the plant journal
SEB Society for Experimental Biology

The Plant Journal (2018) **96**, 1299–1308 doi: 10.1111/tpj.14102

RESOURCE

PlantFAdb: a resource for exploring hundreds of plant fatty acid structures synthesized by thousands of plants and their phylogenetic relationships

John Ohlrogge[1,2,]*, Nick Thrower[1,2], Vandana Mhaske[3], Sten Stymne[4], Melissa Baxter[1], Weili Yang[1], Jinjie Liu[1], Kathleen Shaw[1], Basil Shorrosh[5], Meng Zhang[6], Curtis Wilkerson[1,2] and Bertrand Matthäus[7,]*
[1]*Department of Plant Biology, Michigan State University, East Lansing, MI, USA,*
[2]*Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI, USA,*
[3]*Independent Biotechnology Professional, Pune, India,*
[4]*Department of Plant Breeding, Swedish University of Agricultural Sciences, Alnarp, Sweden,*
[5]*Biotechnology Professional, Fort Collins, CO, USA,*
[6]*Northwest A&F University, Shaanxi, China, and*
[7]*Department of Safety and Quality of Cereals, Working Group for Lipid Research, Max Rubner-Institut, Detmold, Germany*

Received 9 August 2018; accepted 12 September 2018; published online 19 September 2018.
*For correspondence (e-mails ohlrogge@msu.edu; bertrand.matthaeus@mri.bund.de).

SUMMARY

Over 450 structurally distinct fatty acids are synthesized by plants. We have developed PlantFAdb.org, an internet-based database that allows users to search and display fatty acid composition data for over 9000 plants. PlantFAdb includes more than 17 000 data tables from >3000 publications and hundreds of unpublished analyses. This unique feature allows users to easily explore chemotaxonomic relationships between fatty acid structures and plant species by displaying these relationships on dynamic phylogenetic trees. Users can navigate between order, family, genus and species by clicking on nodes in the tree. The weight percentage of a selected fatty acid is indicated on phylogenetic trees and clicking in the graph leads to underlying data tables and publications. The display of chemotaxonomy allows users to quickly explore the diversity of plant species that produce each fatty acid and that can provide insights into the evolution of biosynthetic pathways. Fatty acid compositions and other parameters from each plant species have also been compiled from multiple publications on a single page in graphical form. Links provide simple and intuitive navigation between fatty acid structures, plant species, data tables and the publications that underlie the datasets. In addition to providing an introduction to this resource, this report illustrates examples of insights that can be derived from PlantFAdb. Based on the number of plant families and orders that have not yet been surveyed we estimate that a large number of novel fatty acid structures are still to be discovered in plants.

Keywords: chemotaxonomy, unusual fatty acid, biochemical evolution, biodiversity, seed oil, triacylglycerol.

INTRODUCTION

The plant kingdom (Plantae) with more than 350 000 species provides much of the natural product diversity on Earth. Over 450 different fatty acid structures have been reported to occur in vascular plants (this study). In contrast with the relatively small number of fatty acid structures that occur in plant membranes (which are often referred to as ‘common’ fatty acids and have, with a few exceptions, 16 or 18 carbons and 0, 1, 2 or 3 double bonds) the storage lipids of seeds and other non-membrane lipids are the primary source for the great diversity in the ‘unusual’ fatty acid class. Variations in structure include chain length, methyl branches, number, position and configuration of unsaturation (double and triple bonds), oxidations (hydroxy, keto, epoxy), cyclic structures (cyclopropane, cyclopentene, furan, phenyl), fluoro, cyano, and more (Smith, 1971; Badami and Patil, 1980).

Interest in these unusual fatty acids arises from several factors. First, many unusual fatty acids have valuable properties that have led to their use in industry. Examples include; lauric acid (from coconut, palm kernel) and its derivatives are major ingredients of soaps and detergents; erucic acid (from rapeseed, etc.) provides lubricants and slip agents for plastic films; conjugated fatty acids (e.g. tung oil) provides paints and coatings; ricinoleic acid (from castor) is used for polymers, lubricants, greases and plastics (https://plantfadb.org/pages/Specialty_Uses). Other specialty fatty acids have cosmetic, antimicrobial and pharmaceutical applications (McKeon, 2016). The value of these special structures has prompted numerous efforts to search for additional fatty acids and additional plant species that could provide renewable feedstocks for industry and additional markets for agriculture. In particular, several large projects, (mainly in the years 1960–1990) were carried out to screen thousands of plant species for new, potentially valuable structures. Selected references to these screening efforts and reviews have been summarized (see https://plantfadb.org/pages/key_literature).

Second, many unusual fatty acids are synthesized by enzymes that have evolved as variants of enzymes of membrane fatty acid metabolism. Comparing the structure of these proteins can reveal insights into enzyme reaction mechanisms and provides a window into 'catalytic plasticity' of fatty acid biosynthetic enzymes (Shanklin and Cahoon, 1998). Likewise, studies of the auxiliary enzymes that evolved to efficiently channel the unusual fatty acids into oil will give information about the stepwise evolutionary events leading to oil with high amounts of an unusual fatty acid. This in turn provides tools that can guide rational engineering of novel enzymes and products.

Third, the occurrence of unusual fatty acid structures in different branches of plant phylogeny can provide information about parallel versus convergent evolution of pathways and reveal new insights into how biosynthetic pathways evolved. Correlations between phylogenic distribution and ecology can also sometimes provide clues to the possible functions of unusual fatty acids.

Fourth, engineering the 'common' fatty acid composition of oilseed crops has been extremely successful (Damude and Kinney, 2008). Achieving the same success with high-value unusual fatty acids is an important goal for society that could provide new income for farmers and new renewable resources to replace petroleum with agricultural sources of chemicals for industry (Dyer *et al.*, 2008; Carlsson *et al.*, 2011; Vanhercke *et al.*, 2013).

In 2003, the Seed Oil Fatty Acid (SOFA) database became available (Aitzetmüller *et al.*, 2003). SOFA compiled data collected at the Institute for Chemistry and Physics of Lipids (Münster, Germany, a former institute of the Max Rubner-Institut) together with data from over 1400 publications spanning more than 40 years of research. An update to the database software and literature surveys became available in 2012 (Matthäus, 2012).

