OCL
Oilseeds & fats Crops and Lipids

Available online at:
www.ocl-journal.org

RESEARCH ARTICLE – DOSSIER

OPEN ACCESS

DOSSIER

VARIETAL SELECTION OF OILSEEDS: THE PROSPECTIVE NUTRITIONAL AND TECHNOLOGICAL BENEFITS
PERSPECTIVES OFFERTES PAR LA SÉLECTION VARIÉTALE SUR LA QUALITÉ NUTRITIONNELLE ET TECHNOLOGIQUE DES OLÉAGINEUX

Using modern plant breeding to improve the nutritional and technological qualities of oil crops

Denis J. Murphy*

Head of Genomics and Computational Biology Research Group University of South Wales, CF37 4AT, United Kingdom

Abstract – The last few decades have seen huge advances in our understanding of plant biology and in the development of new technologies for the manipulation of crop plants. The application of relatively straightforward breeding and selection methods made possible the "Green Revolution" of the 1960s and 1970s that effectively doubled or trebled cereal production in much of the world and averted mass famine in Asia. During the 2000s, much attention has been focused on genomic approaches to plant breeding with the deployment of a new generation of technologies, such as marker-assisted selection, next-generation sequencing, transgenesis (genetic engineering or GM) and automatic mutagenesis/selection (TILLING, TargetIng Local Lesions IN Genomes). These methods are now being applied to a wide range of crops and have particularly good potential for oil crop improvement in terms of both overall food and non-food yield and nutritional and technical quality of the oils. Key targets include increasing overall oil yield and stability on a per seed or per fruit basis and very high oleic acid content in seed and fruit oils for both premium edible and oleochemical applications. Other more specialised targets include oils enriched in nutritionally desirable "fish oil"-like fatty acids, especially very long chain ω-3 acids such as eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA), or increased levels of lipidic vitamins such as carotenoids, tocopherols and tocotrienes. Progress in producing such oils in commercial crops has been good in recent years with several varieties being released or at advanced stages of development.

Résumé – Utilisation de la sélection variétale moderne pour améliorer les qualités nutritionnelles et technologiques des cultures oléagineuses. Les dernières décennies ont connu de grandes avancées dans la compréhension de la biologie des plantes et dans le développement de nouvelles technologies de manipulation des semences. L'application de méthodes de croisements variétal et les méthodes de sélection ont rendu possible la révolution verte des années 1960 et 1970 qui ont efficacement doublé voire triplé la production céréalière dans une grande partie du monde et évité des famines de masse en Asie. Durant les années 2000, davantage d'attention a été portée à des approches génomiques en sélection des plantes avec le déploiement d'une nouvelle génération de technologies, comme la sélection assistée par des marqueurs, le séquençage de la génération suivante, la transgénèse (ingénierie génétique ou OGM) et la mutagénèse/sélection automatique (ILLI NG, TargetIng Local Lesions IN Genomes). Ces méthodes sont désormais appliquées à un large éventail de semences et offrent, pour les oléagineux, un potentiel particulièrement intéressant à la fois en termes de rendement alimentaire et non-alimentaire, en qualité nutritionnelle et technique des huiles. Les objectifs clés incluent une amélioration du rendement global en huile et de la stabilité par graine ou par fruit, et un contenu très élevé en acide oléique dans l'huile des graines ou fruits à la fois pour les utilisations alimentaires qui sont prioritaires et les applications oléochimiques. Au rang des autres objectifs plus spécifiques, on peut citer les huiles enrichies en acides gras polyinsaturés recherchés pour leur qualité nutritionnelle de type « huile de poisson », notamment les acides gras à longue chaîne de type oméga-3 comme l'acide eïcosapentaènoïque (EPA) et docosahexaènoïque (DHA), ou des niveaux accrus de vitamines lipophiles comme les caroténoïdes, les tocophérols ou les tocopriènes. Les progrès réalisés dans l'introduction de telles huiles dans les semences commerciales ont été notables ces dernières années, avec le lancement de plusieurs variétés ou le développement à des stades avancés.

* Correspondence: denis.murphy@southwales.ac.ak

1 Introduction

Oil-producing crops currently occupy about 10% of global arable land and are second only to carbohydrate crops in terms of their importance as providers of calories for both humans and their livestock. In addition to their edible roles, oil crops also provide a wide range of industrial products, known collectively as oleochemicals, and are increasing being used as sources of biofuels, especially biodiesel (Gunstone, 2011). There are two major categories of oil crop, namely oilseeds and oil-rich fruits. Most oilseeds are annual and grown in temperate regions with notable examples including soybean, maize, rapeseed and sunflower. Major oil-rich fruit crops include tree species such as oil palm, coconut, olive and avocado.

Most domesticated oilseeds typically accumulate 20−50% of the dry weight of their seed tissue as storage oil, normally in the form of triacylglycerol (TAG). However, there are some oilseed species, such as candlenut, sesame, oiticica and ucuhuba, which contain as much as 60−76% oil in their seeds (Murphy, 1996). As discussed below, the existence of substantial amounts of genetically determined variation in seed oil content means that, in the next few decades, it may be possible to significantly increase oil yield in some of major oilseed crops by metabolic engineering or other advanced breeding techniques (Barthole *et al.*, 2012; Marchive *et al.*, 2014; Rahman *et al.*, 2013). The immense potential of genetic variation for increasing oil yield has been demonstrated in reports (as yet unconfirmed elsewhere) from groups in China that claim discovery of a new very high-oil rapeseed variety containing 64.8% w/w seed oil, rather than the normal average of about 42% (Hu *et al.*, 2013; Li *et al.*, 2011).

Most of the major oilseed crops store their oil in their embryo cotyledons. One exception to this is castor bean where the embryo only occupies a small fraction of the seed and most of the oil is stored in a persistent endosperm layer surrounding the tiny embryo. In oil-rich cereals such as maize most of the oil is stored in the embryo and in the outer aleurone layer of the grain. In desiccation tolerant seeds the storage oils accumulate in the cytosol in the form of small (ca. 1μm diameter) spherical organelles termed lipid droplets (previously called oil bodies or oleosomes). These lipid droplets are surrounded by specific proteins, of which the major groups are oleosins and caleosins. These proteins stabilise the small lipid droplets and enable them to withstand the rigours of seed desiccation, sometimes-prolonged dormancy, and subsequent rehydration after germination. In principle it may be possible to select genetic variants of such crops in which the amount of lipid stored per cell is increased at the expense of other components such as storage proteins or carbohydrates.

In oil-rich fruits such as oil palm, olive and avocado, most of the oil is stored in the fleshy mesocarp tissue although the kernel or seed can also accumulate significant amounts of oil. The composition of kernel and mesocarp oil can be very different as in the case of oil palm, or it might be rather similar as in olive fruits. It should be noted that mesocarp and seed/kernel oils have completely different functions and are regulated by different genetic and physiological mechanisms. Seed oils are long term storage reserves that provide a highly concentrated form of energy and carbon skeletons for use by the rapidly growing seedling immediately after germination and before it has acquired net autotrophic capacity via photosynthesis. In contrast, mesocarp oil is part of the fleshy fruit that mainly acts as a lure for animal vectors attracted by the ripening nutrient-rich fruit. Following consumption of the fruit, the seeds normally pass intact through the gut of the animal and are thereby spread well away from the parent plant.

