Genomics Approaches to Lipid Biosynthesis

John Ohlrogge[1], Vandana B Mhaske, Frederic Beisson, Sari Ruuska

Department of Plant Biology, Michigan State Univ, East Lansing, MI 48824-1312
[1]http://www.plantbiology.msu.edu/ohlrogge.shtml Email: ohlrogge@msu.edu

Abstract
The vegetable oils produced by soybean, palm, canola and other crops provide approximately 25% of the calories consumed by industrial nations. In addition to their dietary significance, vegetable oils are a major agricultural commodity, with worldwide production of 90 billion pounds, worth nearly $50 billion per year. This large market size and the fact that the fatty acid composition of vegetable oils influences both their physical properties and nutritional characteristics has attracted considerable interest toward modifying plant fatty acid production for both food and non-food uses. Many successes have already been achieved in altering the chain length and saturated/unsaturated fatty acid content of dietary plant oils by transgenic methods. Future efforts will lead to plant oils rich in omega-3 structures found in fish oils. Genomic approaches, including EST sequencing, microarrays and bioinformatics are now contributing greater understanding of the underlying metabolism of oilseeds and the regulatory networks that determine the quality and quantity of oils produced.

Media summary
The genetic engineering of plant oils has proven to be technically feasible, with successes in altering the fatty acid profiles of several crops, including soybeans, canola, and cotton.

Key words
Metabolic engineering, oleic, linolenic, nutrition, omega-3

Introduction

The increasing role of vegetable oils

The oils produced by plants provide a major source of calories for human nutrition. In addition, they provide the most concentrated source of energy of any food, supply essential fatty acids (which are precursors for prostaglandins and other similar hormones), are carriers for fat soluble vitamins, and serve to make foods more palatable and to the feeling of satiety after eating.

Consumption of fats differs drastically in different parts of the world. As individuals and societies become more affluent, the proportion of calories from fats and oils in the diet tends to increase. Country-specific analysis of data for 1988-1990 (FAO, 1994) found a range of 7-46% for the % of dietary calories from fat. A total of 19 countries fell below the minimum recommendation of 15% dietary energy supply from fat, the majority of these being in sub-Saharan Africa and the remainder in South Asia. In contrast, 24 countries were above the maximum recommendation of 35%, the majority of these countries being in North America and Western Europe. An additional trend relating to fat/oil consumption has been the transition from consumption of animal to vegetable sources for fat. For example, in 1920 animal sources were responsible for 81% of fat calories consumed in the U.S., whereas by 2000, vegetable oils provided over 54% of fat/oil calories (Figure 1)**.** The major uses of vegetable oils in the U.S. as reflected by industry shipments are cooking, frying and salad oils (47%), shortening (45%), and margarines (5%) **http://www.iseo.org/statistics.htm**

26 Sep – 1 Oct 2004, Brisbane, Australia. Published on CDROM. Web site www.cropscience.org.au

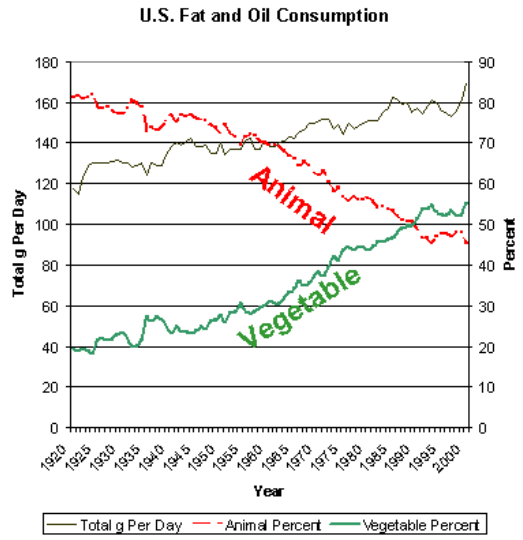

Figure 1: U.S. fat and oil consumption from 1920 to 2000. Data from http://www.usda.gov/cnpp/nutrient_content.html **and** http://147.208.9.134/default.htm

Major oil crops and their fatty acid composition

The major crops which produce approximately three-fourths of vegetable oil for human consumption are soy, palm, canola and sunflower (Figure 2). However, a much larger number of oil crops are consumed because of their special properties, or are important in local economies. For example, sesame, olive and peanut oils are often used for the special flavours they impart to foods whereas cottonseed provides a major edible oil in Australia and China and other regions where cotton is a major crop and the seeds are a low-cost by-product of the fibre. For large-scale use in baking and frying, the food industry chooses commodity vegetable oils largely based on price, which is related to the location of production and whether the oil is a by-product. For example, in the U.S. soybean oil is often considered a by-product of the soy meal industry, and its 80% "market share" dominates the vegetable oil market, making it the largest single source of calories in most U.S. diets. In contrast, palm and rapeseed are crops grown almost exclusively for their oil and are particularly prevalent in Asia and Europe, respectively, near their site of production.

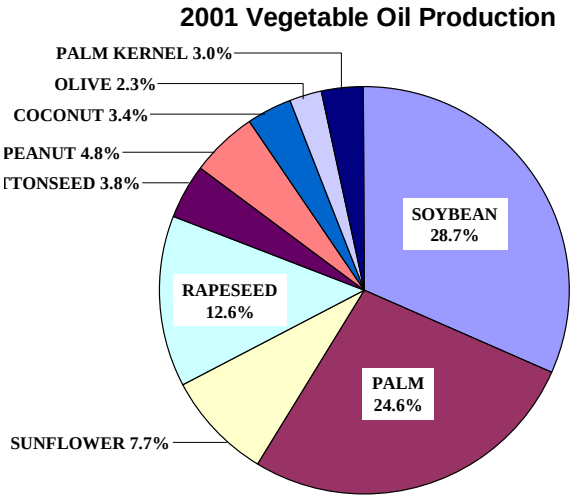

Figure 2: Major world oil crops. Data from www.fas.usda.gov/oilseeds

Although hundreds of different fatty acid structures can occur in seeds of different plant species, only four fatty acids; palmitic, oleic, linoleic and linolenic make up 95% or more of the composition of the major edible vegetable oils (Figure 3). However, the relative proportions of these fatty acids influence both the physical properties and the nutritional impact of the oils. For example, the saturated fatty acid content has a major impact on the melting point of the oil or fat. To produce shortening and margarines, in most cases vegetable oils are hydrogenated which increases their saturated fatty acid content, provides more solid physical properties and also leads to side reactions which yield trans fatty acids. Saturated and *trans*

fatty acids also influence the nutritional properties of the oil, leading to increased cholesterol levels. Other issues related to unsaturated fatty acid content of vegetable oils are discussed below.