Building on SOFA, we have developed PlantFAdb as a tool with several new functions that increase the ease of use and add the ability to visualize taxonomic relationships among fatty acid structures and the plants that produce them. The software and user interface have been completely revised, and the number of publications, plants and fatty acid structures have been updated and substantially increased. Unique to PlantFAdb, users are able to easily explore connections between fatty acid structures and plant species by displaying these relationships on dynamic phylogenetic trees. In this paper we present an overview of the website and database and highlight some insights from its use.

RESULTS AND DISCUSSION

The establishment of the SOFA database in 2003 represented a major accomplishment in meticulous accumulation of published and unpublished datasets from several decades of research. Over a period of more than 40 years the Institute for Chemistry and Physics of Lipids has collected more than 1400 publications from hundreds of journals and manually extracted and recorded fatty acid composition and other lipid data. In addition, hundreds of unpublished analyses from the Institute for Chemistry and Physics of Lipids in Münster (now Max Rubner-Institut) were included. From 1997 to 2002 these data were transferred into an electronically searchable relational database and the results became available on the Internet in 2003 as SOFA (Aitzetmüller *et al.*, 2003). After the retirement of Dr. Aitzetmüller, Bertrand Matthäus oversaw the database and provided an updated version of SOFA (http://sofa.mri.bund.de) (Matthäus, 2012).

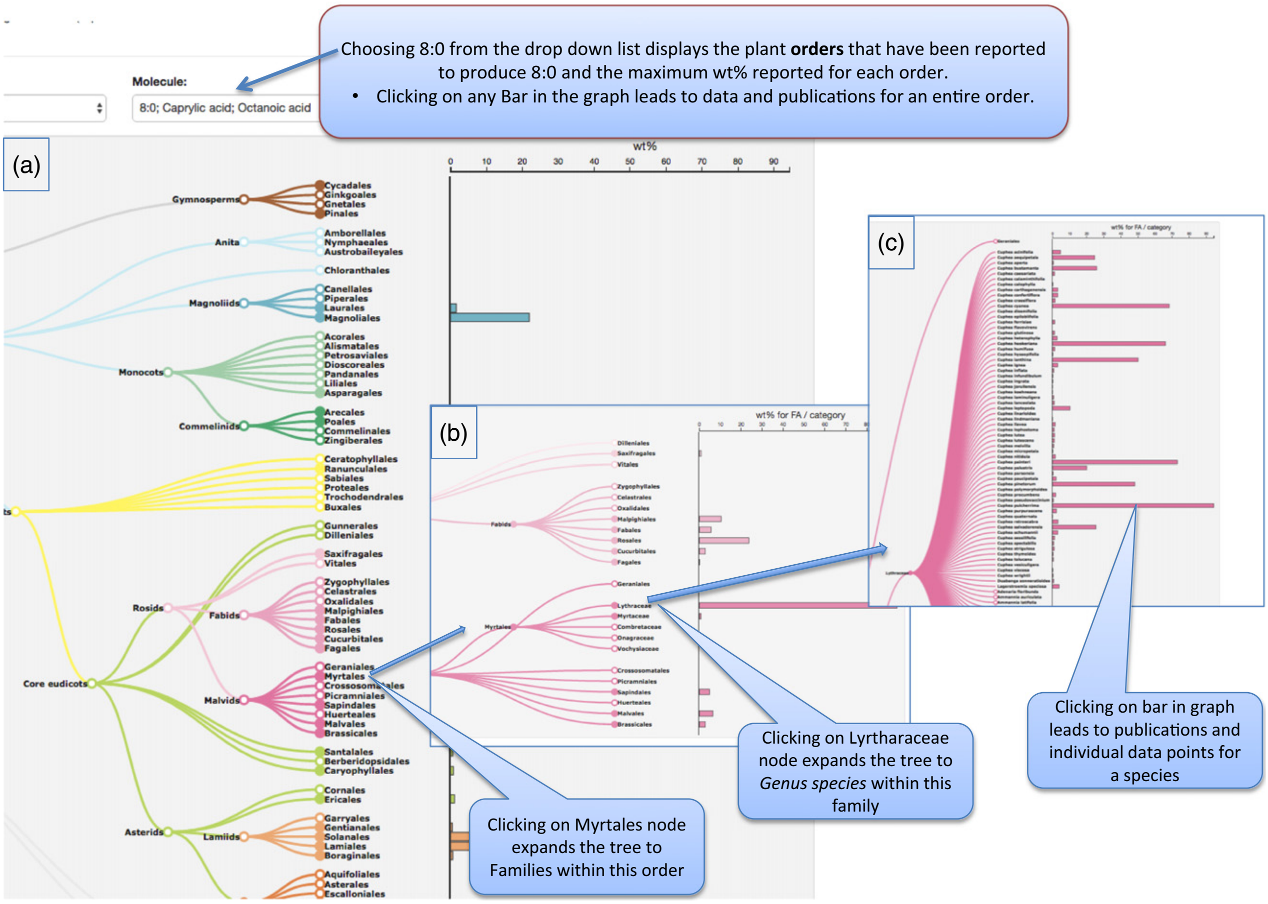
PlantFAdb was developed to further update and expand the information available at SOFA and to add the ability to link fatty acid composition data to phylogenetic relationships. All plant names were reviewed and updated, common names were added, and specific fields in the database were added for variety/cultivar and for source of plant material. For fatty acids, nomenclature has been updated and structure images, mass, molecular formula and links to other databases are now included. For publications, full citation information, including abstracts, digital object identifiers and links to full text (when available) were added. PlantFAdb also contains information from a large number of publications not available via the Internet because they have not been digitized. Therefore the database is the most comprehensive source for information on fatty acids produced by plants. In addition, database fields have been added to indicate the type of lipid analyzed (total lipid, triacylglycerol, etc.) and the origin of datasets. Simple and advanced search forms are available for literature and for structures. Data entry forms have been developed with dropdown menus and an autofill ability to allow rapid additions of new datasets to PlantFAdb.

A summary of the additions to SOFA is presented in Table S1 and at: https://plantfadb.org/pages/about. Prior to submission of this manuscript, PlantFAdb has been updated with literature through early 2018. In addition, data from over 1000 publications and five books in the Chinese literature, much of which is not indexed by the most used literature databases has been added (Li *et al*., 2017). Results from PlantFAdb can be downloaded as csv files, allowing easy import into Excel and other software.

