A Mutation Breeding Program to Improve the Quality of the Oil Crop *Crambe abyssinica*

Jihua Cheng

Thesis committee

Promotor

Prof. Dr R.G.F. Visser

Professor of Plant Breeding

Wageningen University

Co-promotors

Dr. E.M.J. Salentijn

Researcher, Wageningen UR Plant Breeding

Wageningen University & Research Centre

Dr. E.N. van Loo

Researcher, Wageningen UR Plant Breeding

Wageningen University & Research Centre

Other members

Prof. Dr. H.J. Bouwmeester, Wageningen University

Prof. Dr. G. Eggink, Wageningen University

Dr. H.A. Mooibroek, Wageningen University

Prof. Dr. I. Feussner, Georg-August-University Goettingen, Germany

This research was conducted under the auspices of the Graduate School of Experimental Plant Sciences.

A Mutation Breeding Program to Improve the Quality of the Oil Crop *Crambe abyssinica*

Jihua Cheng

Thesis

submitted in fulfilment of the requirement of the degree of doctor

at Wageningen University

by the authority of the Rector Magnificus

Prof. Dr. M.J. Kropff,

in the presence of the

Thesis Committee appointed by the Academic Board

to be defended in public

on Thursday 12 June 2014

at 11 a.m. in the Aula.

Jihua Cheng

A mutation breeding program to improve the quality of the oil crop *Crambe abyssinica*

180 pages.

PhD thesis, Wageningen University, Wageningen, NL (2014)

With references, with summaries in English and Dutch

ISBN: 978-94-6173-962-9

Contents

Chapter 1 General introduction..... 5

Chapter 2 Isolation and characterization of the omega-6 fatty acid desaturase (FAD2) gene family in the allohexaploid oil seed crop Crambe abyssinica Hochst..... 21

Chapter 3 Functional analysis of the omega-6 fatty acid desaturase (CaFAD2) gene family of the oil seed crop *Crambe abyssinica*..... 57

Chapter 4 High-throughput discovery of induced mutations in the allohexaploid oil crop *Crambe abyssinica* using 454 sequencing..... 85

Chapter 5 Detection of induced mutations in *CaFAD2* genes by next generation sequencing leading to the production of improved oil composition in *Crambe abyssinica*..... 113

Chapter 6 General discussion..... 149

Summary..... 163

Samenvatting..... 167

Acknowledgement

About the author

Chapter 1

General introduction

Industrial use of plant oils

Since the last century, our society relies heavily on petroleum, which so far provides not only the energy for transportation and industry but is also a substantial feedstock in the manufacture of all kinds of products which are used on a daily basis in food and feed applications. Due to the finite nature of petroleum resources, the price and supply of petroleum are major issues worldwide. Another problem is that many petroleum derivatives are highly polluting for the environment. These issues motivated researchers, companies and governments to seek sustainable alternatives to petroleum. A transition from an economy based on mineral oil to a plant-based economy for the supply of fuel, building materials and all kinds of industrial products seems the most logic one. Most plant oils are comprised of a range of fatty acids with chain lengths from 16 to 18 carbons such as palmitic (C16:0), stearic (C18:0), oleic (C18:1), linoleic (C18:2) and linolenic acids (C18:3) (Carlsson 2009). Furthermore, there are a large number of unusual fatty acids produced by specific plant species. Such fatty acids include those with chain lengths between 8 and 24 carbons (Carlsson 2009). Chemically, plant oils have structures and properties that are similar to petroleum and therefore offer an excellent alternative (Carlsson 2009).

There are several criteria for considering a crop as a platform for industrial oil. Firstly, the produced oil should be of high quality with a desirable fatty acid profile; secondly, the cost of oil should be competitive, which requires that the oil production depends on low inputs; lastly, the oil production should have no significant impact on the environment including, for example the oil crop used should have a low risk of out-crossing to prevent that these industrial oils mix in the food chain.

