.

LITERATURE CITED

Berry, S.T., Leon, A.J., Hanfrey, C.C., Challis, P., Bukholz, A., Barnes, S.R., Rufener, G.K., Lee, M., and Caligari, P.D.S. 1995. Molecular marker analysis of *Helianthus annuus* L. 2. Construction of an RFLP linkage map for cultivated sunflower. Theor. Appl. Genet. 91: 195-199.

Berry, S.T., Leon, A.J., Challis, P., Livini, C., Jones, R., Hanfrey, C.C., Griffiths, S., and Roberts, A. 1996. Construction of a high density, composite RFLP linkage map for cultivated sunflower (*Helianthus annuus* L.). pp. 1155-1160. *In* Proc. 14TH Int. Sunflower Conf. Vol. 2, Beijing, China.

Berry, S.T., Leon, A.J., Peerbolte, R., Challis, C., Livini, C., Jones, R., and Feingold, S. 1997. Presentation of the Advanta sunflower RFLP linkage map for public research. pp. 113-118. *In* Proc. 19TH Sunflower Res. Workshop, Fargo, North Dakota.

Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Lane, C.R., Lim, E.P., Kalyanaraman, N., Nemesh, J., Ziaugra, L., Friedland, L., Rolfe, A., Warrington, J., Lipshutz, R., Daley, G.Q., and Lander, E.S. 1999. Characterization of single- nucleotide polymorphisms in coding regions of human genes. Nature Genetics 22: 231-238.

Cho, R.J., Mindrinos, M., Richards, D.R., Sapolsky, R.J., Anderson, M., Drenkard, E., Dewdney, J., Reuber, T.L., Stammers, M., Federspiel, N., Theologis, A., Yang, W.H., Hubbell, E., Au, M., Chung, E.Y., Lashkari, D., Lemieux, B., Dean, C., Lipshutz, R.J., Ausubel, F.M., Davis, R.W., Oefner, P.J. 1999. Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. Nature Genetics 23: 203-207.

Escary, J.-L., Lonie, L., Walter, Cox, R.D., Walter, J.M., and Lathrop, M. 1999. Screening strategy for mutation detection using denaturing high-performance liquid chromatography (DHPLC). Intl. Biotech. Laboratory 17: 19.

Gedil, M.A., S.K. Berry, R. Jones, A. Leon, C. Wye, J. Peleman, and S.J. Knapp. 2000. The development of an integrated RFLP-AFLP map for cultivated sunflower. Genome (submitted).

Gedil, M.A., M.B. Slabaugh, S.K. Berry, R. Jones, R. Michelmore, J.F. Miller, T. Gulya, and S.J. Knapp. 2000. Candidate disease resistance genes in sunflower cloned using conserved nucleotide binding site motifs: genetic mapping and linkage to the *Pl1* gene for resistance to downy mildew. Genome (submitted).

Gentzbittel, L., Vear, F., Zhang, Y.X., and Berville, A. 1995. Development of a consensus linkage RFLP map of cultivated sunflower (*Helianthus annuus* L.). Theor. Appl. Genet. 90: 1079-1086.

Gentzbittel, L., Mouzeyar, S., Badaoui, S., Mestries, E., Vear, F., De Labrouhe, D.T., and Nicolas, P. 1998. Cloning of molecular markers for disease resistance in sunflower, *Helianthus annuus* L. Theor. Appl. Genet. 96: 519-525.

Gentzbittel, L., Mestries, E., Mouzeyar, S., Mazeyrat, F., Badaoui, S., Vear, F., Tourvieille de Labrouhe, and Nicolas, P. 1999. A composite map of expressed sequences and phenotypic traits of the sunflower (*Helianthus annuus* L.) genome. Theor. Appl. Genet. 99: 218-234.

Hongtrakul, V., Slabaugh, M.B., and Knapp, S.J. 1998. A seed specific Δ-12 oleate desaturase gene is duplicated, rearranged, and weakly expressed in high oleic acid sunflower lines. Crop Sci. 38: 1245-1249.