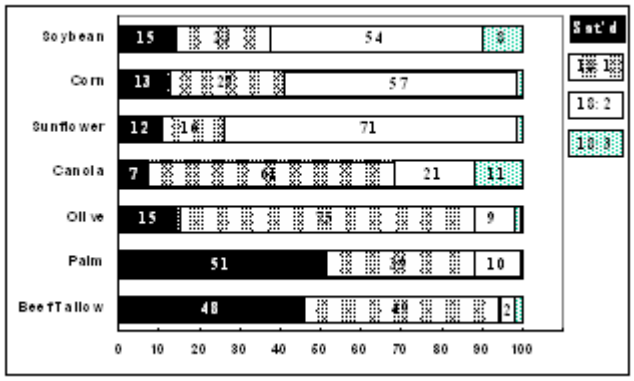

Figure 3: Fatty acid composition of major vegetable oil crops compared to beef tallow.

Dietary issues: With fat, you are what you eat.

Although human tissues have the capacity for fatty acid synthesis, when fatty acids are abundant in the diet this biosynthetic capacity is suppressed. Consequently, most of the fatty acids present in human tissues are derived from the diet rather than from *de novo* synthesis. However, after consumption, dietary fatty acids can be acted on by selective enzymes which target certain structures to specific membrane components (or to other lipids) and they can also be modified in their structure by further desaturation, elongation or beta-oxidation (chain-shortening) reactions. Nevertheless, despite these selections and modifications, the lipid composition of humans is strongly influenced by the fatty acid composition present in the diets. One example of how the composition of human storage and membrane lipids is influenced by the diet is shown in Figure 4 where the distribution of double bond positions of *trans* isomers in humans clearly reflects the distribution found in dietary hydrogenated vegetable oils (Ohlrogge et al. 1982).

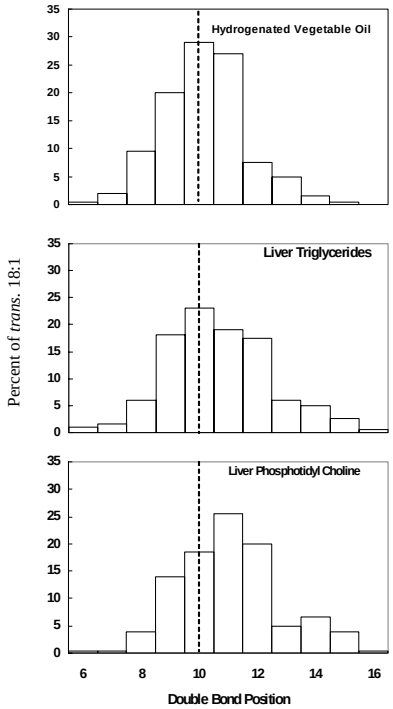

Figure 4. Double bond distribution in the *trans* octadecenonate fraction of hydrogenated vegetable oils and from human tissue lipids. Data from Ohlrogge et al, 1982.

Because different human populations have very different diets, and because these same populations have very different causes of mortality, some members of the medical research community conclude that dietary fatty acids have a major impact on diseases (e.g. Wolfram, 2003). For decades, an emphasis on risks associated with saturated fatty acids and its relation to cholesterol metabolism has encouraged the consumption of more vegetable oil rather than animal fats. More recently, concern has increased about the role of the *trans* fatty acid isomers which are produced during the hydrogenation of vegetable oils and may have negative consequences. Also, in recent years evidence has accumulated that the balance of omega-3 and omega-6 unsaturated fatty acids in diets influences a wide range of human physiological responses including coronary heart disease (CHD). The dominance of plant oils with high omega-6 18:2 in many diets (e.g. U.S.) has led to omega-6/omega-3 consumption ratios near 10:1 in many western diets whereas populations which consume ratios near 1:1 (e.g. Greenland, Japan) have strikingly lower incidence of CHD. Possibly these different fatty acid compositions in diets may in part be causally related to the very different CHD levels shown in Table 1.

Table 1. Deaths due to cardiovascular disease in populations with different omega-6/omega-3 content of diets.

Population	Cardiovascular Deaths per 100,000
U.S.	200
Mediterranean	90
Japan	50
Greenland	20

One explanation offered for the strikingly different CHD mortality associated with high omega-6/omega-3 consumption ratios is the different fates of these structures in the biosynthetic pathways leading to prostaglandins and other eicosenoids. The omega-6 fatty acids are precursors to arachidonic acid which in turn produces prostaglandins and thromboxanes which promote blood clotting. In contrast, the omega-3 fatty acids are poor substrates for prostaglandin biosynthesis and act as competitive inhibitors of the arachidonic pathway.

The above issues regarding the potential influence of dietary fatty acids on human health have stimulated interest in use of genetic engineering to design healthier plant oils. Such efforts have led to some impressive technical successes in reducing the saturated fatty acid and omega-6 content of edible oils and in the future may provide abundant and inexpensive alternatives to fish oil for omega-3 fatty acids.

Genomic approaches: Why genomics has been used to study plant fatty acid synthesis
In addition to their nutritional significance, vegetable oils are a major agricultural commodity, with worldwide production of 90 billion pounds, worth nearly $50 billion per year. This market size has attracted considerable interest to modify plant fatty acid production for both food and non-food uses.

Major aims for molecular genetic improvement of plant oils have been to improve the total oil content of seeds, and to increase the content of specific target fatty acids in the oil to improve nutritional or functional properties and expand uses. In several instances, a functional genomics approach has been successfully used to identify the genes responsible for the production of specific fatty acids. More recently, the development of genomics technology has opened doors for large-scale approaches aiming to uncover and understand the regulatory factors that govern synthesis and accumulation of fatty acids in tissues. In the following, we will discuss examples of these approaches.