Several other large, lipid-focused databases are also available on the Internet (Table S2). In addition, *The Lipid Handbook*, 3rd edition (Gunstone *et al*., 2007) includes a CD-ROM database with information on several hundred fatty acids including biological sources and references. In comparison with these other resources PlantFAdb appears to be the largest in terms of its presentation of primary datasets on fatty acids. Although a number of excellent reviews cover aspects of plant unusual fatty acid synthesis (e.g. Baud, 2018) the most comprehensive reviews on plant unusual fatty acids were published 25 or more years ago (Smith, 1971; Badami and Patil, 1980; van de Loo *et al*., 1993). A comparatively small number of new plant fatty acids have been discovered over the past 20–30 years and there has been no recent general survey of this topic covering all types of structures. PlantFAdb now provides such a survey. A very recent example of discovery of new fatty acids in a Brassicaceae species is illustrative of how some structures have been missed by conventional GC analysis (Li *et al*., 2018).

Display of results on phylogenetic trees

To our knowledge PlantFAdb is the only database of its type that provides the ability to dynamically present chemotaxonomic data on phylogenetic trees. From the 'https://plantfadb.org/tree' page, two dropdown menus are available that allow users to select displays of different categories of unusual fatty acid (e.g. cyclopropene) or examples of individual fatty acids (e.g. octanoic). As an example, Figure 1



Molecule: 8:0; Caprylic acid; Octanoic acid

Choosing 8:0 from the drop down list displays the plant **orders** that have been reported to produce 8:0 and the maximum wt% reported for each order.

- Clicking on any Bar in the graph leads to data and publications for an entire order.

Clicking on Myrtales node expands the tree to Families within this order

Clicking on Lyrtharaceae node expands the tree to *Genus species* within this family

Clicking on bar in graph leads to publications and individual data points for a species

Figure 1. Overview of display of fatty acid chemotaxonomic information in PlantFAdb. The occurrence of octanoic acid is presented as an example. The initial presentation (panel A) displays results for 68 plant orders. Clicking on an order name in the tree expands the display for that order to families (panel B) and clicking on a family name further expands to genus and species (panel C). Clicking on the species or the horizontal bar within the graph opens a new page where the primary data and publication details are presented (not shown).

illustrates the display of information for occurrence of octanoic acid throughout the plant kingdom. PlantFAdb includes over 500 data points for octanoic acid derived from over 300 plant species and 165 publications. The bar graph presents the maximum weight % reported for the selected fatty acid in each taxon. The initial presentation (Figure 1, panel A) displays results for 68 plant orders. Clicking on an order name in the tree expands the display for that order to families (panel B) and clicking on a family name further expands to genus and species (panel C). Clicking on the species or the horizontal bar within the graph opens a new page where the primary data and publication details are presented. All results and graphs in PlantFAdb are immediately updated each time a new dataset is added.

The **main 'Fatty Acids' page** includes a list of more than 450 fatty acids that have PlantFAdb datasets together with structure images, shorthand nomenclature, formula, molecular weight (MW) and number of datapoints. Data on this and other pages can be sorted by clicking on column headings. An advanced search option is available for fatty acids allowing users to enter MW range and other parameters. Clicking on the fatty acid name displays a page with links to other databases, provides tables of all plants that produce the fatty acid, publications, and more. Clicking 'View Tree' on the main fatty acid page opens a new page showing the phylogenetic distribution of plants that produce the selected fatty acid.

The **main 'Plants' page** provides a list of all plants included in PlantFAdb, together with common names, oil content, and the number of publications and datapoints associated with the plant species. The page is initially sorted alphabetically by genus but users can sort on other columns. Data can also be filtered by searching other criteria such as names of a plant order, family, genus, etc. Clicking on the plant name will open a new page where fatty acid composition data from multiple publications are compiled in a single graph (e.g. https://plantfadb.org/plants/10019).

The **main 'Literature' page** lists over 3000 publications in the PlantFAdb database. Abstracts and/or links to full text are available for more than 2000 publications. The simple search box searches all fields, while an advanced query form allows searching independently for author, title, journal, abstract or date (or combinations). Diacritical marks in author's names often complicate searches. We have largely addressed this issue by including multiple spellings; for example a user search for Aitzetmüller, Aitzetmuller or Aitzetmueller will return results for all three.

The landscape of unusual fatty acid occurrence in seed plants

The organization of data in PlantFAdb allows users to visualize an overview of the current state of understanding of plant fatty acid diversity. The ~450 individual fatty acid structures represented in PlantFAdb can be placed into 20 categories. Table 1 presents a summary of these categories with links that display the phylogenetic distribution of each category of unusual fatty acid (or a representative structure). Additional information on specific structures within each category with examples of plant species and references are provided here (https://plantfadb.org/pages/Unusual_FA).

Figure 2 presents a high-level overview of how the structure categories are distributed throughout the plant kingdom. Unusual fatty acids are produced throughout the plant kingdom with at least one structural category reported in most plant orders that have been analyzed (Figure 2). However, there are striking differences among the different types of unusual fatty acids in how widely they occur. Some unusual fatty acid structures are distributed throughout the plant kingdom, appearing in many families and orders. For example, medium chain (C8–C12) structures occur at high levels in seeds from more than 20 orders that are distributed throughout the monocot, dicot and gymnosperm clades of the plant kingdom (https://plantfadb.org/tree?category=C8-C12). Hydroxy fatty acids occur in approximately 15 orders although these are primarily restricted to eudicots and magnolids, and are largely absent from monocots and gymnosperms (https://plantfadb.org/tree?category=mono-hydroxy). The very broad distribution of medium chain and hydroxy fatty acids in most branches of plant evolution almost certainly indicates multiple independent events during evolution that led to biosynthesis of these structures in seeds.

At the other extreme, there are several structures that have so far only been detected in a single plant family or order. These include: fluoro, trihydroxy, furan and cyanolipid structures. Phenyl fatty acids are reported to occur in only two orders that are closely related (Acorales and Alismatales). Several other fatty acid structures are intermediate in their phylogenetic distribution. Structures that likely evolved from the activity of modified desaturases (acetylenic, hydroxy, conjugated, epoxy) are generally found broadly distributed.