Hongtrakul, V., M.B. Slabaugh, and S.J. Knapp. 1998. DFLP, SSCP, and SSR markers for Δ9-stearoyl-acyl carrier protein desaturases strongly expressed in developing seeds of sunflower: intron lengths are hypervariable among elite inbred lines. Mol. Breeding 4: 195-203.

Hoogendoorn, B., M.J. Owen, P.J. Oefner, N. Williams, J. Austin, and M.C. O'Donovan. 1999. Genotyping single nucleotide polymorphisms by primer extension and high performance liquid chromatography. Hum. Genet. 104: 89-93.

Jan, C.C., Vick, B.A., Miller, J.F., Kahler, A.L., and Butler, E.T. 1998. Construction of an RFLP linkage map for cultivated sunflower. Theor. Appl. Genet. 96: 15-22.

Knapp, S.J., Berry, S.T., and Rieseberg, L.H. 2000. Genetic mapping in sunflowers. pp. (in press) *In* Phillips R.L. and Vasil I.K. (ed.) DNA markers in plants. Kluwer, Dordrecht, the Netherlands.

Meyers, B.C., Chin, D.B., Shen, K.A., Sivaramakrishnan, S., Lavelle, D.O., Zhang Z., Michelmore, R.W. 1998. The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. The plant cell 10:1817-1832.

Mouzeyar, S., Roeckel-Drevet, P., Gentzbittel, L., Philippon, J., Tourvieille De Labrouhe, D., Vear, F., and Nicolas, P. 1995. RFLP and RAPD mapping of the sunflower Pl1 locus for resistance to *Plasmopara halstedii* race 1. Theor. Appl. Genet. 91: 733-737.

Oldenburg, M.C. and Siebert, M. 2000. New cleavase fragment length polymorphism method improves the mutation detection assay. BioTechniques 28: 351-357.

Orita, M., H. Iwahana, H. Kanazawa, K. Hayashi, and T. Sekiya. 1989. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. Proc. Natl. Acad. Sci. USA 86: 2766-2770.

Orita, M., Y. Suzuki, T. Sekiya, and K. Hayashi. 1989. Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. Genomics 5: 874-879.

Roeckel-Drevet, P., Gagne, G., Mouzeyar, S., Gentzbittel, L., Philippon, J., Nicolas, P., De Labrouhe, D.T., and Vear, F. 1996. Colocation of downy mildew (*Plasmopara halstedii*) resistance in sunflower (*Helianthus annuus* L.). Euphytica 91: 225-228.

Shen, K.A., Meyers, B.C., Islam-Faridi, M.N., Chin, D.B., Stelly, D.M., and Michelmore, R.W. 1998. Resistance gene candidates identified by PCR with degenerate oligonucleotide primers map to clusters of resistance genes in lettuce. Mol. Plant Microbe Interact. 11:815-823.

Slabaugh, M.B., Huestis, G.M., Leonard, J., Holloway, J.L., Rosato, C., Hongtrakul, V., Martini, N., Toepfer, R., Voetz, M., Schell, J., and Knapp, S.J. 1997. Sequence-based genetic markers for genes and gene families: Single-strand conformational polymorphisms for the fatty acid synthesis genes of *Cuphea*. Theor. Appl. Genet. 94: 400-408.

Sreevatsan, S. 1998. Comparative evaluation of cleavase fragment length polymorphism with PCR-SSCP and PCR-RFLP to detect antimicrobial agent resistance in *Mycobacterium tuberculosis*. Molecular Diagnosis 3: 81.

Vear, F., Gentzbittel, L., Philippon, J., Mouzeyar, S., Mestries, E., Roeckel-Drevet, P., Tourvielle de Labrouhe, D., and Nicolas, P. 1997. The genetics of resistance to five races of downy mildew (*Plasmopara halstedii*) in sunflower (*Helianthus annuus* L.). Theor. Appl. Genet. 95: 584-589.

Wang, D.G., Fan J.B., Siao, C., Berno, A., Young, P., Sapolsky, R., Ghanour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittman, M., Morris, M.S., Shen, N.P., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T.J., Lipshutz, R., Chee, M., and Lander, E.S. 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280: 1077-1082.