Lipid metabolism in the model plant Arabidopsis thaliana: a first draft of the metabolic maps and a catalog of the proteins/genes possibly involved
There is now a growing line of evidence that cellular metabolism is spatially organized and not just a collection of enzymes, substrates and effectors randomly dispersed in the aqueous phase of the cell. Evidences include enzyme arrays adsorbed to membranes or the cytoskeleton, multienzyme complexes and even metabolite pools that are not in free solution. Moreover, enzymatic reactions can often be catalyzed by several isoforms (about 2/3 of the genes in the model plant *Arabidopsis thaliana* belong to multiple gene families). The genes encoding different isoforms may correspond to a true functional redundancy, but could also bear specific information in their coding or regulatory sequences. Thus, different isoforms might be expressed in different tissues or organs at different stages, targeted to different organelles or subdomains within an organelle, regulated by different effectors and may interact

specifically with other proteins or cell structures. Therefore, only a set of specific isoforms may be relevant to a trait of interest and their identification can be a critical step in designing successful strategies of rational metabolic engineering.

For the goal of engineering metabolic traits, the first step is to make a map of all possible reactions and a catalog of the proteins/genes that may be involved in each reaction. In recent years, this task has been greatly facilitated by the sequencing of entire genomes. *Arabidopsis thaliana* was a fortuitous choice for oilseed researchers to make such a catalog of reactions/enzymes because this plant is:

- the best-studied model organism for plant biologists and the first plant to have its genome completely sequenced
- an oil seed plant (about 35% of lipids per seed dry weight)
- a species closely related to a major oil crop (canola).

A first survey of reactions, proteins and genes that may be involved in the metabolism of fatty acids and lipids derived from fatty acids (acyl-lipids) in Arabidopsis has been performed in our lab and reported before the complete sequencing of the Arabidopsis genome (Mekhedov et al 2000). This partial survey covers 65 reactions and compares Arabidopsis with rice. Based on this initial survey, we have recently reported a more complete catalog based on the fully sequenced genome, which covers most (120) reactions of acyl-lipid metabolism in Arabidopsis (Beisson et al. 2003).

The gene catalog is based on sequence similarity searches between the proteins predicted from the Arabidopsis genome and proteins from any organism that have been shown experimentally to be involved in a reaction of acyl-lipid metabolism. The candidate proteins have been subjected to multiple alignments and searches for conserved protein motifs. Over 600 proteins and encoding genes covering more than 120 reactions have been thus identified and classified according to predicted function, subcellular location and alternative splicing. This catalog is part of the Arabidopsis Lipid Gene Database that we have made available on the Web (http://www.plantbiology.msu.edu/lipids/genesurvey/index.htm). Candidate proteins with links to sequence databases and literature are available for each reaction by browsing either metabolic maps or a list of reactions of lipid metabolism sorted by organelles and main cellular functions.

Only 13% of the 644 predicted proteins have been experimentally demonstrated to have lipid-related activity, so the vast majority of the genes represent uncharacterized proteins. Moreover, about 1/3 of these proteins were previously annotated as “unknown function” or with functions unrelated to acyl lipid metabolism. For each reaction, the catalogue gives a list of most probable candidate genes. For example, regarding lipid synthesis, there are 4 main candidate genes for the endoplasmic 2-lysophosphatidate acyltransferase and 7 additional ones. This enzyme catalyzes the formation of phosphatidic acid, a key intermediate in membrane and storage lipid assembly. It is not known if phosphatidic acid and other precursors are synthesized by the same enzymes and at the same subcellular location for these two types of lipids. One of these acyltransferases might thus be specifically involved in oil accumulation in seeds.

In addition to the catalog, the database gives information about the distribution of ESTs in various organs for all genes of the catalog. A statistical analysis of EST abundance in organs is provided and lists of enzymes that are possibly up regulated in some organs is given. A tentative list of enzyme isoforms involved in storage lipid synthesis in seeds has for example been suggested (Beisson et al. 2003). Clearly these lists of differentially expressed genes are only a first step in the search for isoforms of biotechnological interest and must be checked by experimental methods (northern blots, etc.).

It is expected that this Web-based database will provide a first common platform of information about reactions/enzymes of acyl-lipid metabolism in plants and that will facilitate functional studies in Arabidopsis and also in crop plants. This database can be particularly helpful in view of proteomics studies, where it will be important to recognize and distinguish different polypeptides from gene families for a particular reaction of acyl-lipid metabolism. It will also guide the search for knockout phenotypes and help in the design of gene knockout strategies as well as in the analysis of *in vivo* metabolic fluxes in mutants of lipid synthesis.

EST sequencing projects and gene discovery
ESTs (expressed sequence tags) are defined as single-pass DNA sequences from cDNA clones. Initial success was achieved in targeted gene discoveries using small scale EST sequencing projects where the tissues examined were enriched for tissue-specific transcripts. For example, genes responsible for production of ricinoleic acid and conjugated fatty acids were identified and cloned from ESTs prepared from oil producing tissues of castor bean, *Momordica charantia* and *Impatiens balsamina* (Van de Loo et al., 1995, Cahoon et al., 1999). These fatty acids are industrially important for production of nylons, paints etc.

In recent years, due to greatly lower costs and improvements in DNA sequencing, larger scale sequencing projects have become common. The data generated from high throughput genomics can be used for annotating genes and understanding their spatial and temporal regulations. This information would be particularly helpful in understanding traits governed by multiple genes, especially quantitative traits. Furthermore, advanced computer aided sequence analysis and annotations have made it possible to use EST datasets to study activities of many genes simultaneously and to understand their interaction (e.g. Lee et al, 2002).

One of the first large-scale EST collections for plant seeds was generated at Michigan State University (White et al. 2000). Metabolism of seeds is tailored toward the accumulation of economically valuable storage compounds such as oil and the EST collection therefore provided an opportunity to simultaneously study the transcription of various genes involved in biosynthesis and accumulation of plant lipids. A cDNA library was constructed from developing Arabidopsis seeds (between 5 and 13 days old) actively accumulating storage compounds. After subtraction of the library using highly abundant mRNAs (such as seed storage proteins and oleosins) altogether 10 500 clones were sequenced. These ESTs were annotated and about 40% did not have a match in other databases and thus at the time represented novel genes expressed possibly only in seeds. Recently, the ESTs have been mapped to the Arabidopsis genome, and they represent about 3500 genes.

"Data mining" of this EST data provided insights into the import of photosynthates into developing embryos, its conversion into seed oil, and the regulation of this pathway which could be keys to understand what determines oil content of a seed. The main findings of this exercise can be summarized as follows.

