Current and Future Innovations in Soybean (*Glycine max* L. Merr.) Oil Composition

R.F. WILSON*1, J.W. BURTON*1, W.P. NOVITZKY*1 and R.E. DEWEY*2

*1 *USDA-ARS, North Carolina State University*

*2 *Crop Science Department, North Carolina State University*

(100 Derieux St., Raleigh, North Carolina 27695-7620, USA)

Edited by T. Nakahara, AIST, Tsukuba, and accepted January 20, 2001 (received for review October 23, 2000)

Abstract : Biotechnology often is regarded strictly as transgenic research. In practice, it involves a team effort among plant breeders, genomicists, and molecular geneticists. Genetic improvement of soybean (*Glycine max*) began over 5000 years ago when wild soybeans (*Glycine soja*) were introduced into China. These ancestors of cultivated soybean exhibit a wide range of genetic diversity for traits that range from seed size and color to genes that govern oil composition. However, the 'gene pool' for US varieties comes from less than 12 of the 18000 types of *Glycine max*. Thus, the genetic base for modern soybeans is rather narrow. For example, wild soybeans contain desaturase genes (FAD3) are not present in domesticated soybean that contribute to elevated polyunsaturates in *Glycine max*. How can we extract and utilize untapped genetic diversity in soybean or any other crop? The answer is through biotechnology. Soybean breeders have an arsenal of natural gene mutations (recessive alleles) that influence fatty acid composition. Modern genetic technology helps identify the gene, gene product and the exact nature of the mutation in each allele. Transgenic research provides 'proof of concept' and also may be used to create genetic diversity for novel traits. Gene markers, maps and micro-array technology help to locate selected genes in segregating populations. These tools accelerate breeding progress and enable variety development in a socially acceptable manner. These concepts are applied to the development of agronomic soybeans with lower-palmitic acid, higher oleic acid and lower-linolenic acid concentration. In the near future, the means to create natural mutations that fine-tune regulation of metabolic enzyme activities for specific traits will be in hand. Such technical advances may lessen social concern for biotechnology, through more effective use of natural genetic diversity to achieve goals now thought possible only by application of transgenes in commercial food/feed products.

J. Oleo Sci. ***50***, 353-358 (2001).

Key words : breeding, DNA markers, fatty acid alleles, oil quality, saturated-fat, trans-isomers

1 Introduction

Soybean is a good example to show how biotechnology can be used to exploit genetic diversity in any crop species. The extent of natural genetic diversity in soybean is quite broad. We may think of soybean only as having a round yellow seed, but all the seed shown in **Fig. 1** are soybeans. These differences in seed color and size are just a small demonstration of the abundance of natural genetic variation that exists in this crop.

The natural genetic diversity in domesticated soybean (*Glycine max*) traces back to its ancestor (*Glycine soja*) or wild soybean. About 5000 years ago, wild soybean was introduced to China. Perennial breeding and selection by Chinese and Japanese farmers over the millennia eventually led to the proliferation of soybean throughout the world. Today, the United States Department of Agriculture (USDA) maintains a Collection of about 18,000 different types of soybean. Information on this collection is documented by GRIN, the USDA germplasm re-

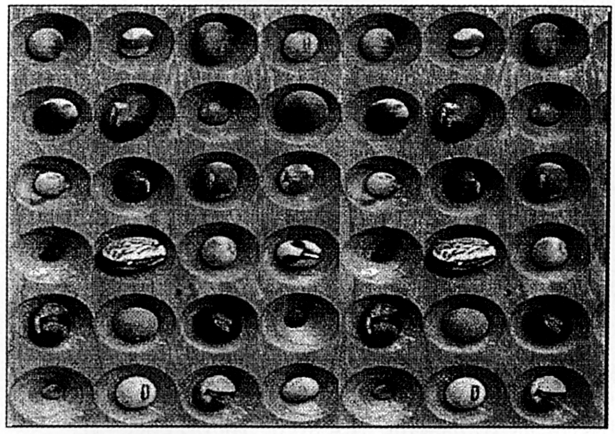

Fig. 1 Genetic Variation for Seed Size and Color in Soybean.

source information network (1).

It may be of interest to know that the very first modern soybean (*Glycine max*) varieties had black seed, a characteristic derived from wild soybean. The only reason soybeans are primarily yellow today, is because that is what the Japanese market wanted. It was a difficult task, but breeders worked together, and eventually broke the strong genetic linkage between 7 genes that determined black seed coat color. In effect, they found a way to give consumers what they wanted.