The lipid droplets in ripe fruits tend to lack a defined protein coat and often coalesce to form large irregular oily inclusions in mesocarp cells. During ripening the walls of these cells break down and enzymes such as lipases and lipoxygenases begin to metabolise the oil to create a range of oxidised lipid products, including a suite of volatile components that are released and serve to attract animals. One of the challenges in maximising oil yield from ripe fruit tissues such as mesocarp is the strong tendency for storage TAG to be partially broken down by such lipases and/or lipoxygenases, especially after the fruit undergoes dehiscence or is mechanically separated from the parent plant. In some crops such as oil palm, the release of free fatty acids by lipase action is a major factor in the spoilage of ripe fruits, which can rapidly become unsuitable for further processing.

The ability to manipulate the activity of genes encoding key lipases and/or lipoxygenases could result in substantial increases in overall oil yield of the crop concerned (Morcillo *et al.*, 2013). Lipoxygenases (LOX) can also play important roles in the determination of oil quality in mature fruits. In some crops such as olives, several lipoxygenases are responsible for the creation of the volatile lipid derivatives responsible for the highly favourable organoleptic qualities of the freshly pressed oil (Kalua, 2007). However, lipoxygenases can also adversely affect oil quality and seed viability via lipid peroxidation. Peroxidation of unsaturated fatty acids due to enhancement of LOX activity which directly leads to reduction in seed vigour and deterioration of grain nutritional quality. In a recent study in rice it was found that attenuating LOX activity using RNAi technology had a positive effect on seed oil stability and seed viability – in particular GC-MS analysis revealed that reduction of free fatty acid levels in non-transgenic seeds during storage was higher when compared with that of transgenic rice seeds, which might be relevant to the situation in ripe palm fruits (Gayen *et al.*, 2014).

2 Creating/discovering new genetic variation in oil yield/quality

Conceptually, at least, breeding is a fairly straightforward process. The two keys to the successful breeding of any organism are *variation* and *selection*. All that a breeder requires is some degree of genetic variation for the trait in question in a given population, plus a means of identifying and selecting the most suitable variants. These more useful variants can then be mated or crossed with each other to produce a population that is now enriched in the newly selected genetic variety. Most traits involved in oil yield or quality are complex, *i.e.* they are regulated by the combined activity of several/many

genes, and sometimes also by a range of environmental factors. This makes it challenging for breeders to manipulate such complex traits via simple Mendelian crosses even if there is significant genetic variation in the crop concerned. Another challenge for modern breeders dealing with established crops is that, following centuries of intensive breeding and selection, many of the major commercial crops now have very narrow genetic bases so that there is often only limited genetic variation for key agronomic traits.

As discussed above, in some of the major oil crops like rapeseed, there is still significant genetic diversity in oil yield which has the potential to be used by breeders to create new commercial varieties by the introgression of such germplasm into existing elite breeding lines that are already optimised for other key agronomic traits such as disease resistance. For example, at present, the oil content of the main rapeseed cultivars is only 45−48% in Canada and 41−42% in China and Australia. However, elsewhere in the world, many higher yielding rapeseed lines have been reported such as Major in France (50.7%), Zephyr in Canada (51.4%), Zhongyou0361 in China (54.7%), and a non-edible high-erucic line in Canada (54.8%). More recently, in 2009, a Chinese breeder reported a rapeseed line with 60% followed in 2011 by the rapeseed line YN171 with 64.8%, which is 55% greater than the oil content of the main rapeseed cultivars in China (Hu *et al.*, 2013; Li *et al.*, 2011).

Unfortunately many established oilseed crops, particularly those with a long history of intensive breeding, do not exhibit the kind of genetic variation in oil yield that is described above for rapeseed (which is a relatively new crop that has yet to be fully domesticated as discussed by Murphy, 2007a). This means that modern breeders often need to create new genetic variation either via manipulating the existing crop genome (*e.g.* using wide crosses or mutagenesis) or by adding new genes to the crop genome that result in improved phenotypes (*e.g.* using genetic engineering or transgenesis). The ability of plant breeders to create new genetic variation was greatly enhanced in the mid-20th century by several technological advances. Improved knowledge of plant reproduction made it easier to set up sexual crosses with a wider range of varieties within crop species or with wild relatives from other species. The increasing use of inbred lines for commercial varieties and the loss of many landraces meant that breeders often needed to access new genetic variation from outside the race or species of the crop in question. Their efforts were assisted by the development of new forms of tissue culture as a way of manipulating and propagating plants plus the increasingly refined use of synthetic growth regulators and induced mutagenesis.

These technologies enabled new traits to be added to crops via wide genetic crosses, using methods such as embryo rescue, asymmetric cell fusion, nuclear implantation, and somatic embryogenesis. Previous attempts at wide crossing between distantly related species were frequently frustrated by genome incompatibility. Two important methods to overcome this are the "rescue" of hybrid embryos that would otherwise abort in the seed, and chemically induced chromosome doubling. As well as making possible much wider genetic crosses, chromosome doubling has enabled the use of methods such as somatic hybridisation and haploid breeding (Murphy, 2007b).

3 Transgenic breeding technologies

Transgenic technologies give breeders addition tools to manipulate genomes using recombinant DNA methods that are continually being improved and refined. It is important to stress that breeders never employ transgenesis on its own as a tool for crop improvement. Transgenic technologies simply create new phenotypic variants that still require selection and introgression into elite line. More specifically, transgenesis is used in combination with other breeding technologies such as tissue culture/regeneration, hybrid creation, mutagenesis, backcrossing, and marker-assisted selection. This means that it can be misleading to speak of a new crop variety as "transgenic" or "GM" as if it had only been created using transgenic technologies. As shown in Table 2, in 2013, almost 180 Mha comprising > 10% of the global arable land area was reported as being planted with transgenic/GM crops. However, each of these crops has also benefited from one or more of the non-transgenic technologies listed above. For example, well over three quarters of all crops grown, including most transgenic varieties, have resulted from some form of hybridisation and backcrossing.

Although transgenic crops, including major oilseed species such as soybean, maize, rapeseed and cotton, are now planted extensively in much of the world outside Europe, in all cases the transgenic phenotypes relate to traits such as herbicide tolerance and pest resistance rather than oil yield or quality. Indeed, even where novel oil-related traits are available in transgenic crops, these traits have been produced via non-transgenic breeding approaches such as mutagenesis or wide crosses and then crossed into transgenic lines modified for input traits such as herbicide tolerance. As discussed below, while transgenic methods have been successful in producing high oleic oils in several major crops, it has proved much more difficult to engineer commercially relevant levels of the kinds of novel fatty acids that might be used as industrial oleochemical feedstocks.

One factor that has delayed the uptake of transgenic crops in some parts of the world is that although transgenesis is simply one of several alternative strategies for variation enhancement in breeding programmes, the resultant plants are treated very differently from almost-identical non-transgenic varieties developed via methods such as mutagenesis or assisted wide crossing by government agencies and by some sections of the general public, especially in some parts of Europe. Transgenic varieties have a different legal status and are subject to much more complex regulatory systems in various regions of the world, which can hinder their development and uptake by farmers, processors, retailers and consumers. Indeed, despite almost 15 years of successful cultivation on a global scale, transgenic crops still banned or heavily restricted in some countries. For this reason, we need to look at the development of transgene technology in a different way to other technologies. As we will see below, some developments such as so-called "clean gene" technologies are aimed more at satisfying generalised public concerns rather than addressing proven safety issues or wider aspects of crop improvement *per se*.

There are several ways in which transgene technology can be improved to make it technically easier, more efficient, wider in its scope, and better able to address concerns expressed by certain sections of the public, especially in some parts of

Europe. Some technical issues and areas of public concern are listed below:

– In the future, it will be desirable to generate transgenic crops that do not contain selection markers, such as genes for antibiotic or herbicide tolerance.
– Until now transgenic plants have been created using random insertion of transgenes, which can lead to variations in transgene behaviour and other unpredictable pleiotropic effects. In order to achieve stable and predictable transgene expression under a variety of field conditions, transgene introduction technologies need improvement.
– The spread of transgenes into wild populations via cross pollination can be prevented using genetic use restriction technologies (GURTs).
– Biocontainment strategies should be incorporated into certain types of transgenic plants, *e.g.* expressing non-edible or pharmaceutical products to prevent risk of contamination of human or animal food/feed chains.