It is also possible to ask whether some branches of plant evolution are particularly rich in unusual fatty acid structures. The Asteraceae and Fabaceae families are distinctive in this regard, each with the occurrence of at least 8–10 categories of unusual structures. However, it should be noted that these families are among the largest in the plant kingdom (~30 000 and ~20 000 species, respectively) and have been surveyed more extensively than many other families. The Sapindales order is perhaps more distinctive with six categories of fatty acid structure represented but only ~5700 species. The fatty acids found in monocots (~60 000 species) appear to be substantially less diverse than those of dicots and gymnosperms.

Table 1 Overview of categories of unusual fatty acids

			Phylogenetic Distribution				
Category of Fatty Acid Structure	Example (with link)	Shorthand	Tree (Link)	# Orders	Pathway described?	Enzyme(s) cloned?	Transgenic plants?
Acetylenic	Crepenynic	18:2-Δ-9c,12a	Tree	6	Yes	Yes	Yes
Allenic	Laballenic	18:2-Δ-5,6allene	Tree	3	No	No	No
Branched chain	14-Methylpalmitic	ai-17:0	Tree	>10	Yes	-	-
Conjugated double bonds	Eleostearic	18:3-Δ-9c,11t,13t	Tree	8	Yes	Yes	Yes
Cyano conjugates	Type IV cyanolipid 20:1 ester	CN-IV-20:1-delta-13c	Tree	1	No	No	No
Cyclopentenyl	Hydnocarpic	16:1cy	Tree	4	No	No	No
Cyclopropane	Dihydrosterculic	9,10-cpa-19:0	Tree	4	Yes	Yes	Yes
Cyclopropene	Sterculic	9,10-cpe-19:1	Tree	>10	No	No	No
Dicarboxylic	Phellogenic	21-COOH-21:0	Tree	1	No	No	No
Epoxy	Vernolic	12,13-O-18:1-Δ-9c	Tree	>10	Yes	Yes	Yes
Fluoro/ halogenated	18-fluoro-oleic	18-F-18:1-Δ-9c	Tree	1	Yes	No	No
Furan	Furocarpic	9,12-O-18:2-Δ-9,11	Tree	1	No	No	No
Mono-hydroxy	Ricinoleic	12-OH-18:1-Δ-9c	Tree	>10	Yes	Yes	Yes
Di-hydroxy	9,10-dihydroxy-octadecanoic	9,10-di-OH-18:0	Tree	4	Yes	Yes	Yes
Tri-hydroxy	9,10,18-trihydroxy-octadecanoic	9,10,18-tri-OH-18:0	Tree	1	No	No	No
Oxo/keto	Licanic	4-O=18:3-Δ-9c,11t,13t	Tree	6	No	No	No
Phenyl-	13-Phenyltridecanoic	13-Ph-13:0	Tree	2	No	No	No
Medium chain length (C8-C12)	Lauric	12:0	Tree	>10	Yes	Yes	Yes
Unusual double bond positions	Petroselinic	18:1-Δ-6c	Tree	6	Yes	Yes	Yes
Very-long chain length (>C20)	Erucic	22:1-Δ-13c	Tree	>10	Yes	Yes	Yes

Many new plant fatty acid structures remain to be discovered

Are there more fatty acid structures still to be discovered? Examination of all plant species represented in PlantFAdb together with recent phylogenetic information allows us to identify branches in plant evolution whose seeds appear not to have been analyzed for fatty acid composition.

The plant kingdom is estimated to comprise more than 350 000 species. The fatty acid composition of less than 5% of these species has been examined. Within the Spermatophytes (seed plants) there are approximately 420 families in 68 orders (http://www.theplantlist.org/). Of these, PlantFAdb includes data on fatty acid composition of 250 families in 60 orders. Therefore, fatty acid composition data are **not** available for 170 families and eight orders (representing 40% of plant families and 12% of plant orders). Consequently, examination of plants within these non-surveyed families is likely to yield discoveries of many new fatty acid structures. Table 2 presents information on the branches of the plant kingdom that have had little or no fatty acid analysis and therefore may be rich sources for new discoveries. Special efforts that are focused to access and analyze the unexplored plant species represented in Table 2 can be expected to yield many new discoveries. Advances in analytical technology, such as GC–GC and LC–MS can be expected to enhance the speed, resolution, sensitivity and accuracy of future analyses.

Biosynthetic pathways for many unusual fatty acids remain to be discovered

It is notable that the biosynthesis of at least four categories of unusual fatty acids can be attributed to modifications in the oleate desaturase (FAD2) gene or related desaturases. This is documented for hydroxy (van de Loo *et al*., 1995), acetylenic (Lee *et al*., 1998), epoxy (Lee *et al*., 1998) and conjugated (Cahoon *et al*., 1999) fatty acids. In addition, although the biosynthetic pathways of the allenic and oxo categories have not been described, these fatty acids may also derive from modifications of desaturase active sites to confer new specificities and products. Considering that up to 100 different structures in PlantFAdb may have originated from FAD2-like enzymes, evolutionary modifications of this active site can be considered a major engine of plant fatty acid diversity.