Like seed color, most traits in soybean are governed by several genes, and natural mutations in those genes provide a great wealth of natural genetic diversity for value-added traits (2). However, biotechnological methods are needed to do a better job of extracting and using the untapped natural genetic diversity in soybean. These efforts will eventually be needed to respond more quickly to changes in consumer preferences, and will provide the flexibility needed to ensure a crop remains competitive in global markets.

As an example, altering oil composition in response to consumer preferences is a current goal for soybean research. The primary target is the development of vegetable oil with improved flavor and frying stability (**Table 1**). Such oil probably will exhibit lower saturated fat, higher oleic acid and adequate levels of linoleic acid to maintain desired flavor. However, other goals also are being pursued which involve the development of a highly saturated soybean oil as an ingredient for low-trans isomer margarine base stocks, and soybean oil with enhanced reactivity for industrial applications. This paper describes the contributions of conventional breeding, genomics and molecular genetics in the pursuit of these 'biotechnical' objectives.

Table 1 Primary Goals for Genetic Alteration of Fatty Acid Composition in Soybean.

Fatty Acid	Normal Oil	Products		
		Cooking Oil	Margarine	Paints
	%Crude Soybean Oil			
Saturated	15	7	42	11
Oleic	23	60	19	12
Linoleic	53	31	37	55
Linolenic	9	2	2	22

All targets may be achieved by natural gene section, and are non-GMO

2 Results & Discussion

2·1 Altering 'Drying Oil' Properties of Soybean

For example, let's assume that a client wants soybean oil with higher levels of linolenic acid (18:3). We know that oil from wild soybean seed is distinguished by relatively high levels of 18:3. We know that 18:3 concentration is determined by the activity of n-3 desaturases. We also know that n-3 desaturase is the product of a gene described as 'FAD3'. So, wild soybean appears to be a good genetic resource for genes that determine higher linolenic acid. This assumption is supported by natural genetic variation for FAD3 gene expression among accessions of wild soybean. The biological or causal basis for this variation partially is shown in Southern blots of DNA fragments from FAD3 genes taken from cultivated and wild soybean (**Fig. 2**). There are distinct differences in the gene structure, as determined by DNA fragment size and mobility, between cultivated and wild soybeans. Similar distinction also may be made among wild soybean lines. This information implies that differences in n-3 desaturase structure lead to greater 18:3 synthetic activity, and that alternative n-3 desaturases were lost or modified during the domestication of *Glycine max* (3).

The proof of that hypothesis comes when plant breeders intermate *Glycine soja* and *Glycine max*. Such a population typically yields progeny with much greater polyunsaturate levels than normal soybean oil (4). Thus, the reintroduction of specific alternative n-3 desaturase genes into *Glycine max* may someday be used to make special soybean oils that are more

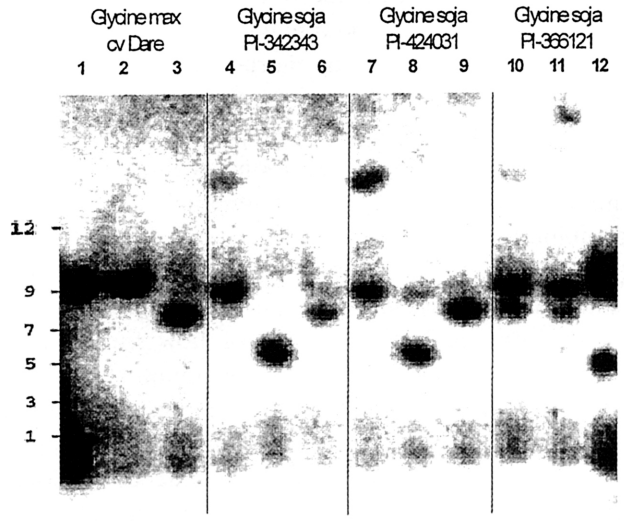

Fig 2. Differences in FAD3 Gene Structure among Soybeans. Southern-blot of DNA polymorphisms that hybridize with a soybean cDNA probe for the FAD3 gene. Genomic DNA was digested with the same restriction enzyme in: Lanes 1,4,7, 10; Lanes 2,5,8,11; and Lanes 3,6,9,12.

attractive for industrial applications, such as inks and coatings.

2·2 Identifying Other Genes That Determine Changes in Fatty Acid Composition

Plant breeders have discovered a considerable arsenal of recessive alleles that influence the level of each fatty acid in soybean oil (**Table 2**). All of these genes are recessive alleles. More are found each year, either as spontaneous events or through natural gene recombination. None are transgenic. All are heritable and may be combined through hybridization to achieve changes in oil composition.