In general the relative abundance of the cDNAs encoding different enzymes of fatty acid synthesis was found to be similar in the seed and non-seed EST sets, suggesting that seeds do not alter to a substantial degree the relative expression of genes encoding pathway components to accomplish the increased flux through the pathway in seeds. Rather, the entire pathway is apparently up-regulated, as suggested by the overall higher relative abundance of ESTs noted for fatty acid synthesis in the seed compared with the non-seed sets (Mekhedov et al., 2000). The data suggest that at least nine genes encoding enzymes or subunits involved in this pathway are co-ordinately regulated. This fact would not have been uncovered in studies involving single genes.

ESTs encoding cytosolic glycolytic enzymes were more abundant than ESTs of their plastid counterparts. Transcripts for transporters of phosphoenolpyruvate (PEP) but not for pyruvate could be identified. Together these findings led to a hypothesis (White et al, 2000) that a major route of carbon in developing Arabidopsis seeds involves cytosolic glycolysis from sucrose to PEP, and then PEP is imported to plastids via its transporter and finally converted to pyruvate and acetyl-CoA. This data set also identified >500 known and putative regulatory factors (transcription factors, protein kinases + phosphatases) which represent new, potential candidates for seed-specific metabolic regulators

ESTs for other oilseeds:
Soybean and sesame are two important oilseeds for which large EST databases are publicly available. Soybean (*Glycine max*) is the world's largest source of vegetable oil and a large public genomics initiative has been developed for soybean. (soybean genomics initiative, http://soybean.ccgb.umn.edu). As of November 2003 this database contains 344,524 EST sequences originating from over 80 different libraries. This includes ESTs from various sources such as leaves, flowers, shoots, seeds, and somatic embryos. Together they constitute a rich resource for this important agronomic crop.

Sesame (*Sesamum indicum*) is a highly-valued oilseed crop of Asia that produces seeds with 50% of dry weight oil, and which is commonly used in oriental cooking. Sesame oil also contains lignans that act as antioxidants, stabilizing the oil. In order to identify genes involved in storage product accumulation, as well as lignan synthesis, a total of 3328 ESTs from 5-25 days old immature sesame seeds were sequenced (Suh et al., 2003). To search for novel or seed-specific sequences, the ESTs were analyzed by BLASTX to the non-redundant protein database of GenBank, and also compared against the Arabidopsis proteome. The project identified several putative genes or specific isoforms of fatty acid synthesis gene families that may be characteristic to sesame seed. Differences in the expression levels of different isoforms encoding seed storage proteins and oleosins were also observed in comparative analysis of ESTs between sesame and Arabidopsis developing seeds. It was further noted that there is a relatively high expression of cytosolic malate dehydrogenase and low expression of plastidic pyruvate kinase in sesame. These observations suggest that a major pathway of carbon flux for plastid fatty acid synthesis in sesame may involve synthesis of malate in cytosol and its import to plastids, as opposed to PEP import as suggested for Arabidopsis.

ESTs clearly can serve as an important tool for data mining and gene annotations. Comparison of ESTs across different tissues, developmental stages; validated with appropriate statistical analysis could provide insights into regulation of gene(s) and their interactions. However, it is important to recognize various limitations to conclusions based on the EST approach. For example, it is hard to identify genes with low expression levels, such as many transcription factors. Furthermore, even cDNAs of a central fatty acid synthesis enzyme enoyl-ACP reductase were not found in the Arabidopsis EST collection, although the mRNA must be present in seed tissue. Moreover this approach does not take into account other aspects of regulations such as post-translational modifications, differences in enzyme activities etc. In summary, although EST abundance is an imperfect estimate of gene expression they can lead to useful preliminary insights and hypotheses, which can be further tested by other experiments.

Microarrays for large scale gene expression analysis

The technology of DNA microarrays allows for the first time the simultaneous measurement of thousands of transcripts, and direct comparison of RNA populations from different sources. This technology has revolutionized the ability of researchers to think broadly about gene expression research. Recently, oligonucleotide based arrays have been developed with the complete genome of Arabidopsis represented. Expression profiling of seeds active in fatty acid synthesis has attracted less attention, but a few attempts are described here.

(a) Arabidopsis Developing Seed Microarrays

The first plant fatty acid-related transcriptional profiling project was derived from the MSU Arabidopsis developing seed EST project described above. For the first generation arrays about 3000 DNA elements were chosen for the construction of cDNA based microarrays (summarized in Girke et al., 2000). These microarrays were initially used for tissue profiling in Arabidopsis to find genes that are expressed highly in seeds by comparing RNA isolated from developing seeds, leaves and roots. The microarray experiments demonstrated that most seed-expressed transcripts, including lipid biosynthesis related genes, were also expressed in other tissues and so could not be classified seed-specific. Of the >100 lipid genes studied, only 10 were found to be expressed at least 10 fold more in seeds. These genes did not include actual biosynthetic genes, but were more related to FA modification or (such as fatty acid elongase or oleosins). However, a subset of >250 highly seed-specific transcripts were identified, and provided candidates for further studies.

A second generation of Arabidopsis seed cDNA microarrays contained ~6000 DNA elements, which were later mapped to about 3500 Arabidopsis genes. These microarrays were used in time-course experiments to study the temporal patterns of gene expression during Arabidopsis seed filling, as well as comparing gene expression between wild type seeds and seeds from a mutant with a defect in seed storage oil accumulation (wri1, Focks and Benning, 1998). Some major observations from this study are listed below; for details see Ruuska et al. (2002):

- Genes associated with different storage products had distinct expression patterns (see Figure 5). For example, transcripts for many core fatty acid synthesizing enzymes had a bell-shaped expression

pattern, and began to be down-regulated during the active oil synthesis. Examples of this group include ACCase subunits, KAS I and FAD2 (oleate desaturase).

- In contrast, transcripts for some fatty acid modifying enzymes, such as linoleate desaturase (FAD3), elongase (FAE1), as well as for oleosins and storage proteins were induced later and remained high throughout the time course.
- Transcripts for starch metabolism enzymes such as amylases were highest in young seeds, which also transiently accumulate starch.
- Proteins involved in carbon delivery through cytosolic glycolysis via PEP to plastid fatty acid synthesis were co-ordinately expressed. The expression of many proteins (but not all) involved in sucrose import, metabolism and cytosolic glycolysis increased towards seed filling or followed the profile of FAS enzymes. In addition, the expression of PEP transporter followed the profile of the core FAS enzymes, as did the plastid pyruvate dehydrogenase, further strengthening the idea that cytosolic carbon delivery for FAS and fatty acid synthesis enzymes are co-ordinately expressed.
- Several lipid- and carbohydrate metabolism related genes were down-regulated in the *wri1*- mutant with reduced seed lipid content. Most of these genes belonged to the group with a bell-shaped expression pattern during seed development, and included BCCP, KAS I, FAD2, and chloroplast pyruvate dehydrogenase E1a subunit. In contrast, for genes such as oleosins and fatty acid elongase that are involved in lipid synthesis but did not belong to the bell-shaped expression group were not altered in *wri1*.