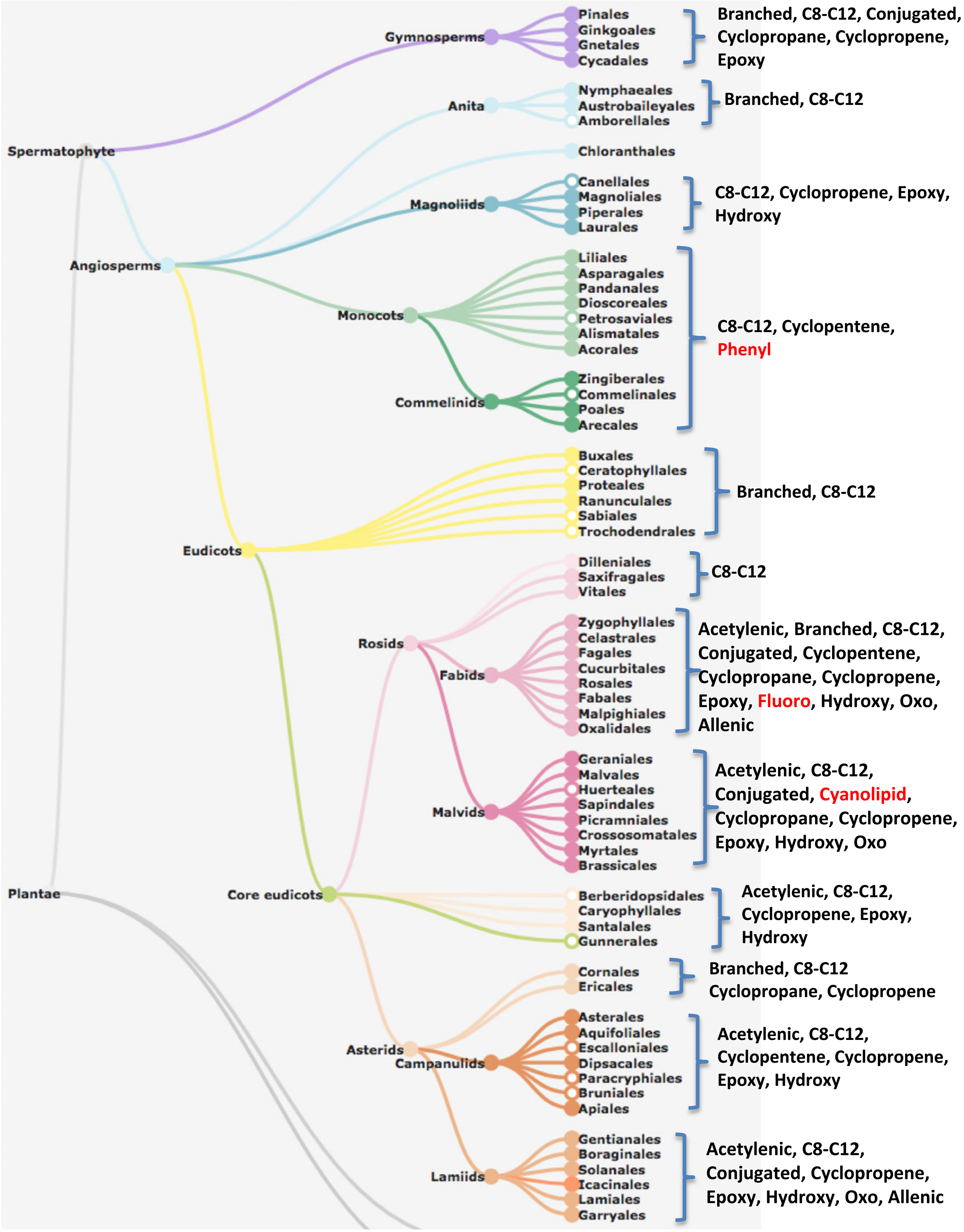


Figure 2. Overview of unusual fatty acid distribution in seed plants. Structures found in only one or two orders are highlighted in red. Structures present at trace or minor levels are not included. The absence of structures in a clade may reflect lack of analysis. Very long-chain fatty acids and structures with unusual double bond locations are widespread and are not displayed here.

Table 2 What branches of plant evolution are missing FA data? https://plantfadb.org/pages/Whats_Missing

Orders without FA data in PlantFAdb	Notes	Families	Genera	Species	Family names
Amborellales	*Amborella trichopoda* is only species. Molecular analyses place the genus at or near base of the flowering plant lineage	1	1	1	Amborellaceae
Berberidopsidales	**Berberidopsidales** is an order of Southern Hemisphere woody flowering plants.	2	3	4	Berberidopsidaceae; Aextoxicaceae
Bruniales	**Bruniales** not used at the rank of order until 2008 sister to the Apiales, and one of the asterid taxa.	2	14	79	Bruniaceae; Columelliaceae
Canellales	**Canellales** is one of four orders of the magnoliids; which comprise 136 species of fragrant trees and shrubs	2	9–13	136	Canellaceae; Winteraceae
Ceratophyllales	**Hornwort** order of flowering plants, consisting of a single family with one cosmopolitan genus (*Ceratophyllum*) containing 10 species	1	1	10	Ceratophyllaceae
Escalloniales	Member of the Asterids clade; ~130 species in 7 genera	1	7	130	Escalloniaceae
Huerteales	Shrubs or small trees found in most tropical or warm temperate regions. The flowers of *Perrottetia* have been studied in detail but otherwise the genera are poorly known	4	6	24	Petenaeaceae; Gerrardinaceae; Tapisciaceae; Dipentodontaceae
Paracryphiales	Order of woody shrubs and trees native to Australia, SE Asia, and New Caledonia.	1	3	36	Paracryphiaceae
Petrosaviales	Very small order of rare leafless achlorophyllous, mycoheterotrophic plants found in dark montane rainforests in Japan, China, Southeast Asia and Borneo	1	2	4	Petrosaviaceae
Sabiales	Represented by a single family (Sabiaceae), which is also considered by some as a member of Proteales.	1	3	66	Sabiaceae
Vahliales	Herbs and shrubs in Africa and Indian subcontinent. This family previously placed in Saxifragales order, and reassigned to Vahliales in 2016 by APG IV	1	1	5	Vahliaceae

Table 1 also indicates whether the biosynthetic pathway (s) are known, whether genes have been cloned, and whether transgenic plants have been engineered to produce the unusual fatty acids. For approximately one-third of the 20 structural categories, a biosynthetic pathway has not yet been established (including for cyclopropene, cyclopentene, phenyl) and therefore these structures represent attractive subjects for new discoveries.

Why there is great diversity in fatty acid structures in seed oils and how did it evolve?