There are many cases where breeders actually want to knock out a particular gene in order to create a favourable phenotype. In the many cases where the identity of the target gene is unknown, mutagenesis/TILLING is an effective method to achieve knockouts. However, as we understand more about plant molecular biology, there is an increasing list of traits where the identity of the target gene is known to a high degree of probability. In such cases, transgene-induced gene mutation can be used. Two major methods are RNA interference (RNAi) and zinc-finger mutagenesis.

RNAi can be triggered by generating transgenic lines expressing RNAs capable of forming a double-stranded hairpin. Compared to previous antisense approaches, RNAi is usually much more effective at reducing levels of target gene transcripts and has other advantages as follows. RNAi-encoding transgenes are inherited in a much more stable way, and the degree of downregulation of the target gene can be modulated by varying the strength of the transgene promoter. Also, unlike other technologies, RNAi can knock out all members of a multigene family. Finally, the DNA constructs are relatively easy to make and can be used in the latest generation of transgene vectors. Practical RNAi technology only dates from about 2003, but it has already been demonstrated to be useful in generating variation for important lipid-related traits in oil crops (Cheng *et al.*, 2013; Gayen *et al.*, 2014). However, it is worth pointing out that RNAi is not a radically novel technology in terms of creating genetic variation. It simply an alternative way to generate knockout mutations that can in principle be generated in several other ways, albeit more crudely and expensively.

Zinc-finger mutagenesis is an even more recently refined method for the targeted knockout of plant genes (Shukla *et al.*, 2009). The technology is based on construction of a zinc-finger nuclease that recognises a specific DNA sequence in a genome and induces a double-stranded break at this target locus. Zinc-finger nucleases are engineered proteins made up of zinc-finger-based modules that recognise specific DNA sequences. These DNA-binding modules are fused to an endonuclease domain. Once the zinc-finger nuclease binds to a particular stretch of DNA, the nuclease domain will introduce a double-stranded DNA break in two different places.

This form of DNA cleavage results in an overhang that is difficult to repair, leading to a loss of function in the target gene. Alternatively, if a homologous gene is introduced at the same time as the zinc-finger nuclease, the result will be replacement of the target gene with another gene.

Although first reported in the mid-1990s, zinc-finger mutagenesis was initially just a research tool. However, several reports in the late 2000s have demonstrated that newer versions of the technology may soon be applicable for the efficient generation of targeted mutations in crop plants and/or replacement of poorly performing genes with improved versions. An advantage is that mutagenised plants can be backcrossed to their respective wild type varieties to create genetic lines that contain the desired mutation but are devoid of any transgenic DNA sequences. A comparison of the various gene silencing/mutagenesis technologies is shown in Table 1.

4 Crop genomics

Probably the most dramatic example of technology improvement in the 21st century has been in DNA sequencing where the cost per base has decreased by a remarkable 100 000-fold since 2000, as shown in Figure 1 (Mardis, 2008; Shendure and Ji, 2008). The first plant genome to be fully sequenced was the model species, *Arabidopsis thaliana*, published in 2001, while the first crop genome was rice, where a high quality sequence was published in 2005. The sequencing of the much larger maize genome required a massive effort by company and public labs and the results were published in a series of papers in 2009. Other large-scale projects are currently underway for developing country crops such as sorghum and foxtail millet and sequence data are now being publicly released at an increasingly rapid pace. Advances in next-generation sequencing technologies are enabling the genomes of even comparatively minor crops to be characterised (Edwards and Batley, 2010).

In some cases, a single method has been used but, more commonly, several sequencing technologies are used in combination for best results. For example, Roche 454 technology was used to sequence the 430 Mb genome of cocoa, *Theobroma cacao*, and the 1700 Mb genome of oil palm. In contrast, a combination of Sanger and Roche 454 sequencing was used for the apple and grape (500 Mb) genomes. A combination of Illumina Solexa and Roche 454 sequencing was used for the genomes of polyploid cotton. Roche 454 sequencing has been used for Miscanthus, while Sanger, Illumina Solexa, and Roche 454 sequencing are being used for banana. Illumina GAII sequencing has been used for the *Brassica rapa* genome, while Sanger and Illumina Solexa technologies were used for the cucumber genome.

These powerful combined approaches are now making it feasible to tackle very large cereal genomes such as barley (5500 Mb) and breadwheat (17 000 Mb) where their massive size had previously ruled out full-genome sequencing. The cheapness and speed of genome sequencing is also making it possible to sequence, not just single reference genomes, but many individual genomes in a population. This approach will be used to uncover genome-wide variations that underlie some

Table 1. Comparison of different gene silencing/mutagenesis technologies.

Mutagenic agent	Effect on DNA structure or gene expression	Comments
Ethylmethane sulphonate (EMS)‡	Point mutations, CG to AT transitions	Large range of mutations including gain and loss of gene function
Di-epoxy butane (DEB)‡	Point mutations, 6−8 bp deletions	High efficiency, hundreds of mutations per genome Difficult to locate mutations in genome
Gamma-rays or X-rays‡	Chromosome breaks and rearrangements	Large insertions, deletions, and rearrangements Mostly loss of function mutations
Fast neutrons‡	Deletions, < 1 kb	Medium efficiency Difficult to locate mutations in genome
T-DNA*	DNA insertion	Insertion of specific DNA sequences Mostly loss of function mutations
Transposons*	DNA insertion/deletion	Low efficiency Easy identification of mutations
RNAi or zinc-finger constructs*	DNA insertion so that transcribed gene product causes gene silencing	Targets single genes Causes loss of function Very effective Requires knowledge of target gene function

‡ Non-transgenic method; * transgenic method.

Table 2. Major global transgenic crops and traits in 2013. Data from James (2013) Global Status of Commercialized Biotech/GM Crops: 2013. *ISAAA Brief* 46, http://www.isaaa.org/resources/publications/briefs/46/default.asp.

Crop	Area, Mha	%
Soybean	84.5	48
Maize	57.4	33
Cotton	23.9	14
Rapeseed/canola	8.2	5
Sugar beet	0.5	< 1
Alfalfa	0.1	< 1
Others	0.6	< 1
Total	175.2	100

Trait	Area, Mha	%
Herbicide tolerance	100	57
Insect resistance (Bt)	31	18
Herbicide tolerance + Insect resistance	44	25
Others	< 0.1	< 1
Total	175	100

Fig. 1. Decreasing DNA sequencing costs over the past decade has been mirrored by a huge increase in the proportion of uncurated proteins in public databases. Since the early 2000s the cost of sequencing genomes has plummeted by more than 4 orders of magnitude. This has created a glut of raw data, much of which has yet to be curated in terms of definite identification of functional genes and proteins. By 2012, for each curated protein sequence in public databases there were several million uncurated sequences. Data from Pubmed and NCBI.

of the more complex developmental and agronomic traits of interest to researchers and breeders.