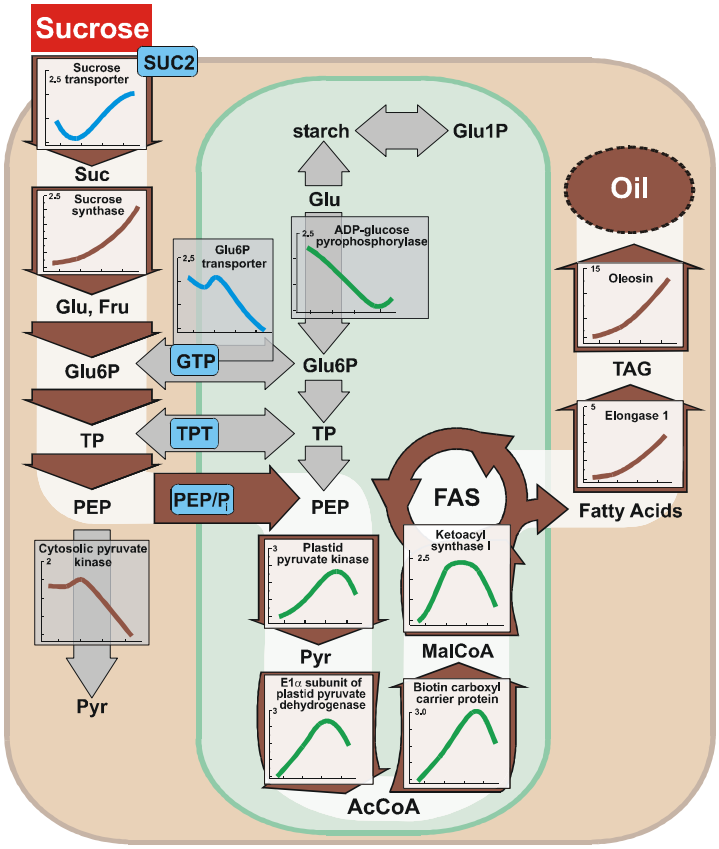

Figure 5. Temporal Patterns of Gene Expression for Proteins Involved in Conversion of Sucrose to Oil in Developing Arabidopsis Seeds. Adapted from Ruuska et al. (2002). The inserts show changes in relative levels of expression of transcripts between 5 and 13 days after flowering. Microarray expression patterns for enzymes in the cytosol are shown with red lines, reactions in the plastid with green lines and membrane transporters in blue. Grey arrows indicate reactions or transporters whose expression decreases during development and are believed to carry less flux than the major pathway represented by brown arrows. The shift during development from starch biosynthesis and cytosolic metabolism of PEP toward oil biosynthesis and plastid uptake and metabolism of PEP is reflected in the coordinated expression patterns of enzymes and transporters. Not shown in this figure are additional contributions to oil metabolism such as dark and light reactions of photosynthesis, the oxidative pentose phosphate pathway and others.

(b) Brassica napus and Maize Embryo FAS Gene Analysis

Recently O'Hara et al. (2002) used RNA blots and DNA filter arrays together with EST analysis to study the expression of a group of fatty acid synthesis genes in developing *Brassica napus* embryos. They found that mRNAs for the biotin carboxylase subunit of plastid ACCase, as well as 3-ketoacyl-ACP reductase, enoyl-ACP reductase, and thioesterase all accumulated in a coordinate manner. Their expression profile peaked around 3 and 4 weeks after flowering, and dramatically decreased afterwards. These expression patterns were similar to the reported mRNA and protein levels of enoyl-ACP reductase and stearoyl-ACP desaturase in Brassica seeds (Fawcett et al., 1994) and also to many FAS enzymes in Arabidopsis seeds (Ruuska et al., 2002). This study again highlighted the fact that increased demand for fatty acid synthesis during embryogenesis was met with the up-regulation of a group of synthesis enzymes. In addition, the transcripts were present in constant molar ratio, such that at all times ACP mRNAs were most abundant, and those of the FAS complex were similar to each other's. The results agree with earlier EST comparisons and a study of developing Arabidopsis siliques, which showed that the accumulation of mRNAs encoding the subunits of the heteromeric ACCase subunits is temporally correlated (Ke et al., 2000).

Expression of FAS related genes has also been studied from developing maize embryos (Lee et al., 2002). Transcript profiles were quantitated using glass based cDNA arrays, nylon membrane arrays, RNA blots, and EST abundance analysis. Similar to previous studies, several FAS genes (ACCAse, KAS III, ACPs, desaturase) exhibited a coordinated expression pattern such that their expression peaked several days before the maximal fatty acid accumulation would be expected. Several oleosin isoforms had a different profile such that their expression was induced at a later stage. The authors also clearly demonstrated that, for a given gene, ESTs numbers from libraries constructed at different stages of seed development correlated well with microarray results. The mean correlation coefficient between electronic northern and nylon array expression profiles was 0.86, and between electronic northern and glass array expression profiles was 0.84. This outcome provided an independent confirmation that ESTs and microarray signals both reflect transcript abundance. Together these studies on Arabidopsis, *Brassica napus* and maize all demonstrate that the up-regulation of several fatty acid synthesizing enzymes during seed development seems to be under similar and coordinated control.

Oilseed metabolic engineering; successes and challenges

Plant lipid metabolism has proven particularly amenable to modifications both by traditional breeding approaches (Ohlrogge et al, 1991) and more recently by transgenic manipulations (for reviews see for example, Voelker and Kinney, 2001, Thelen and Ohlrogge, 2002; Schultz and Ohlrogge, 2001). Many technical successes have been achieved, including high-laurate canola (Voelker et al, 1996); high-stearate canola (Knutson, 1992); reduced saturates (Kinney, 1996) and production of gamma-linolenic (Sayanova et al, 1999). Soybeans, canola, sunflower, and cotton (e.g. Liu et al, 2002) have all been successfully engineered for improved fatty acid composition. The production of high-oleic soybeans is described in more detail below because of its potential nutritional and industrial impact. Although there have been several impressive technical engineering successes, appearance of transgenic oils on the market has been slowed, in part because of concerns regarding antibiotic resistance markers used to produce the original transgenic crops.