The fatty acid composition in membrane lipids must be strictly controlled in order to maintain proper membrane functions. Only small numbers of fatty acids occur in membrane lipids and the fatty acid composition of these lipids is strikingly conserved throughout the plant kingdom. As many of the unusual fatty acids have in-chain polar groups, it can be anticipated that lipids containing such fatty acids would disturb bilayer formation or other membrane physical properties. The major known function of seed oil, conversely, is to serve as an energy and carbon source used in seed germination and early growth before photosynthesis starts. Therefore, it might be expected that there is little evolutionary pressure to maintain a certain fatty acid composition in the oil as long as it can be metabolized to support the germinating seed. Unusual fatty acids with epoxy and hydroxy groups, triple bonds and conjugated double bonds are synthesized from enzymes that have evolved from an ancestral $\Delta12$ fatty acid desaturase and occur in seed oils in many plant taxa and families. Likewise, medium-chain and very-long-chain fatty acids are synthesized by variants of enzymes (thioesterases and fatty acid elongases) expressed in all plants. This could lead to the conclusion that lack of evolutionary pressure has led to the great fatty acid diversity by mutation events in fatty acid metabolizing genes in many species. However, if such evolutionary changes were merely neutral, they would not be maintained but biochemical data clearly indicate that in seeds, where the unusual fatty acid occurs in high amounts, there has been an evolutionary advantage to produce an ever-increasing percentage of this fatty acid in the oil.

Many enzymes producing various unusual fatty acids have now been cloned and expressed in transgenic plants with the aim to produce these potentially valuable fatty acids in crop plants. This has in most cases led to transgenic plants producing seeds with only a fraction of the amount of the particular fatty acids seen in gene donor plants and often with compromised oil yield and seed germination (Napier, 2007). The reasons for these disappointing results are revealed in biochemical studies of seeds producing unusual fatty acids. From earlier studies of production of oil rich in petroselinic from coriander (Suh *et al.*, 2002), in decanoic acid from *Cuphea* (Bafor *et al.*, 1990) and ricinoleic acid in castor bean (Bafor *et al.*, 1991) it became evident that these plants have many enzymes (e.g. acyl-ACP thioesterases, phospholipid:diacylglycerol acyltransferases, diacylglycerol acyltransferases, lysophosphatidic acid acyltransferases) with high activity towards the unusual fatty acid. The low or modest amounts of unusual fatty acids produced in transgenic seeds expressing genes for production of the particular fatty acid have been significantly enhanced by co-expressing genes encoding some of these auxiliary enzymes (Burgal *et al.*, 2008; Kim *et al.*, 2015; Iskandarov *et al.*, 2017). Thus, seeds with high amounts of unusual fatty acids have recruited, or developed, a number of highly specialized enzymes to ensure that the produced fatty acid is efficiently channeled into the oil and excluded from membranes. This indicate that there is an evolutionary advantage for the plant to contain the unusual fatty acid in the oil and that the magnitude of this advantage has increased simultaneously with the stepwise increase in the amount of this fatty acid that must have taken place throughout evolution. A recent analysis of genes expressed during seed development in *Physaria fendleri* that accumulates ~60% hydroxy fatty acids and a closely related Brassicaceae species, *Camelina sativa*, concluded that there are likely more than 20 genes within several lipid pathways that co-evolved to facilitate high unusual fatty acid levels (Horn *et al.*, 2016). Vernolic acid, an epoxy fatty acid, is present in seed oils in plants from many orders. *Crepis palaestina,* in the Asterales order accumulates about 60% vernolic acid (9-epoxy-18:1Δ9) in its seed oil which it synthesizes from linoleic acid via a divergent FAD2 desaturase-like enzyme (Lee *et al.*, 1998). However, *Euphorbia lagascae* seeds (Malpighiales order) also accumulate vernolic acid, but it is synthesized from linoleic acid by a P450 enzyme that is unrelated to the desaturase type of enzymes (Cahoon *et al.*, 2002). This illustrates that the same unusual fatty acid in seed oils can be achieved by the action of evolutionary unrelated enzymes. It can therefore be speculated that plants also could have adopted different strategies in the evolution of many auxiliary enzymes that boost the level of the particular unusual fatty acid in their seed oils.

What is the evolutionary advantage of unusual fatty acids? The oil of castor bean, is comprised of 90% of the hydroxylated ricinoleic acid, which is a potent laxative. In addition to this inedible oil, castor seeds also contain the highly toxic ricin protein and thus have two lines of effective defense towards seed predators. Jojoba seeds contain oil that is nearly exclusively comprised of wax esters, instead of triacylglycerols, and cannot be digested by most animals. The cyclopentene fatty acids, such as chaulmoogric acid, occurring in high amounts in *Hydnocarpus* species, are both emetic and highly toxic fatty acids (Emerson, 1937). Crepenynic acid, occurring in high amounts in seed oils in at least four plant orders is a precursor for a great number of acetylenic compounds that have antibacterial and insecticide properties that are found in vegetative tissues in many plants (Bohlmann *et al.*, 1973). Interestingly, sunflower seeds, which do not contain any acetylenic fatty acid, express a gene encoding a Δ9 acetylenase upon fungal infection (Cahoon *et al.*, 2003), indicating that the enzyme responsible for crepenynic acid in seed oil of many Asteracae species likely evolved from an enzyme involved in pathogen defense.

Although biological properties of most unusual fatty acids have not been studied, it can be assumed that many other unusual fatty acids are toxic for microorganisms and animals or cause digestion problems for seed predators. For others, such as medium chain fatty acids (C8–C14) and very long chain fatty acids (e.g. erucic acid) this is unlikely. Instead they might have some yet unknown physiological advantages for the plant during seed development or seedling establishment.

Limitations to PlantFAdb

Although PlantFAdb is the most comprehensive database of its type and literature has been included through early 2018, some data have inevitably been missed. Some publications may have been overlooked because they did not match our keyword-based search strategies. Publications from sources that are not indexed or that are published in less common languages are most likely to have been overlooked. PlantFAdb also certainly includes some errors that arise from mistakes during manual entry of tabular data, from mistakes made by researchers in the identification of fatty acids, and from mistakes in the identification of plant species. Users are strongly encouraged to carefully review the original publications from which data in PlantFAdb were derived. In addition, it is important to note that some unusual fatty acids may not have been detected for a variety of reasons including that they are unstable during analysis, do not elute from gas chromatograms, etc. Spitzer (1999) provides a particularly useful guide for systematic analysis of unusual fatty acids in seed oils.

In addition, although some data are included from leaves or other non-seed tissues, PlantFAdb is focused almost entirely on analyses of seeds and fruits. Therefore, a number of unusual structures that occur in vegetative tissues, or in cutin, suberin, waxes or other extracellular lipids are not included in PlantFAdb. If other researchers have interest and expertise, additional structures and datasets can be added to PlantFAdb in the future.