4.1 Beyond the genome: other "omic" technologies

In order to move beyond gene composition through to gene expression, protein function and their ultimate manifestations as phenotypes in an organism, it is often necessary to analyse structural and functional molecules, such as proteins, membrane lipids, and carbohydrates in particular plant cells or tissues. At a more detailed level, there are many thousands of smaller metabolites whose composition differs greatly according to tissue, developmental stage, and in response to environmental conditions. The ability to simultaneously analyse large numbers of often-complex molecules is the basis of the so-called "omic" technologies. Hence, transcriptomics is the analysis of transcribed genes in the form of mRNAs; proteomics is the analysis of protein composition; lipidomics is the analysis of lipid composition; metabolomics is the analysis of small metabolites, and so on. Several automated analytical techniques have been developed to separate and identify each of these classes of biomolecules.

The transcriptome is a comprehensive list of the genes expressed in a particular tissue at a particular stage of development and/or in response to particular environmental stimuli. In many ways transcriptome sequences can be much more useful than genome sequences because they only include the particular fraction of the tens of thousands of genes that are expressed under specific conditions. In the case of oil palm, the analysis of the fruit transcriptome during oil accumulation is already proving very useful in identifying key genes that may regulate this vital process (Bourgis *et al.*, 2011; Dussert *et al.*, 2013; Tranbarger *et al.*, 2011).

The metabolome is the complete list of metabolites found in a particular organelle, cell, or tissue under a specific set of conditions. The identification of important plant metabolites in a plant such as oil palm, which include carotenoids, phenols, and fatty acyl components, used to be a very slow process relying on bulky, expensive equipment that could only be operated by a few skilled specialists. However, new lightweight devices, supplemented by robotic and informatics approaches, now make it possible to automate the process and even to assign accurate identities to complex mixtures of such molecules.

Metabolome analysis can help uncover the details of fruit development at a molecular level (Neoh *et al.*, 2013; Teh *et al.*, 2013). Metabolome studies can also indicate how plants are reacting at the molecular level to specific stimuli, *e.g.* by comparing stressed and unstressed plants we can gain important information about how some plants can tolerate certain stresses while others cannot. In other cases, metabolome analysis can give useful information about molecular changes caused by the addition of transgenes to plants. This kind of analysis is often used as part of the process of regulation of transgenic crops where it may be necessary to test whether a transgenic variety is "substantially equivalent" to non-transformed varieties of the same crop (Beale *et al.*, 2009).

The proteome is defined as the expressed protein complement of an organism, tissue, cell or subcellular region (such as an organelle) at a specified stage of development and/or under a particular set of environmental conditions. Perhaps the most important molecules in cells are the proteins, some of which are structural while many others act as enzymes responsible for the biosynthesis of most of the other molecules in a cell. Proteins are the direct products of gene expression and the timing and spatial distribution of their accumulation and function results in the phenotype of a particular organism. However, patterns of gene expression as measured by transcription, *i.e.* the formation of mRNA, are not always reflected by patterns of accumulation or activity of the corresponding proteins.

In some cases, the mRNA may not be efficiently translated to protein. In other cases the protein might be synthesised but is then either broken down or remains inactive. A protein might be present in a cell but is inactive due to incomplete post-translational processing, *e.g.* failure to bind a ligand, or due to inhibition, *e.g.* by phosphorylation. There are many examples where proteins might be present in a cell but remain in an inactivated state until they are activated by a specific stimulus. In such cases, both transcriptome and proteome data would indicate that the gene was active and the protein was being synthesised but this would be misleading in terms of function if the protein was not active. Ideally, the information in the proteome

should therefore include any post-translational processing undergone by each protein analysed. 1st-generation proteomics was mainly concerned with identifying the gross protein composition of samples, but new-generation technologies are beginning to focus on questions such as post-translational processing and the biological activities of such proteins (Murphy, 2011). Despite these advances, it remains a significant challenge to identify which proteins within a given proteome are partially or completely functionally active.

It is only by addressing these latter questions that we can verify not only that a particular protein has been synthesised and is in the right location, but also that it has the appropriate biological function. Therefore, we can learn a lot about the actual function of a genome in a specific cell or tissue by examining its proteome. Like the metabolome, the proteome in a plant sample can vary greatly according to genotype, tissue location, developmental stage, and environmental conditions (Gómez-Vidal *et al.*, 2009; Zamri, 2013). The full proteome will comprise thousands of proteins, some of which may be present in high abundance while others are at very low levels. The analysis of low-abundance proteins poses considerable difficulties for proteomics that have yet to be resolved but given the rate of progress it is likely that automated or semi-automated methods will be developed for the near-complete description of the oil palm proteome in the not too distant future.

4.2 Bioinformatics

Bioinformatics is a relatively new discipline that brings together biologists, mathematicians and computer scientists to make sense of the avalanche of data generated by genome sequencing and profiling programmes, and from the other 'omic technologies described above (Kanehisa and Bork, 2003). The sheer volume of data generated by these methods often makes it virtually impossible to analyse raw results manually. For example, a next-generation DNA sequencer can generate thousands of sequence fragments making up millions of base pair readouts per day. These fragments need to be analysed for overlaps and then assembled into "contigs", or continuous sequences of many fragments that will eventually be collated to make up an entire chromosome. This process is now done automatically using algorithms, or repetitive step-by-step mathematical procedures.

Other algorithms are used in genome annotation. This involves the identification of putative genes, including their promoters, regulatory elements, introns, exons, and mRNA/protein products. Other software can detect possible regions encoding small, non-coding RNAs and specific repetitive elements in genome sequences. Such sequences are now known to play important roles in several aspects of genome function in complex eukaryotes such as higher plants and animals (López-Flores and Garrido-Ramos, 2012; Van Wolfswinkel and Ketting, 2011). Software is also used to drive robotic and other automated systems used in tasks such as mass profiling of large populations. Advances that enable non-specialists to use sophisticated software have been facilitated by improved computing technology and more powerful linked networks. This has been especially crucial in enabling massive amounts of data, often measured in many terabytes

$(10^{12}$ bytes), that are generated by some of the new technologies. For example, a single two-hour run on an Illumina GAII DNA sequencer can generate 10 terabytes of data.

One potential problem here is that the vast amounts of raw data generated by DNA sequencers are beyond the ability of many labs, or even companies, to archive. Therefore, the raw data are often immediately processed by proprietorial software developed by instrument manufacturers and only the much-reduced processed data are saved. Even with the most advanced computing technology, the costs of storing the original raw data can be greater than the cost of repeating the entire sequencing run. Another challenge for future software development is to improve the assembly of processed sequence data for the increasingly diverse applications required by researchers. To address this, new forms of open-source bioinformatics software, such as SOLiD, are being developed where members of the community can adapt and improve software tools to fit their own applications.

In the future it will be desirable for bioinformatics researchers to work closely with biologists to develop a wider range of broadly applicable tools for the extraction of useful information from the various 'omics databases related to oil palm. Because this is pre-competitive research, both the databases and analytical tools such as algorithms should ideally reside in the public domain as open-source products. The magnitude of this problem is demonstrated in Figure 1, which shows how the massive decrease in sequencing costs has resulted in an avalanche of gene and protein sequence data in public repositories. However, there are currently more than one million uncurated protein sequences for each curated sequence. Clearly there is an urgent need to curate and validate many more genes and proteins in these databases.

5 Selection of favourable genetic variants

As discussed above, novel genetic diversity in an oil crop species can be created by several different methods including transgenesis, mutagenesis/TILLING, wide crosses, and introgression from wild relative. However, it is still often necessary to identify the genetic source of the novel trait and to develop easily used markers. Several types of genetic marker can be used to assist the selection of favourable traits in plant breeding. Morphological and biochemical markers, such as fruit colour, fatty acid composition, or dwarfism, are relatively easy to observe or measure but many other key agronomic traits such as disease resistance are not so easily assessed in this way. By far the most useful class of genetic markers are those based on DNA sequences. Such markers are now being applied to almost every aspect of plant and animal breeding, and also in medicine, basic research and even in forensic science. The use of modern techniques like association genetics and quantitative trait loci (QTL) analysis are enabling chromosomal regions and individual genes involved in the regulation of important traits to be mapped and identified (Rafalski, 2010; Xu, 2010). These methods have recently been used to map the lipase gene involved in oil deterioration in ripe palm fruits (Morcillo *et al.*, 2013) and QTL analysis of genes regulating the fatty acid composition of palm oil (Montoya *et al.*, 2013).