The fact that so many 'recessive alleles' are being discovered for fatty acid traits in soybean, makes it imperative to determine how these putative genes differ in form and function. Such knowledge helps ensure that unique combinations of recessive alleles are achieved. Part of this information comes from knowing the specific gene and gene product (enzyme activity) that is controlled by a given allele, and also by knowing the nature of the mutation that alters the gene product or enzyme activity. Several soybean cDNA probes for genes that encode these enzymes in fatty acid synthesis, and similar probes for many of the genes in the glycerolipid synthetic pathway are now available. Let us assume that we wish to determine the gene and type of mutation that is associated with recessive *fap* alleles that determine palmitic acid (16:0) concentration in soybean. A likely candidate enzyme for gene action is the 16:0-ACP thioesterase. Hence, the FAT-B probe should help determine if natural mutations, in the gene that encodes the 16:0-ACP thioesterase, have anything to do with phenotypic changes in palmitic acid concentration in germplasm exhibiting *fap1*, *fap2* or *fap-nc* alleles. **Figure 3** shows

Table 2 Reccessive Alleles that Determine Fatty Acid Composition in Soybean.

Phenotype		Alleles
16:0	High	fap_2 fap_{2b} fap_4 fap_5
	Normal	Fap
	Low	fap_1 fap_3 fap_6 fap_{nc}
18:0	High	fas fas_a fas_b fas_{nc}
	Normal	Fas
	Low	None
18:1	High	$fad\text{-}1_{nc}$ $fad\text{-}2_{nc}$
	Normal	Fad
	Low	$fad\text{-}W_{soja}$ $fad\text{-}X_{soja}$
18:3	High	$fan\text{-}Y1_{soja}$ $fan\text{-}Y2_{soja}$
	Normal	Fan
	Low	fan fan_1 fan_2 fan_3 fan_{nc}

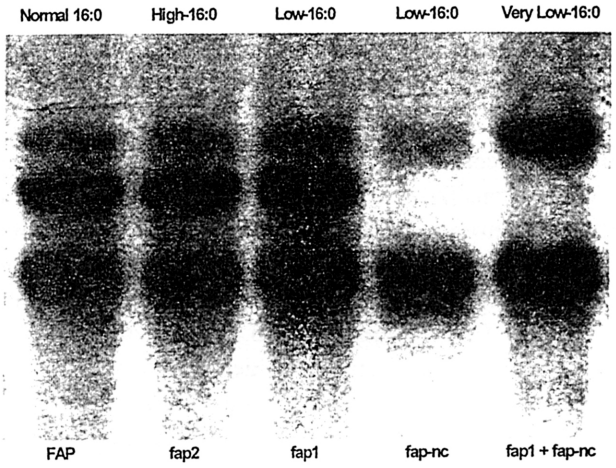

Fig. 3. Genotypic Differences in Structure of the 16:0-ACP Thioesterase Gene in Soybean.
Genomic DNA from soybeans exhibiting alleles at FAP loci that determine phenotypic differences in palmitic acid concentration was hybridized with the soybean FAT-B cDNA probe. Normal 16:0 (FAP), High-16:0 (fap2), Low-16:0 (fap1 and/or fap-nc). The recessive fap1 allele contains a point mutation. The recessive fap-nc allele has a null mutation in a second copy of the gene that encodes 16:0-ACP thioesterase.

the DNA fragments from these lines that hybridize to the FAT-B probe. No differences in fragment size or DNA sequence are detected in this gene from normal and high-16:0 soybeans. This suggests that the *fap2* allele (for high-16:0) does not encode the 16:0-ACP thioesterase (5). Although this gene from soybeans carrying the *fap1* allele (for low-16:0) appears to have similar fragment size, sequence analysis shows a 'point' mutation where a leucine residue is substituted for a tryptophan residue near the active site of the enzyme (unpublished). This subtle change causes reduced synthesis of 16:0-CoA. Hence, the recessive *fap1* allele encodes the 16:0-ACP thioesterase.

The *fap-nc* allele is associated with a different type of mutation in a second copy of the 16:0-ACP thioesterase gene. Part of this gene is totally missing. This is called a 'null' mutation. It is a natural gene deletion event. And, it is heritable. When recessive *fap1* and *fap-nc* alleles are combined through Mendelian gene action, the oil has about a 4-fold reduction in 16:0 concentration compared to normal soybean oil (6). This sort of molecular genetic information helps the plant breeder choose the best parental material for a particular selection goal. It also helps the breeder avoid crosses between lines that have two or more very similar alleles.