Future

There have been no large-scale analytical surveys of plant fatty acid composition in recent years. As noted above, an effort to collect and analyze seeds from unsurveyed orders of the plant kingdom (highlighted in Table 2) should lead to the discovery of many new structures and expand our understanding of the diversity and evolution of fatty acid biosynthetic pathways. In addition to the unsurveyed orders, approximately 40% of all plant families have not yet been examined. Furthermore, much of the data represented in PlantFAdb were collected over 50 years ago, and in many cases without the benefits of high-resolution capillary GC columns or mass spectrometers. Thus, re-examination of some plant species is warranted and in some cases structures should be confirmed using more recent analytical technology.

Major advances in high-throughput and low-cost DNA sequencing now offer new opportunities to achieve a more precise genetic understanding of the steps underlying evolution of plant fatty acid diversity. For example, the 10 000 Plant Genome (10KP) project aims to build an annotated reference genome for a member of every genus of the Viridiplantae by 2023 (Cheng *et al.*, 2018). It is hoped that PlantFAdb will be helpful in bioinformatics strategies to expand the links between chemotaxonomy and emerging sequence data for plants producing all fatty acid structures. New tools will probably need to be developed to allow the automated integration of databases such as PlantFAdb with the ongoing release of new sequence datasets.

Request for community involvement

From 2015 to 2017 the development of PlantFAdb was supported by the Great Lakes Bioenergy Research Center. Future updates and development will depend on voluntary contributions. Users interested in contributing additional data are encouraged to contact the authors. If other plant chemists have databases for other classes of phytochemicals (terpenes, polyphenols, alkaloids, etc), they might consider creating similar tools. Software underlying PlantFAdb is freely available at: https://github.com/PlantFattyAcidGroup/plantFAdb.

METHODS

Literature

To retrieve published data on plant fatty acids, the literature was surveyed by keyword searches of the ISI-Web of Knowledge (https://webofknowledge.com/) and Google Scholar. Older literature in the SOFA database did not include titles, abstracts, and DOI. To retrieve these, author name, year and source information from SOFA was used to search the ISI-Web of Knowledge, Google Scholar, or Google. In some cases from older or lesser known sources, abstracts could not be found in electronic form and scanned copies of the publications were requested via interlibrary loan services and abstracts were entered manually.

Website and database

The PlantFAdb website was built using the Ruby on Rails web application framework (https://rubyonrails.org/) connected to a Postgres relational database management system (https://www.postgresql.org/). All data are parsed and imported into the relational database to generate unique identifiers and support advanced queries. The database has separate tables for entities including plants, publications, fatty acids, datasets and results. This framework supports dynamic data queries including the generation of data for the phylogenetic tree and direct links between various elements on the site such as the dynamic fatty acid links to phylogenetic data. The tree visualization is built with the d3 JavaScript library (http://d3js.org) and dynamically renders a custom hierarchy stored in the database combined with a bar chart of data for each leaf node in the tree. The bar chart values are generated by querying the stored result data for total count or maximum yield across the displayed plant and fatty acid selections. The tree hierarchy is based on the Angiosperm Phylogeny website (http://www.mobot.org/MOBOT/research/APweb/) and the Angiosperm Phylogeny poster (http://www2.biologie.fu-berlin.de/sysbot/poster/poster1.pdf). A drag and drop interface is available to manage the hierarchy including adding and removing items as needed. Initial datasets were parsed and loaded into the database using custom scripts. Data in the SOFA database were downloaded and reformatted to fit the new database schema. All plant names were reviewed and updated based on information at the Taxonomic Name Resolution Service from the iplant collaborative (http://tnrs.iplantcollaborative.org/). Common names of plants (English) were added based on the USDA database (plants.usda .gov/). Fatty acids were updated with additional CAS registry numbers, formula and mass curated with information from scifinder (https://scifinder.cas.org) or the Lipid Handbook CD-ROM. LipidMaps and Pubchem identifiers were manually curated from the LipidMaps (https://lipidmaps.org) and Pubchem databases and inchi, stdinchi, stdinchikey and smiles identifiers were imported from OPSIN (https://opsin.ch.cam.ac.uk/). Publications were cross-referenced with the Web of Science database and manually curated using journal websites, Google Scholar to update metadata, add DOI numbers and remove duplication. Additional datasets were entered online using forms available for manual data entry. All data can be manually curated and updated using the available interfaces. These changes are submitted as drafts for review before final publication and changes are stored permanently in a historical log. The data management interfaces are accessible on the website only from authenticated user accounts managed by the site administrator. Data entry forms are designed to enable collaboration among disparate researcher groups and support future contributions. Code for PlantFAdb is available at: https://github.com/PlantFattyAcidGroup/plantFAdb.

ACKNOWLEDGEMENTS

We thank Devin Higgins of MSU libraries for assistance with batch electronic searches of literature databases and Ajay W Tumaney (Central Food Technological Research Institute, Mysore, India) for assistance with retrieving information from the Journal of the Oil Technologists' Association of India. Lucas Busta (Univ. of Nebraska) provided helpful comments and suggestions. This work was supported in part by Department of Energy-Great Lakes Bioenergy Research Center Cooperative Agreement DE-FC02-07ER64494.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Table S1. Additions to SOFA that have been incorporated into PlantFAdb.

Table S2. Selected websites focused on lipids.

REFERENCES

Aitzetmüller, K., Matthäus, B. and Friedrich, H. (2003) A new database for seed oil fatty acids – the database SOFA. *Eur. J. Lipid Sci. Technol.* **105**, 92–103.

Badami, R.C. and Patil, K.B. (1980) Structure and occurrence of unusual fatty acids in minor seed oils. *Prog. Lipid Res.* **19**, 119–153.

Bafor, M., Jonsson, L., Stobart, A.K. and Stymne, S. (1990) Regulation of triacylglycerol biosynthesis in embryos and microsomal preparations from the developing seeds of *Cuphea lanceolata*. *Biochem. J.* **272**, 31–38.

Bafor, M., Smith, M.A., Jonsson, L., Stobart, K. and Stymne, S. (1991) Ricinoleic acid biosynthesis and triacylglycerol assembly in microsomal preparations from developing castor-bean (*Ricinus communis*) endosperm. *Biochem. J.* **280**(Pt 2), 507–514.