5.1 QTL analysis

Quantitative trait loci (QTL) are chromosome regions containing genes that regulate complex traits. In several cases, genetically complex traits of agricultural interest are mostly regulated by one or a few major QTL. For example, it was known that the grain-shattering trait in rice is regulated by numerous genes and several potentially important QTL were found. However, more detailed analysis showed that just one of these QTL is responsible for 69% of the genetic variation in grain shattering. This locus coincides with the *sh4* gene, which encodes a transcription factor regulating the expression of several other genes involved in grain shattering. By isolating the *sh4* gene, researchers were able to learn a great deal about the process of grain shattering in rice as well as possible new ways to manipulate this key trait in rice and other crops. QTL analysis has now been applied to many other crop species and their wild relatives (Agarwal *et al.*, 2008; Bernard, 2008; Guimarães *et al.*, 2007) and is being applied in oil crops to the identification of complex lipid-related traits such as fatty acid composition (Montoya *et al.*, 2013).

In order to carry out QTL analysis, two parents with widely different genotypes are crossed to create a segregating population. The parent plants might be different varieties of the same species, or a crop plant crossed with a wild relative. This method is known as biparental crossing. Using such populations, a series of genetic markers can be assembled at intervals along each chromosome. Major QTL involved in traits of interest can then be pinpointed with respect to these markers. The initial resolution of such mapping is relatively low as it can only localise QTL to chromosome regions of 10−30 cM, which might correspond to several hundred genes. In order to localise a gene of interest, it is necessary to perform finer resolution genetic analysis, *e.g.* by crossing near-isogenic lines that only vary in the QTL region. This might narrow the region to just a few genes that can then be tested for their effects on the original trait by knockouts or overexpression studies. Eventually, a single gene might be characterised that could enable a relatively complex trait to be manipulated in order to improve crop performance.

DNA-based marker assisted selection can save time and money in crop breeding programmes as follows. In order to select most characters of interest, it is normally necessary to grow up and analyse each new generation of the crop before it is possible to perform phenotypic selection of appropriate plants. Many traits, such as disease resistance or salt-tolerance cannot be measured until plants have been grown, often to full maturity, and then tested in the field. A DNA-based molecular marker is used to identify a segment of genomic DNA within which allelic variation in sequence has allowed its location to be genetically mapped. In breeding programmes, such markers are chosen because of their close proximity to a gene of interest so that the marker and target gene are inherited together. This enables breeders to use the marker as a relatively straightforward way of screening very large populations for the presence of a target gene without needing to perform complex phenotypic tests. Hence, MAS can be used to track the presence of useful characters in large segregating populations in crop-breeding programmes. Using molecular markers, breeders can screen many more plants at a very early stage and save

several years of laborious work in the development of a new crop variety.

This is especially useful for crops like oil palm where it can take 3−4 years or more for a fruit phenotype to become fully apparent. Molecular markers have now been developed for most of the major commercial crops, including several tree species. In addition to their increasingly prominent role in genetic improvement of crops, molecular markers are useful for many other applications such as characterising crop genetic resources, management of gene banks, and disease diagnosis. At present marker assisted selection systems are being developed for oil palm and comprehensive genetic and physical maps of the genome are now available. Genetic maps have recently been used to localise oil palm genes involved in the regulation of important traits such as fatty acid composition (Montoya *et al.*, 2013; Singh *et al.*, 2009), embryogenesis and callogenesis (2013), seed coat thickness (Singh, 2013b).

5.2 Association genetics

Association genetics, or association mapping, was initially developed as a research tool for mapping and characterising genes of interest, especially those regulating complex traits. It is therefore related to methods such as QTL mapping. Previously, genetic mapping in plants was normally done by selecting two dissimilar parents to create a biparental segregating population. In contrast, association genetics uses collections of individuals from diverse sources, such as wild populations, germplasm collections, or specific breeding lines. It uses new methods to accelerate crop genetic profiling, including large-scale SNP discovery and high-throughput sequencing, and the availability of increasing availability of plant genomic sequences.

Association genetics involves searching for statistically significant associations between changes in a DNA sequence and changes in the phenotype of a trait in a large panel of unrelated genetic lines of a species. So far it has been used as research tool to study the genetic basis of complex traits in human and animal systems, and more recently in plants. It was initially focused on single gene traits in plants, but is increasingly used to analyse quantitative traits. Unlike traditional linkage mapping studies of populations created by crossing two parents, association genetics can explore all the recombination events and mutations in large and diverse populations. It can also achieve higher mapping resolutions, which facilitates identification of major genes regulating complex traits.

More recently, association genetics has been used for more practical applications such as commercial crop breeding (Rafalski, 2010). So far, it has mostly been applied to maize for the analysis of such traits as starch composition, anthocyanin biosynthesis, oleic acid content, and carotenoid content. It has been used to study flowering time in pearl millet and barley. In the future, as genome characterisation becomes more detailed and informatics tools become more sophisticated, association genetics will be applied to more agronomic traits and to more crop species. The technique is especially useful for characterising desirable alleles that are moderately abundant in populations. For the identification of rare alleles, such as some disease resistance genes or genes introgressed from exotic germplasm (such as wild relatives of a crop) conventional segregating populations formed from biparental crossing will still be required for genetic mapping.

6 Modifying oil composition

6.1 High oleic oils

One of the most striking features of oil crop breeding over recent years has been the move to developing very high oleic acid varieties in all the principal commodity crops. Oleic acid-rich oils are desirable for several reasons. First, their high monounsaturated fatty acid content confers advantages in edible markets, especially in comparison with either highly saturated or highly polyunsaturated oils. Second, high oleic acid oils have favourable industrial properties including good lubricating performance over a wide temperature range and superior performance as biodiesel methyl esters at low temperatures.

This means that developing high oleic varieties of the major global oil crop, oil palm, is now an important R&D priority for the industry. Palm mesocarp oil typically contains about 35−40% oleic acid plus 40−50% palmitic and about 10% linoleic acids. The development of a much higher oleic acid composition in the mesocarp oil would open up new markets for edible palm oil and could eventually lead to the displacement of existing less efficient high-oleic oilseeds such as soybean, rapeseed and sunflower.

There are several precedents for the creation of completely new market opportunities by breeding high-oleic varieties of oil crops. Perhaps the most dramatic example is that of rapeseed, which like other brassica oilseeds, such as mustard and crambe, historically produced a seed oil that consisted of > 50% erucic acid. While erucic acid can be used in some food applications, several studies claimed that its consumption by rats was associated with cancers and in the 1960s rapeseed oil was banned from use in the USA. This led Canadian breeders to develop new rapeseed varieties with 60−65% oleic acid and very low levels of erucic acid. These new forms of high-oleic rapeseed were called "canola" and soon dominated the market, creating an entirely new multi-billion dollar export crop for Canada and a new relatively cheap and nutritious vegetable oil for consumers around the world (Murphy, 2007b).