High-oleic soybean oil: a genetically engineered product with benefits for both food and non-food uses.
Most soybean varieties produce an oil rich in polyunsaturated fatty acids - about 50% linoleic acid (18:2) and 10% linolenic acid (18:3). These fatty acids, particularly 18:3, make the oil unstable and easily oxidized. When heated, the oil develops objectionable flavors and odors. Thus unprocessed soybean oil is unsuitable for many applications, and therefore, for most edible uses it is chemically hydrogenated. This process adds to the cost of the oil and also introduces side reactions such as conversion of double bonds from the cis to *trans* configuration creating trans-fatty acids.

The biosynthesis of polyunsaturated fatty acids in plants is catalyzed by a series of enzymes with the first step carried out by an enzyme that converts oleic acid (18:1) to linoleic acid (18:2). In 1994, the gene (FAD2) for this enzyme was isolated in Arabidopsis by screening mutants generated by T-DNA insertions (Okuley et al, 1994). Shortly afterward, molecular biologists at DuPont succeeded in isolating

and suppressing the expression of the gene in soybean. This strategy led to a major decrease of the 18:1 to 18:2 conversion step and almost completely eliminated polyunsaturated fatty acids in the soybean oil.

The new transgenic soybean oil has 85% oleic acid, one of the highest oleic acid contents found in nature. The absence of polyunsaturated fatty acids eliminates the need for hydrogenation to stabilize the oil. Furthermore, an unanticipated benefit of the oleic increase was that the saturated fatty acid content of the oil fell from approximately 15% to less than 8%. The new soybean oil has a composition similar to olive and other high-oleic oils, which are considered to provide health benefits, compared to other plant and animal oils. The fatty acid trait was stable in field trials, and the oil yield of the crop was identical to the control lines. Thus neither the transformation process nor the major change in fatty acid composition was detrimental to the high yield of the soybean line. This example is also instructive because it demonstrates how quickly some discoveries can be translated into new crops. With the resources of a major corporation, genetic engineers only needed less than five years from gene isolation to a field-tested transgenic soybean crop ready for commercialization of a new product (Kinney, 1996a; 1996b).

(c) Possible Health Benefits of High-Oleic Soybean Oil.

As noted above, current medical understanding indicates a strong impact of dietary fatty acids on cardiovascular disease and human health. Consequently, there is much interest in tailor-producing healthier vegetable oils and such products may help to balance consumer opposition to "GMO" foods. As mentioned above, health concerns regarding vegetable oil-derived foods include the presence of saturated (particularly palmitic) and *trans*-unsaturated fatty acids. Industrial hydrogenation increases saturated fatty acid content and also results in production of *trans*-isomers of unsaturated fatty acids that are normally not found in vegetable oils and have been associated with coronary heart disease. For many food applications, vegetable oils with a reduced amount of *trans*-unsaturated and saturated fatty acids are desirable to improve human health. The transgenic high-oleic soybean oil composition provides these benefits in a crop which contributes the major source of fatty acids in American diets. One added consumer benefit to wide future use of the engineered high-oleic oils may be reduction in the pathologies associated with high omega-6 fatty acid consumption, as described above.

Omega-3 Fatty Acid Production in Oil crops.

Due to its prominent role in photosynthetic membranes, the 18 carbon, omega-3 fatty acid, alpha-linolenic ($18:3^{9,12,15}$) is likely the most abundant fatty acid produced in nature. This fatty acid is also found in some oilseeds, although at low concentrations (see figure 4). One notable exception is the oil of linseed (flax) which contains over 50% 18:3. As mentioned above, linolenic acid is easily susceptible to oxidation which initially led to the major use of linseed oil as a “drying oil” in paints, varnishes and other products where the oxidation eventually leads to useful polymerization. However, food uses of high 18:3 linseed were greatly limited by this tendency to oxidize which also led to negative flavors. In the 1980's, Allan Green at CSIRO in Australia used mutation breeding and a clever screening procedure to eliminate two genes responsible for 18:3 biosynthesis (Green, 1986). As a result, important new cultivars of linseed are available that produce an edible oil high in 18:2, and referred to as Linola.
http://www.newcrops.uq.edu.au/newslett/ncnl3-92.htm

Although oils high in omega-3 polyunsaturated fatty acids are highly susceptible to oxidation and development of flavour and odor problems, the recognition of the health benefits of omega-3 rich fish oils has changed the perspective of some segments of the food industry from emphasizing their elimination toward the development of specialty food oils which are rich in omega-3 PUFA. Such oils would be expected to be marketed as nutritional supplements or as specialty ingredients with a premium price that would reflect the storage/stability issues as well as identity preservation costs and consumer demand.

The beneficial impact of fish oils seems to be most related to their content of the long-chain omega-3 fatty acids docosahexaenoic acid (DHA) and eicosenoic acid (EPA). These fatty acids (or their precursors) are derived from the omega-3 fatty acids of photosynthetic membranes present in algae consumed by fish. Although the 18:3 found in plant oils appears to have some health benefits similar to those of DHA/EPA rich fish oils, the 18:3 presumably must be elongated and further desaturated to influence the eiscosanoid biosynthetic pathways. Therefore, DHA and EPA are considered more desirable components than 18:3 to promote health benefits. Several groups around the world use genomics to discover genes in marine algae, mosses and fungi that could be used to produce arachidonic acid, DHA or

EPA in oilseed crops (e.g. Tonon et al., 2002; Domergue et al, 2003) and a number of research and development efforts are underway to introduce these pathways into plant seeds (Wallis et al, 2002; Drexler et al, 2003).

Progress toward increasing seed oil content

For both edible and industrial uses, an increase in seed oil content is desirable and has been a major goal of oilseed engineering. However, to be economically useful, such a change must not come at the expense of overall seed yield or at the loss of other high-value components. For example, soybean is the largest source of vegetable oil comprising 30% of the world market and now constitutes over 80% of all dietary vegetable oils in the U.S. Although termed an oilseed, soybean contains only 18-22% oil on a seed dry weight basis and is grown principally as a high-protein meal for animal feeds. Thus, increasing oil in soybean will in most cases not be useful if it comes at the expense of high-value soy protein which drives the crop's economics. By comparison, most other oilseed crops (except cotton) are grown primarily for their oil and produce seeds with 40 to 60% oil. The wide range of seed oil percentage observed in nature suggests that this pathway might be amenable to metabolic engineering, particularly in 'low-oil' oilseeds, provided the key mechanisms which control oil content are identified.