2·3 Technology for Finding Genes in a Segregating Population

Making a cross and analyzing a segregating population starts a lengthy process of sorting out and selecting the progeny that contain desired homozygous gene combinations for a given genetic trait. This process typically involves growing the population for at least three generations. To expedite selection, DNA markers to help breeders find desired genes for specific traits at early stages of inbreeding. These markers define the location and position of a gene on genomic maps, and serve as guideposts for all genes in a given linkage group or chromosome.

One way to use this marker in a breeding program is to assemble many gene markers in a micro-satellite array. A single array may be used to test genetic variation at 250 gene loci with SSR markers and the DNA from 50 progeny lines of a soybean population that is segregating for each of marked gene. The reaction identifies progeny with homozygous alleles of interest with DNA from leaf tissue of F1 plants or DNA from F2 seed. Identification of lines with homozygous genes allows the breeder to proceed directly to pure inbred line selection by a method known as 'single-seed decent'. This innovation can trim up to 3 years off the timeline for soybean variety development. Hence, gene markers may significantly improve breeding efficiency.

As an example, the inheritance of genes for low-16:0 plus low-18:3 concentration involves at least 3 different recessive alleles. These alleles have been combined in a new soybean variety, called 'Satelite' (**Table 3**). The crude oil has less that 7% total saturates and should not require hydrogenation for many food applications. Tests of how these genes interact with the rest of the genome show that agronomic traits are unaffected by changes in 16:0 and 18:3. Hence, this justifies moving this line along toward commercial production. Indeed, yielding ability of low-16:0 low-18:3 soybeans is very competitive with established normal varieties. Yielding ability is always a deciding factor. Major improvement of oil quality is of little value, if the variety has poor yielding ability. The cv 'Satelite' and similar low-16:0 low-18:3 soybeans are being developed by 12 breeding programs in the US (including Monsanto). In a few years, this type of soybean will be available to farmers in all of the US production areas from Minnesota to Georgia.

2·4 Increasing Oleic Acid Concentration in Soybean Oil

At the same time, these breeding programs also are working on the next innovation in soybean oil quality: the addition of genes to elevate oleic acid (18:1) in low-16:0 low-18:3 soybeans. Soybean oil with higher 18:1 should be available in the next 3 to 5 years. The source of the high-oleic alleles is a germplasm line called, N97-3363-4 (**Fig. 4**). This germplasm that was developed by the USDA-Agricultural Research Service at Raleigh, North Carolina though natural gene recombination. This line and its descendents are the only non-transgenic high oleic soybean breeding lines in existence.

Table 3 Fatty Acid Composition and Yielding Ability of a New Soybean Variety Exhibiting Low-Palmitic and Low-Linolenic Acid Concentration.

Line	Phenotype	16:0	18:0	18:1	18:2	18:3	Yield
		%crude oil					*bu/A*
Brim	Normal	11	4	23	53	9	47
Satelite	Lo-Sat Lo-Lin	3	3	37	54	3	51

16:0, palmitic acid; 18:0, stearic acid; 18:1, oleic acid; 18:2, linoleic acid; 18:3, linolenic acid

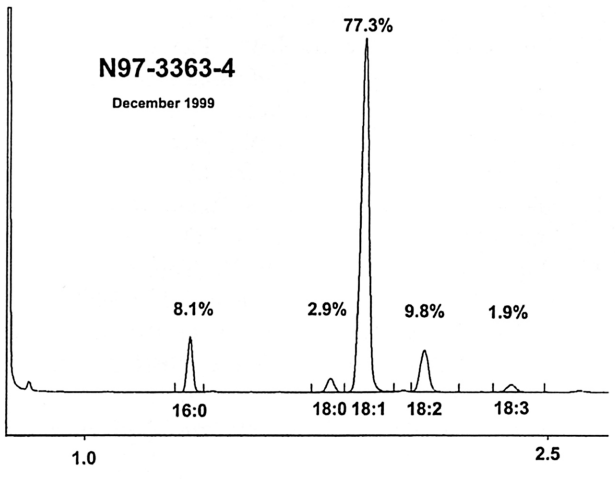

Fig. 4. Fatty Acid Composition of a Non-Transgenic High-Oleic Soybean.