The new canola varieties were the result of naturally occurring mutations that inactivated the fatty acid elongase system responsible for converting C18:1 oleic acid to C22:1 erucic acid. In order to isolate and characterise these mutants, many thousands of seeds from diverse accessions were laboriously screened in a process that took almost a decade. A similar approach was used to screen seeds of sunflower, which normally contain high levels of the polyunsaturate, linoleic acid. In a few cases, mutated sunflower seeds were found where inactivation of oleate desaturase genes resulted in a greatly reduced ability to form linoleic acid and the accumulation instead of an oil containing 60−75% oleic acid. In other cases, induced mutagenesis has been used to create new genetic variation in seed oil content. For example, this mutagenesis approach was used to develop high-oleic versions of linseed where the seed-specific desaturases responsible for converting oleic acid to linolenic acid were inactivated by several mutations (Murphy, 2007b).

More recently, similar conventional (*i.e.* non-transgenic) breeding approaches have led to the development of very-high oleic oils such as rapeseed/canola with 75%; soybean with 83%; sunflower with 80−90%; safflower with 75%; and olive with 75% oleate. The use of induced mutagenesis in crop breeding, including fatty acid manipulation, has recently been made much more effective by the automated mutagenesis/selection system termed TILLING (Murphy, 2011; Shu, 2009; Xu, 2010). In other cases, transgenic (GM) approaches have been used by commercial companies to produce very high oleate and low polyunsaturate varieties of some of the major annual oilseed crops. Examples include rapeseed/canola (89% oleate); Indian mustard (73% oleate); soybean (90% oleate); and cottonseed (78% oleate). These transgenic lines are based on antisense or RNAi technologies and several other gene deletion technologies are also under development (Murphy, 2011). In one recent example, targeted mutation of Δ-12 and Δ-15 desaturase genes in hemp resulted in major alterations in seed oil profile including the production of a 70% oleate variety (Bielecka *et al.*, Bielecka). In another case, transcription activator-like effector nucleases (TALENs) were engineered to recognize and cleave conserved DNA sequences in two seed-specific oleate desaturase genes in soybeans, which resulted in an increase in the oleate content of the seed oil from 20% to 80% (Haun *et al.*, 2014).

6.2 Very long-chain polyunsaturated oils

As outlined above, very long-chain polyunsaturated oils enriched in ω-3 fatty acids are widely regarded as having significant nutritional benefits. The major source of such oils is oily fish, in particular species such as trout and salmon. As with many marine species, stocks of these fish in the wild are becoming depleted. As an alternative they are increasingly being raised as a livestock resource in fish farms. However, in some cases fish farms have been criticised for the amount of disease and pollution created. Moreover, even farmed fish are still relatively expensive and are not suitable for all types of diet, such as when a person has an allergy to fish. Another alternative strategy is to produce very long-chain polyunsaturates in crop plants instead of fish. However, plants do not normally accumulate such fatty acids and numerous additional enzymes are needed for the conversion of a typical plant 18-carbon polyunsaturates to very long-chain polyunsaturates such as DHA or EPA as shown in Figure 2.

Despite the considerable technical barriers involved in the metabolic engineering of very long-chain polyunsaturates in crop plants, encouraging progress has been made. For example, in one experiment, nine genes from various fungi, algae, and higher plants were inserted into the oilseed, *Brassica juncea*, with the resultant accumulation of as much as 25% arachidonic acid and 15% EPA. While this was a notable achievement, much higher levels of these lipids will need to be produced before transgenic plants can be viable alternative sources of marine oils (Ruiz-López *et al.*, 2012; Wu *et al.*, 2005; Xue *et al.*, 2013). One notable recent achievement has been the engineering of transgenic *Camelina sativa* plants with < 31% EPA or 14% DHA (Ruiz-López *et al.*, 2014). These ω-3 fatty acid-enriched camelina varieties have now
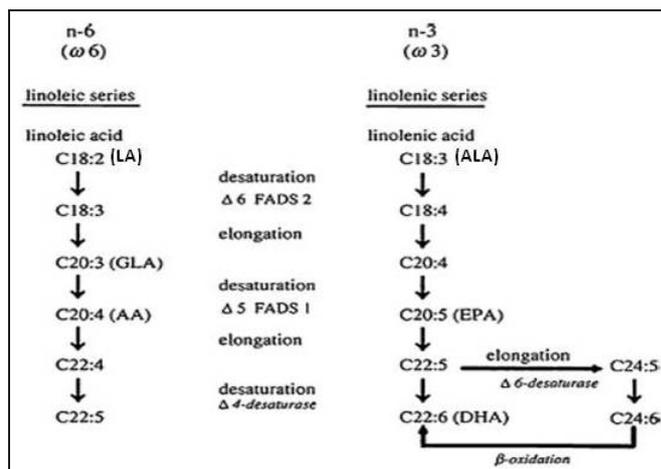
Fig. 2. Biosynthetic pathways for conversion of linoleic acid to very long chain polyunsaturated fatty acids.

been approved for field trials in the UK and in the future they could provide a niche source of speciality oils for nutraceutical markets (Rothamsted, 2014).

6.3 Other novel oils

The major oilseed crops have relatively restricted lipid profiles and are almost exclusively made up of 16-carbon and 18-carbon acyl species with between 0 and 3 double bonds. This narrow range of fatty acids is not always optimal for various downstream uses, which is why there is great interest in extending the diversity of fatty acids in commercially available sources of lipids. In contrast to the commonly used crop and livestock sources of lipids, there are many other species that accumulate more exotic fatty acids that could be used in many types of industrial applications. Such novel fatty acids can range from 8-carbon to 24-carbon chains with a host of interesting functional groups including conjugated double bonds, triple bonds, hydroxy and epoxy groups. Most of these plants are unsuitable to be grown as crops themselves but they could be sources of genes encoding the enzymes responsible for their ability to accumulate the novel acyl groups. If such genes could be successfully transferred to some of the major oil crop species then it might be possible to produce commercially viable quantities of the novel acyl lipids (Murphy, 2010).

Despite over twenty years of trying and some encouraging recent results (Mietkiewska *et al.*, 2014; Nguyen *et al.*, 2014), it has proved much harder to achieve this aim of "designer GM oil crops" than was first imagined (Murphy, 1994; 2009; 2010; Vanherke *et al.*, 2013; Zanetti *et al.*, 2013). In many cases small amounts of the novel fatty acids are produced but are not efficiently transferred to form TAGs, or accumulate instead on membrane lipids with undesirable side effects. In other cases, the novel fatty acids are broken down by oxidation before they can accumulate in storage oils. It also appears that different plant species often have different mechanisms for processing fatty acids for storage, so that different strategies for the identification and transfer of suitable genes must be used in each

case. In short, the production of high levels of exotic or unusual fatty acids in seed or fruit storage oils behaves like a quantitative trait involving numerous genes, not all of which are currently understood. Therefore, while the achievement of 80−90% levels of such fatty acids in a given oil crop is probably feasible, the practical realisation of this goal is still some way from fulfilment.

6.4 Carotenoids

The lipophilic pigment, β-carotene (provitamin A), which is mostly found in some seed oils and in leaves, is a highly desirable nutrient that is deficient in the diet of many people in developing countries. This is especially true where rice is the main dietary staple. Normal rice grains are white and are almost entirely lacking in β-carotene. In some parts of Asia where poorer people often mainly subsist on rice, with very low intakes of coloured vegetables, the incidence of vitamin A deficiency (leading to night blindness) is very high and is estimated to affect some 124 million children. This led a group of Swiss and German biotechnologists to develop so-called "golden rice". The grains of this GM rice variety are yellow because they have been engineered to accumulate high levels of β-carotene.