The committed step for *de novo* FAS is the production of malonyl-CoA catalyzed by acetyl-CoA carboxylase. Malonyl-CoA production is a control point for this pathway based upon analysis of acyl-CoA and acyl-ACP pool sizes (for review see Ohlrogge and Jaworski, 1997). Since malonyl-CoA levels in plastids are very low (less than 10%) in comparison to acetyl-CoA, it seemed logical that up-regulation of ACCase activity might increase fatty acid flux. This has been clearly shown to be the case in *E. coli* (Davis et al., 2000). Evidence that increased malonyl-CoA pools could increase fatty acid production was obtained by targeting a homomeric ACCase to rapeseed plastids (Roesler et al., 1997). Under the control of a seed-specific promoter this chimeric protein resulted in higher ACCase activities and increased oil yield by 3-5% on a seed dry weight basis. However, the small increase pointed towards additional control points for FAS.

Could fatty acid synthase activities also be limiting FAS? Several labs have partly addressed this question by overexpressing enzymes downstream of malonyl-CoA production. The conclusion from these investigations is that upregulation of any one enzyme does not increase flux through FAS. Indeed, overexpression of some activities actually decreased FAS and fatty acid content as observed with the overexpression of a condensing enzyme. Rather than an increase in fatty acid content a 5-10% decrease was observed when a spinach KAS III was expressed in tobacco (Dehesh et al., 2001). In the same report, a *Cuphea* KAS III expressed in rapeseed seed embryos resulted in a 9% decrease in fatty acid content. In a related study, targeting of an *E. coli* malonyl-CoA:ACP transacylase to rapeseed leucoplasts increased this plastid activity up to 45-fold but did not increase fatty acid content (Verwoert et al., 1995).

Based upon the aforementioned and other studies it seems unlikely that the up-regulation of any single fatty acid synthase enzyme will have a major positive effect on FAS flux. Although not all fatty acid synthase enzymes have been overexpressed to determine the effect on FAS, substantial increases in flux will likely require up-regulation of multiple activities. This conclusion has stimulated more comprehensive efforts to identify transcriptional, protein kinase or other regulatory factors that might up-regulate the entire pathway.

Initial studies also suggest that reactions late in the TAG biosynthetic pathway may provide increased sink strength that could stimulate increased fatty acid production. Overexpression of a yeast long-chain sn-2 acyltransferase resulted in >50% (dry mass/seed) increases in seed oil content of Arabidopsis and rapeseed (Zou et al., 1997). Field trials of these transgenic rapeseeds gave increases ranging between 8.1 and 13.5% (Taylor et al., 2002). In addition, Jako et al (2001) reported that over-expression of an Arabidopsis diacylglycerol acyltransferase in Arabidopsis seeds can also increase seed oil content as well as seed weight. Together these studies suggest that increased flux into oil may be more easily achieved by strategies targeted at the latter steps in the pathway.

Conclusions

Although over 60 million hectares of GMO crops were planted in 2003, almost none of these were engineered for "output traits". However, engineering of plant oil synthesis has been technically highly successful and has progressed more rapidly than manipulation of carbohydrate, vitamin or protein quality. Thus, transgenic plants with improved edible oils are available and soon to be commercialized. Although the manipulation of fatty acid chain length and the level of fatty acid desaturation have been demonstrated, engineering plants with major changes in flux to oil, protein or carbohydrate appears more difficult. This is likely due to the complexity associated with the engineering of primary carbon metabolism and an unclear picture of how pathways are regulated in vivo. Thus, we still do not understand the basis for why some crops accumulate high oil contents in seeds (e.g. canola) whereas others accumulate more protein (e.g. soybean). Thus, one of the challenges that lie ahead is to understand the fundamental mechanisms responsible for partitioning of carbon between alternative storage products in seeds.

References

Beisson F, Koo AJK, Ruuska S, Schwender J, Pollard M, Thelen JJ, Paddock T, Salas JJ, Savage L, Milcamps A, Mhaske VB, Cho Y and Ohlrogge JB (2003) Arabidopsis genes involved in acyl lipid metabolism. A 2003 census of the candidates, a study of the distribution of expressed sequence tags in organs, and a web-based database. Plant Physiology 132: 681-697.

Cahoon EB, Carlson TJ, Ripp KG, Schweiger BJ, Cook GA, Hall SE and Kinney AJ (1999) Biosynthetic origin of conjugated double bonds: production of fatty acid components of high-value drying oils in transgenic soybean embryos. Proceedings of the National Academy of Sciences of the United States of America 96: 12935-12940.

Davis MS, Solbiati J and Cronan JE (2000) Overproduction of acetyl-coa carboxylase activity increases the rate of fatty acid biosynthesis in Escherichia coli. Journal of Biological Chemistry 275: 28593-28598.

Dehesh K, Tai h, Edwards P, Byrne J and Jaworski JG (2001) Overexpression of 3-ketoacyl-acyl-carrier protein synthase IIIs in plants reduces the rate of lipid synthesis. Plant physiology 125: 1103-1114.

Domergue F, Abbadi A, Ott C, Zank TK, Zahringer U, Heinz E. (2003) Acyl carriers used as substrates by the desaturases and elongases involved in very long-chain polyunsaturated fatty acids biosynthesis econstituted in yeast. J Biol Chem. 278:35115-26

Drexler H, Spiekermann P, Meyer A, Domergue F, Zank T, Sperling P, Abbadi A, Heinz E. (2003) Metabolic engineering of fatty acids for breeding of new oilseed crops: strategies, problems and first results. J Plant Physiol. 160:779-802.

FAO Food and Nutrition Paper, No. 57: Fats and oils in human nutrition. Report of a joint expert consultation. Rome, Food and Agriculture Organization of the United Nations, 1994.

Focks N and Benning C . (1998) Wrinkled1 - a novel, low-seed-oil mutant of Arabidopsis with a deficiency in the seed-specific regulation of carbohydrate metabolism. Plant Physiology 118: 91-101.

Girke T, Todd J, Ruuska S, White J, Benning C and Ohlrogge J (2000) Microarray analysis of developing Arabidopsis seeds. Plant Physiology 124: 1570-1581.