Baud, S. (2018) Seeds as oil factories. *Plant Reprod.* **31**, 213–235. https://doi.org/10.1007/s00497-018-0325-6

Bohlmann, F., Burkhardt, T. and Zdero, C. (1973) *Naturally Occurring Acetylenes*. London, New York: Academic Press.

Burgal, J., Shockey, J., Lu, C., Dyer, J., Larson, T., Graham, I. and Browse, J. (2008) Metabolic engineering of hydroxy fatty acid production in plants: RcDGAT2 drives dramatic increases in ricinoleate levels in seed oil. *Plant Biotechnol. J.* **6**, 819–831.

Cahoon, E.B., Carlson, T.J., Ripp, K.G., Schweiger, B.J., Cook, G.A., Hall, S.E. and Kinney, A.J. (1999) Biosynthetic origin of conjugated double bonds: production of fatty acid components of high-value drying oils in transgenic soybean embryos. *Proc. Natl. Acad. Sci. USA*, **96**, 12935–12940.

Cahoon, E.B., Ripp, K.G., Hall, S.E. and McGonigle, B. (2002) Transgenic production of epoxy fatty acids by expression of a cytochrome P450 enzyme from *Euphorbia lagascae* seed. *Plant Physiol.* **128**, 615–624.

Cahoon, E.B., Schnurr, J.A., Huffman, E.A. and Minto, R.E. (2003) Fungal responsive fatty acid acetylenases occur widely in evolutionarily distant plant families. *Plant J.* **34**, 671–683.

Carlsson, A.S., Yilmaz, J.L., Green, A.G., Stymne, S. and Hofvander, P. (2011) Replacing fossil oil with fresh oil - with what and for what? *Eur. J. Lipid Sci. Technol.* **113**, 812–831.

Cheng, S., Melkonian, M., Smith, S.A. *et al.* (2018) 10KP: a phylodiverse genome sequencing plan. *Gigascience*, **7**, 1–9.

Damude, H.G. and Kinney, A.J. (2008) Enhancing plant seed oils for human nutrition. *Plant Physiol.* **147**, 962–968.

Dyer, J.M., Stymne, S., Green, A.G. and Carlsson, A.S. (2008) High-value oils from plants. *Plant J.* **54**, 640–655.

Emerson, G.A. (1937) Influence of dehydrocholic acid (Decholin, NNR) on oral toxicity of chaulmoogra oil. *Int. J. Lepr.* **5**, 159–162.

Gunstone, F.D., Harwood, J.L. and Dijkstra, A.J. (2007) *The Lipid Handbook With CD-ROM*. Boca Raton, FL: CRC Press.

Horn, P.J., Liu, J., Cocuron, J.C., McGlew, K., Thrower, N.A., Larson, M., Lu, C., Alonso, A.P. and Ohlrogge, J. (2016) Identification of multiple lipid genes with modifications in expression and sequence associated with the evolution of hydroxy fatty acid accumulation in Physaria fendleri. *Plant J.* **86**, 322–348.

Iskandarov, U., Silva, J.E., Kim, H.J., Andersson, M., Cahoon, R.E., Mockaitis, K. and Cahoon, E.B. (2017) A specialized diacylglycerol acyltransferase contributes to the extreme medium-chain fatty acid content of *Cuphea* seed oil. *Plant Physiol.* **174**, 97–109.

Kim, H.J., Silva, J.E., Vu, H.S., Mockaitis, K., Nam, J.W. and Cahoon, E.B. (2015) Toward production of jet fuel functionality in oilseeds: identification of FatB acyl-acyl carrier protein thioesterases and evaluation of combinatorial expression strategies in Camelina seeds. *J. Exp. Bot.* **66**, 4251–4265.

Lee, M., Lenman, M., Banas, A. *et al.* (1998) Identification of non-heme diiron proteins that catalyze triple bond and epoxy group formation. *Science*, **280**, 915–918.

Li, C., Cheng, X., Jia, Q., Song, H., Liu, X., Wang, K., Zhao, C., Zhang, Y., Ohlrogge, J. and Zhang, M. (2017) Investigation of plant species with identified seed oil fatty acids in Chinese literature and analysis of five unsurveyed Chinese endemic species. *Front. Plant Sci.* **8**, 224.

Li, X., Teitgen, A.M., Shirani, A. *et al.* (2018) Discontinuous fatty acid elongation yields hydroxylated seed oil with improved function. *Nat. Plants*, **4**, 711–720.

van de Loo, F.J., Fox, B.G. and Somerville, C. (1993) Unusual fatty acids. In *Lipid metabolism in plants* (Thomas S Moore., ed). Milton: CRC Press, pp. 91–126.

van de Loo, F.J., Broun, P., Turner, S. and Somerville, C. (1995) An oleate 12-hydroxylase from Ricinus communis L. is a fatty acyl desaturase homolog. *Proc. Natl. Acad. Sci. USA*, **92**, 6743–6747.

Matthäus, B. (2012) The database seed oil fatty acids (SOFA) is back on the Internet!. *Eur. J. Lipid Sci. Technol.* **114**, 701–702.

McKeon, T.A. (2016) *Industrial Oil Crops*. Amsterdam: Academic Press and AOCS Press.

Napier, J.A. (2007) The production of unusual fatty acids in transgenic plants. *Annu. Rev. Plant Biol.* **58**, 295–319.

Shanklin, J. and Cahoon, E.B. (1998) Desaturation and related modifications of fatty acids. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49**, 611–641.

Smith, C.R. (1971) Occurrence of unusual fatty acids in plants. *Prog. Chem. Fats Other Lipids*, **11**, 137–177.

Spitzer, V. (1999) Screening analysis of unknown seed oils. *Fett-Lipid*, **101**, 2–19.

Suh, M., Schultz, D. and Ohlrogge, J. (2002) What limits production of unusual monoenoic fatty acids in transgenic plants? *Planta*, **215**, 584–595.

Vanhercke, T., Wood, C.C., Stymne, S., Singh, S.P. and Green, A.G. (2013) Metabolic engineering of plant oils and waxes for use as industrial feedstocks. *Plant Biotechnol. J.* **11**, 197–210.