The transgenic rice contains two inserted genes encoding the two enzymes responsible for conversion of geranyl geranyl diphosphate to β-carotene that are missing in normal white rice. In the initial variety of golden rice produced in the late 1990s, the two transferred genes were phytoene synthase from the daffodil, *Narcissus pseudonarcissus*, and carotene desaturase from the soil bacterium *Erwinia uredovora*. These two genes were inserted into rice under the control of an endosperm-specific gene promoter to ensure that they were only expressed in developing grains and not in other tissues. This early variety of golden rice accumulated 1.6 μg/g of carotenoids in its grains and was light yellow in colour, but an adult would need to eat several kilograms of this rice each day in order to obtain their recommended dietary intake of provitamin A.

The relatively low β-carotene yields in the first version of golden rice led to the substitution of a phytoene synthase gene from maize, which was much more effective than the original daffodil gene (Paine *et al.*, 2005). This new version of golden rice contained 37 μg/g of carotenoids in its grains and was a darker orange in colour. It would require only 144 g of this improved rice to provide the recommended daily requirement for provitamin A (Tang *et al.*, 2012). Although the first versions of golden rice were available by 1999, the extensive food safety and environmental checks, plus the need to backcross into local rice varieties have meant that it has so far taken well over a decade for this GM crop to be developed for public release (Beyer, 2010). By 2013, advanced field trials were underway at the International Rice Research Institute in the Philippines and it is planned that the first batches will be released on a trial basis to selected farmers in 2014/15. It is not always necessary to use transgenic methods to increase the carotenoid content of grain crops. For example, naturally occurring genetic variation in lycopene epsilon cyclase activity has been used to create high carotenoid varieties of white maize, which is the major form of the crop in Africa (Harjes *et al.*, 2010).

7 Accumulating storage oil in non-seed/fruit tissues

Most oil crops only accumulate significant amounts of oil in specialised tissues such as seeds and/or fruits. The remainder of the crop plant is generally of relatively low value but can be considerable more abundant than seed/fruit in terms of overall biomass. This makes it attractive to engineer the ectopic accumulation of useful oils in such tissues. Obvious targets are leaves (Vanhercke *et al.*, 2014) and underground tissues such as tubers (Turesson *et al.*, 2010). In both cases, very little net TAG accumulation normally occurs and the tissues are mainly involved in processes such as photosynthesis (leaves) and carbohydrate storage (tubers). However, developmental mutants have been identified in model species such as Arabidopsis where significant amount of TAG accumulation occurs. One example is where a mutation of the PKL (PICKLE) gene led to the upregulation of a variety of transcription factor genes involved in embryo development and the induction of lipid droplet formation in leaf cells (Ogas *et al.*, 1997). Other examples have focused on one of the downstream transcription factor genes from PKL called WRI1 (WRINKLED1) which is involved in the upregulation of parts of the glycolysis and fatty acid biosynthesis pathways that generate the acyl components required for TAG assembly and accumulation (Ma *et al.*, 2013; To *et al.*, 2012).

A very small number of plants can naturally accumulate oils in underground storage tissues and the study of such species might facilitate the engineering of oil accumulation traits in very high biomass crops such as potato, cassava or yams that are currently grown for their starch content. For example, tubers of the yellow nutsedge (*Cyperus esculentus*) begin to accumulate starch and sugars, but later accumulate as much as 24% oil on a dry mass basis in mature tubers (Turesson *et al.*, 2010). If the developmental switch involved in oil accumulation is a relatively simple genetic trait, it may be possible to transfer this very useful phenotype to a wide range of starchy crops and thereby greatly increase their nutritional value either as human food or livestock feed.

References

Agarwal M, Shrivastava N, Padh H. 2008. Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Rep.* 27: 617–631.

Barthole G, Lepiniec L, Rogowsky PM, Baud S. 2012. Controlling lipid accumulation in cereal grains, *Plant Sci.* 185: 33–39.

Beale MH, Ward JL, Baker JM. 2009. Establishing substantial equivalence: metabolomics. *Methods Mol. Biol.* 478: 289–303.

Bernardo R. 2008. Molecular markers and selection for complex traits in plants: learning from the past 20 years. *Crop Sci.* 48: 1649–1664.

Beyer P. 2010. Golden Rice and "Golden" crops for human nutrition. *New Biotechnol.* 27: 478–481.

Bielecka M, Kaminski F, Adams I, *et al.* 2014. Targeted mutation of D12 and D15 desaturase genes in hemp produce major alterations in seed fatty acid composition including a high oleic hemp oil. *Plant Biotechnol*. J. 12: 613–623.

Bourgis F, Kilaru A, Cao X, *et al.* 2011. Comparative transcriptome and metabolite analysis of oil palm and date palm mesocarp that differ dramatically in carbon partitioning. *Proc. Natl. Acad. Sci. USA* 108: 12527–12532.

Cheng J1, Zhu LH, Salentijn EM, *et al.* 2013. Functional analysis of the omega-6 fatty acid desaturase (CaFAD2) gene family of the oil seed crop *Crambe abyssinica. BMC Plant Biol.* 13:146.

Dussert S, Guerin C, Andersson M, *et al.* 2013. Comparative transcriptome analysis of three oil palm fruit and seed tissues that differ in oil content and fatty acid composition. *Plant Physiol.* 162: 1337–1358.

Gayen D, Ali N, Ganguly M, Paul S, Datta K, Datta KS. 2014. RNAi mediated silencing of lipoxygenase gene to maintain rice grain quality and viability during storage. *Plant Cell Tissue Organ Cult.* 118: 229–243.

Gómez-Vidal S, Salinas J, Tena M, Lopez-Llorca LV. 2009. Proteomic analysis of date palm (*Phoenix dactylifera* L.) responses to endophytic colonization by entomopathogenic fungi. *Electrophoresis* 30: 2996–3005.

Guimarães EP, *et al.*, Eds. 2007. Molecular marker-assisted selection. Current status and future perspectives in crops, livestock, forestry and fish. Rome: FAO, available at: http://www.fao.org/docrep/010/a1120e/a1120e00.HTM.

Gunstone FD. 2011. Supplies of vegetable oils for non-food purposes, *Eur. J. Lipid Sci. Technol.* 113: 3–7.

Harjes CE, Rocheford TR, Bai L, *et al.* 2008. Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science* 319: 330–333.

Haslam RP, Ruiz-Lopez N, Eastmond P, Moloney M, Sayanova O, Napier JA. 2013. The modification of plant oil composition via metabolic engineering – better nutrition by design. *Plant Biotechnol. J.* 11: 157–168.

Haun W, Coffman A, Clasen BM, Demorest ZL, Lowy A, Ray E. 2014. Improved soybean oil quality by targeted mutagenesis of the fatty acid desaturase 2 gene family. *Plant Biotechnol. J*. 12: 934–940.

Hu Z-Y, Hua W, Zhang L, *et al.* 2013. Seed structure characteristics to form ultrahigh oil content in rapeseed. *PLoS One* 8: e62099.

ISAAA' Brief 46. 2013. Available at: http://www.isaaa.org/resources/publications/briefs/46/executivesummary/default.asp.

Kalua CM, Allen MS, Bedgood DR Jr, Bishop AG, Prenzler PD, Roberts K. 2007. Olive oil volatile compounds, flavour development and quality: A critical review. *Food Chem.* 100: 273–286.

Kanehisa M, Bork P. 2003. Bioinformatics in the post-sequence era. *Nat. Genet. Suppl.* 33: 305–310.

Li DR, Tian JH, Chen WJ, Zhang WX. 2011. Breeding technologies and germplasm innovation on extra-high-oil content in *Brassica napus*. *Acta Agricultural Boreali-occidentalis Sinica* 20: 83–87.

López-Flores I, Garrido-Ramos MA. 2012. The repetitive DNA content of eukaryotic genomes. *Genome Dyn.* 7: 1–28.