As mentioned earlier, the low-16:0 and low-18:3 traits are to three different gene mutations. The recessive *fap1* and *fap-nc* alleles restrict 16:0-CoA formation at the same enzymatic step, the 16:0-ACP thioesterase. In addition, a recessive *fan* allele reduces conversion of 18:2 to 18:3 at the n-3 desaturase. However, two recessive *fad* alleles determine the high-18:1 phenotype in soybean. The enzyme products of these alleles restrict desaturation of 18:1 to 18:2 at the n-6 desaturase. Thus far, there have been no obvious problems in combining all five of these alleles. Yet, we must be aware of how gene expression is influenced by environmental factors. As an example, a positive relation often exists between 18:1 and growth temperature as the seed develop. Under controlled growth conditions, 18:1 levels in N97-3363 may increase with higher temperature, within a range from about 55 to 75% 18:1. There are many theories as to why 18:1 in plants varies with temperature (7). None explain this phenomenon. However, a new approach may show the biological basis for this phenomenon, and provide a useful means to stabilize desaturase activity against the influence of environmental effects. Specifically, this work involves directed control of metabolic enzyme activity

2·5 Directed Regulation of Enzyme Activity

Many metabolic enzymes have highly conserved regions that contain phosphorylation sites, like this serine residue (8). These sites may become phosphorylated in response to a signal transduction event, such that might arise from a change in growth temperature. These sites are highly conserved among crop species, and there is a strong association between the activity of the enzyme and whether or not the serine residue in this domain is esterified with phosphorus. It is proroposed that directed control of an enzyme activity may then be accomplished by creating a point mutation that removes or replaces Serine at the phosphorylation site. In practice, when this Serine is replaced in the structure of acetyl-CoA Carboxylase (ACCase-I), activity of the enzyme is no longer affected by changes in growth temperature (unpublished). We soon will see if this same approach works for the enzyme that converts 18:1 to 18:2.

To do this, we may employ techniques like Chimeraplasty or Native Gene Surgery (9). These methods are based on a natural process, the DNA repair mechanism, and may be used to create natural gene mutations (via replacement, removal or addition of DNA nucleotides. The simplest example is a 'point-mutation', where a key amino acid is replaced with another. A chimeric RNA/DNA construct may be used to make a recessive allele in a target gene. DNA of the construct emulates the sequence of one allele in the target gene, except it has (for example) a codon for glutamine instead of proline. A chimera or a hair-pin loop of RNA attached to this DNA sequence also contains the complementary codon for glutamine. When inserted in a plant cell, this structure finds the target gene and binds to it. During gene replication, the DNA repair mechanism recognizes that the codons for glutamine and proline do not match. The codon for glutamine is deemed to be correct, and the codon for proline is replaced in both strands of the new DNA molecule. Hence, a site-specific mutation is made in a specific gene. The mutation then segregates in a breeding population, as would any other genetic trait.

3 Conclusion

In conclusion, biotechnology is helping make it possible to develop soybeans with a wide range in oil composition. The first to be marketed will most likely have an oil composition that serves as a low saturated substitute for partially hydrogenated soybean oil. However, the main point we wish to emphasize is that it takes a team effort to achieve such goals, not just one discipline alone. Biotechnology has three main components, plant breeding, genomics and molecular genetics. Information from these disciplines will ensure that biotechnology continues to help us

develop new products, and respond to consumer preferences in a socially acceptable manner.

References

1. USDA, ARS, National Genetics Resources Program. (1999) Available: www.ars-grin.gov/cgi-bin/npgs/html/obvalue.pl?51083.
2. Carter, T.E., Jr., Nelson, R., Cregan, P., Boerma, H., Manjarrez-Sandoval, P., Zhou, X., Kenworthy, W. & Ude, G. (2000) *Proc. Am. Seed Trade Assoc.,* **20**, 35-42.
3. Pantalone, V.R., Rebetzke, G.J., Burton, J.W. & Wilson, R.F. (1997) *J. Am. Oil Chem. Soc.,* **74**, 159-163.
4. Pantalone, V.R., Rebetzke, G.J., Wilson, R.F. & Burton, J.W. (1997) *J. Am. Oil Chem. Soc.,* **74**, 563-568.
5. Wilson, R.F., Burton, J.W., Pantalone, V.R. & Dewey, R.E. (2000) *Lipid Biotechnology*, Marcel Dekker, New York, in Press.
6. Wilcox, J.R., Burton, J.W., Rebetzke, G.R. & Wilson, R.F. (1994) *Crop Sci.* **34**, 1248:1250.
7. Martin, B.A., Wilson, R.F. & Rinne, R.W. (1986) *J. Am. Oil Chem. Soc.,* **63**, 346-352.
8. Huber, S.C., McMichael, R.W., Toroser, D., Bachmann, M., Athwal, G.S., Winter, H. & Huber, J.L. (1998) *Protein Phosphorylation in Plants*, Imprime Univ de Paris-Sud, Orsay, France, ISBN: 2-9512511-0-6. pp. 101-107.
9. Hohn, B. & Puchta, H. (1999) *Proc. Natl. Acad. Sci. USA,* **96**, 8321-8323.