Green AG (1986) Genetic conversion of linseed oil from industrial to edible quality. Journal of the American Oil Chemists Society 63: 464.

Jako C, Kumar A, Wei YD, Zou JT, Barton DL, Giblin EM, Covello PS and Taylor DC (2001) Seed-specific over-expression of an Arabidopsis cdna encoding a diacylglycerol acyltransferase enhances seed oil content and seed weight. Plant Physiology 126: 861-874.

Ke JS, Wen TN, Nikolau BJ and Wurtele ES (2000) Coordinate regulation of the nuclear and plastidic genes coding for the subunits of the heteromeric acetyl-coenzyme a carboxylase. Plant Physiology 122: 1057-1071.

Kinney AJ (1996) Designer oils for better nutrition. Nature Biotechnology 14: 946.

Kinney AJ (1996) Development of genetically engineered soybean oils for food applications. Journal of Food Lipids 3: 272-292.

Knutzon DS, Thompson GA, Radke SE, Johnson WB, Knauf VC, and Kridl JC (1992) Modification of Brassica seed oil by antisense expression of a stearoyl-acyl carrier protein desaturase gene.

Proceedings of the National Academy of Sciences of the United States of America . USA 89: 2624-2628

Lee J, Williams ME, Tingey SV and Rafalski J (2002) DNA array profiling of gene expression changes during maize embryo development. Functional and Integrative Genomics 2: 13-27.

Liu Q, Singh SP and Green AG (2002) High-stearic and high-oleic cottonseed oils produced by hairpin RNA-mediated post-transcriptional gene silencing. Plant Physiology 129: 1732-1743.

Mekhedov S, De Ilarduya OM and Ohlrogge J (2000) Toward a functional catalog of the plant genome. A survey of genes for lipid biosynthesis. Plant Physiology 122: 389-401.

O'hara P, Slabas AR and Fawcett T (2002) Fatty acid and lipid biosynthetic genes are expressed at constant molar ratios but different absolute levels during embryogenesis. Plant Physiology 129: 310-320.

Ohlrogge JB, Browse J and Somerville CR (1991) The genetics of plant lipids. Biochimica et Biophysica Acta. 1082: 1-26.

Ohlrogge JB and Jaworski JG (1997) Regulation of fatty acid synthesis [review]. Annual Review of Plant Physiology and Plant Molecular Biology, 48:109-136.

Okuley J, Lightner J, Feldmann K, Yadav N, Lark E and Browse J (1994) Arabidopsis FAD2 gene encodes the enzyme that is essential for polyunsaturated lipid synthesis. The Plant Cell 6: 147-158.

Roesler K, Shintani D, Savage L, Boddupalli S and Ohlrogge J (1997) Targeting of the Arabidopsis homomeric acetyl-coenzyme a carboxylase to plastids of rapeseeds. Plant Physiology 113 : 75-81.

Ruuska SA, Girke T, Benning C and Ohlrogge JB (2002) Contrapuntal networks of gene expression during Arabidopsis seed filling. Plant Cell 14: 1191-206.

Sayanova O, Davies GM, Smith MA, Griffiths G, Stobart AK, Shewry PR and Napier JA (1999) Accumulation of delta(6)-unsaturated fatty acids in transgenic tobacco plants expressing a delta(6)-desaturase from *Borago officinalis*. Journal of Experimental Botany 50: 1647-1652.

Schultz D, Ohlrogge J (2001) Metabolic Engineering of Fatty Acid Biosynthesis. In Lipid Biotechnology Kuo, T-M and Gardner, eds. Marcel Dekker, Inc. New York.

Shintani D and Ohlrogge JB (1995) Feedback inhibition of fatty acid synthesis in tobacco suspension cells. The Plant Journal 577-587.

Suh, M, Bae, J, Sim, W, Hur, C, Kang, C and Ohlrogge, JB (2003) Comparative analysis of expressed sequence tags between *Sesamum indicum* and *Arabidopsis thaliana* developing seeds. Plant Molecular Biology 52: 1107 – 1123.

Taylor DC, Katavic V, Zou JT, Mackenzie SL, Keller WA, An J, Friesen W, Barton DL, Pedersen KK, Giblin EM, Ge Y, Dauk M, Sonntag C, Luciw T and Males D (2002) Field testing of transgenic rapeseed cv. hero transformed with a yeast sn-2 acyltransferase results in increased oil content, erucic acid content and seed yield. Molecular Breeding 8: 317-322.

Thelen JJ and Ohlrogge JB (2002) Metabolic engineering of fatty acid biosynthesis in plants. Metabolic Engineering 4: 12-21.

Tonon T, Harvey D, Larson TR and Graham IA (2002) Long chain polyunsaturated fatty acid production and partitioning to triacylglycerols in four microalgae. Phytochemistry 61: 15-24.

Verwoert II, van der Linden KH, Walsh MC, Nijkamp HJJ and Stuitje AR (1995) Modification of *Brassica napus* seed oil by expression of the *Escherichia coli* fabh gene, encoding 3-ketoacyl-acyl carrier protein synthase III. Plant Molecular Biology 27: 875-886.

Voelker T and Kinney AT (2001) Variations in the biosynthesis of seed-storage lipids. Annual Review of Plant Physiology and Plant Molecular Biology 52: 335-361.

Voelker TA, Hayes TR, Cranmer AM, Turner JC and Davies HM (1996) Genetic engineering of a quantitative trait - metabolic and genetic parameters influencing the accumulation of laurate in rapeseed. Plant Journal 9: 229-241.

Wallis JG, Watts JL and Browse J (2002) Polyunsaturated fatty acid synthesis: what will they think of next? Trends in Biochemical Sciences 27: 467-473.

White JA, Todd T, Newman T, Focks N, Girke T, De Ilarduya OM, Jaworski JG, Ohlrogge JB and Benning C (2000) A new set of Arabidopsis expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil. Plant Physiology 124: 1582-1594.

Wolfram G (2003) Dietary fatty acids and coronary heart disease. European Journal of Medical Research 8: 321-324.

Zou JT, Katavic V, Giblin EM, Barton DL, Mackenzie SL, Keller WA, Hu X and Taylor DC (1997) Modification of seed oil content and acyl composition in the brassicaceae by expression of a yeast sn-2 acyltransferase gene. Plant Cell 9: 909-923.