Ma W, Kong Q, Arondel V, *et al.* 2013. *WRINKLED1*, a ubiquitous regulator in oil accumulating tissues from *Arabidopsis* embryos to oil palm mesocarp. *PLoS One* 8: e68887.

Marchive C, Nikovics K, To A, Lepiniec L. Baud S. 2014. Transcriptional regulation of fatty acid production in higher plants: molecular bases and biotechnological outcomes. *Eur. J. Lipid Sci. Technol.* DOI:10.1002/ejlt.201400027.

Mietkiewska E, Miles R, Wickramarathna A, *et al.* 2014. Combined transgenic expression of *Punica granatum* conjugase (*FADX*) and *FAD*2 desaturase in high linoleic acid *Arabidopsis thaliana* mutant leads to increased accumulation of punicic acid *Planta* 240: 575–583.

Montoya C, Lopes R, Flori A, *et al.* 2013. Quantitative trait loci (QTLs) analysis of palm oil fatty acid composition in an interspecific pseudo-backcross from *Elaeis oleifera* (H.B.K.) Cortés and oil palm (*Elaeis guineensis* Jacq.), Tree Genet. Genome 9: 1207–1225.

Morcillo F, Cros D, Billotte N, *et al.* 2013. Improving palm oil quality through identification and mapping of the lipase gene causing oil deterioration. *Nat. Commun.* 4: 2160.

Murphy DJ, Ed. 1994. Designer Oil Crops – Breeding, processing and biotechnology. Weinheim, New York: VCH.

Murphy DJ. 1996. Engineering oil production in rapeseed and other Oil crops. *Trends Biotechnol*.14: 206–213.

Murphy DJ. 2007a. People, plants, and genes: the story of crops and humanity. Oxford University Press, UK.

Murphy DJ. 2007b. Plant breeding and biotechnology: Societal context and the future of agriculture. Cambridge, UK: Cambridge University Press.

Murphy DJ. 2009. Industrial oil crops, when will they finally deliver on their promise? *Inform.* 20: 749–754.

Murphy DJ. Manipulation of oil crops for industrial applications, In: Singh BP, Ed., Industrial crops and uses. CABI Press, 2010, pp. 183–206.

Murphy DJ. 2011. Plants, biotechnology, and agriculture. CABI Press.

Nguyen HT, Park H, Koster KL, Cahoon RE, Nguyen HTM, Shanklin J, *et al.* 2014. Redirection of metabolic flux for high levels of omega-7 monounsaturated fatty acid accumulation in camelina seeds. *Plant Biotechnol. J.* DOI:10.1111/pbi.12233.

Neoh BK, Teh HF, Ng TL, *et al.* 2013. Differential metabolite profiling of metabolites in Oil Palm Mesocarp at different stages of oil biosynthesis. *J. Agric. Food Chem.* 61: 1920–1927.

Ogas J, Cheng JC, Sung ZR, Somerville C. 1997. Cellular differentiation regulated by gibberellin in the Arabidopsis thaliana pickle mutant. *Science* 277: 91–94.

Paine JA, Shipton CA, Chaggar S, *et al.* 2005. Improving the nutritional value of Golden Rice through increased pro-vitamin A content. *Nat. Biotechnol.* 23: 482–487.

Rafalski JA (2010) Association genetics in crop improvement. *Curr. Opin. Plant Biol.* 13: 174–180.

Rahman H, Harwood JL, Weselake R. 2013. Increasing seed oil content in Brassica species through breeding and biotechnology. *Lipid Technol.* 25: 182–185.

Ruiz-Lopez N, Haslam RP, Napier JA, Sayanova O. 2014. Successful high-level accumulation of fish oil omega-3 long-chain polyunsaturated fatty acids in a transgenic oilseed crop. *Plant J.* 77: 198–208.

Ruiz-López N, Haslam RP, Venegas-Calerón M, *et al.* 2012. Enhancing the accumulation of omega-3 long chain polyunsaturated fatty acids in transgenic *Arabidopsis thaliana* via iterative metabolic engineering and genetic crossing. *Transgenic Res.* 21: 1233–1243.

Rothamsted. 2014. GM Camelina field Trial: Public information, Rothamsted. Research. Availble at: http://www.rothamsted.ac.uk/camelina.

Shukla VK, Doyon Y, Miller JC, *et al.* 2009. Precise genome modification in the crop species Zea mays using zinc-finger nucleases, *Nature* 459: 437–441.

DOSSIER

Soh AC. 2011. Genomics and plant breeding. *J. Oil Palm Res.* 23: 1019–1028.

Shu QY, Ed. 2009. Induced Plant Mutations in the genomics era, FAO, Rome, Availble at: `http://www.fao.org/docrep/012/i0956e/i0956e00.htm`.

Tang G, Hu Y, Yin SA, *et al.* 2012. ß-Carotene in Golden Rice is as good as ß-carotene in oil at providing vitamin A to children. *Am. J. Clin. Nutr.* 96: 658–964.

Teh LS, Neoh BK, Hong MP, *et al.* 2013. Differential Metabolite Profiles during Fruit Development in High-Yielding Oil Palm Mesocarp. *PLoS One* 8: e61344.

To A, Joubès J, Barthole G, *et al.* 2012. WRINKLED Transcription Factors Orchestrate Tissue-Specific Regulation of Fatty Acid Biosynthesis in *Arabidopsis. Plant Cell* 24: 5007–5023.

Tranbarger TJ, Dussert S, Joët T, *et al.* 2011. Regulatory mechanisms underlying oil palm fruit mesocarp maturation, ripening, and functional specialization in lipid and carotenoid metabolism. *Plant Physiol.* 156: 564–584.

Turesson H, Marttila S, Gustavsson KE, *et al.* 2010. Characterization of oil and starch accumulation in tubers of *Cyperus esculentus* var. sativus (Cyperaceae): A novel model system to study oil reserves in nonseed tissues. *Am. J. Bot.* 97: 1884–1893.

Vanhercke T, Wood CC, Stymne S, Singh SP, Green AG. 2013. Metabolic engineering of plant oils and waxes for use as industrial feedstocks. *Plant Biotechnol. J.* 11: 196–210.

Vanhercke T, El Tahchy A, Liu Q *et al.* 2014. Metabolic engineering of biomass for high energy density: oilseed-like triacylglycerol yields from plant leaves. *Plant Biotechnol. J.* 12: 231–239.

Van Wolfswinkel JC, Ketting RF. 2010. The role of small non-coding RNAs in genome stability and chromatin organization. *J. Cell Sci.* 123: 1825–1839.

Wu G, Truksa M, Datla N, *et al.* 2005. Stepwise engineering to produce high yields of very long-chain polyunsaturated fatty acids in plants. *Nat. Biotechnol.* 23: 1013–1017.

Xu Y. 2010. Molecular plant breeding. Oxford, UK: CABI.

Xue *et al.* 2013. Production of omega-3 eicosapentaenoic acid by metabolic engineering of *Yarrowia lipolytica*. *Nat. Biotechnol* 31: 734–740.

Zamri R. 2010. Differential expressions of proteomes of oil palm (*Elaeis Guineensis* Jacq.) and *Arabidopsis thaliana* tissues during callogenesis. Masters thesis, Universiti Putra Malaysia.

Zanetti F, Monti A, Berti MT. 2013. Challenges and opportunities for new industrial oilseed crops in EU-27: A review. *Ind. Crops Products* 50: 580–595.

Cite this article as: Denis J. Murphy. Using modern plant breeding to improve the nutritional and technological qualities of oil crops. OCL 2014, 21(6) D607.