To date, many oil crops have been highlighted as potentially useful in industry. Their usability as industrial crops was proposed because of the presence of special fatty acids in their oil (Dyer et al. 2008) (**Table 1**). The major oil crops including oil palm, soybean, rapeseed, sunflower, peanut, cottonseed, corn and olive provide most of the food oil used for (human) consumption. Meanwhile, a significant proportion (around 20 %) also serves as industrial (non-food) oil (Vanhercke et al. 2013). One major industrial utilization of plant oils involves medium-chain fatty acids e.g. laurate (C12:0) in surfactant applications, such as in soaps, detergents and related personal care products. Other highly unsaturated food oils, such as linseed and soybean oils, also find uses as renewable alternatives for synthetic drying agents

in surface coatings and inks and are also processed into epoxygenated oils for use in industrial resins and glues (Vanhercke et al. 2013).

In addition, a few minor oil crops can produce novel oil types containing unusual fatty acids, like castor bean (*Ricinus communis*) (Rivarola et al. 2011), crambe (*Crambe abyssinica*) (Lalas et al. 2012), tung tree (*Aleurites fordii*) (Yang et al. 2013) and jojoba (*Simmondsia chinensis*) (Kumar and Sharma 2008) (**Table 1**). The oils from such crops are dedicated to a number of specific applications. Castor oil accumulates in the endosperm and contains very high levels (90%) of the hydroxy fatty acid ricinoleic acid (12-hydroxy C18:1$^{\Delta 9}$) (**Table 1**), which is an important industrial feedstock for the production of the polyamide Nylon 11 and of polyurethane, a high-performance lubricant. Tung oil (or China wood oil) obtained from the nut of the tung tree is highly valuable because of its unique drying property and very high content (82%) of the conjugated fatty acid α-eleostearic acid (C18:3$^{\Delta 9c,11t,13t}$) (**Table 1**). Therefore, this oil is useful for furniture protection. Crambe oil naturally contains up to 60% of erucic acid (C22:1$^{\Delta 13}$), which is a useful raw material for industrial manufacture of plastics and lubricants (**Table 1,** more details on the use of crambe oil will be discussed below). Jojoba is a medicinal- and oil yielding species. The most valuable product of jojoba is the wax ester in its seed oil with a wax ester concentration of up to 50-55% of the seed weight. This oil is used extensively in the cosmetics industry (Kumar and Sharma 2008, Kumar et al. 2012) (**Table 1**).

Table 1 Fatty acid composition (%) in the oil of several crops.

	C14:0	C16:0	C18:0	C18:1	C18:2	C18:3	C20:1	C22:1	C22:1	Others	Reference
Palm	5	36	2	50	8						(Gunstone and Harwood 2007)
soybean		11	4	23	54	8					(Gunstone and Harwood 2007)
Canola		4	2	60	21	10	1				(Gunstone and Harwood 2007)
Sunflower		7	5	19	68						(Gunstone and Harwood 2007)
Linseed		6	2	19	24	47					(Gunstone and Harwood 2007)
Brassica napus		5	1	15	14	9	7		45		(Ackman 1990)
Camelina sativa		7	2.5	15.4	21.5	33.2	12.6	12.6	2.2		(Budin et al. 1995)
Crambe abyssinica		2	1	18	9	6	2		56		(Lazzeri et al. 1997)
Castor		1	1	3	4					89 (Hydroxy FA)	(Ogunniyi 2006)
Jojoba	0.04	0.9	0.1	0.1	3	0.2	0.4		22	50 (Wax ester)	(Miwa 1971)
Tung		2.6	2.7	6	7.4	70					(Suzuki et al. 2001)

Metabolic pathway of seed oil biosynthesis

The broad applications of plant oils motivated researchers to uncover the pathways leading to seed oil, which in most plants consists of fatty acids esterified to three hydroxyl groups of glycerol (so called triacylglycerol, TAG). Therefore, many studies have been done to isolate and characterize critical genes in this pathway. The process of seed oil biosynthesis has been studied extensively and its understanding has increased tremendously. A recent review summarizes the details of pathways in seed oil synthesis (Bates et al. 2013). Overall, the synthesis of fatty acids and TAG involves multiple organelles in cytoplasm. Fatty acid (FA) biosynthesis begins in plastids whereas TAG assembly occurs outside the plastid and TAG may be incorporated into both the endoplasmic reticulum (ER) and the oil body (Chapman et al. 2012, Chapman and Ohlrogge 2012, Liu et al. 2012).

Initially, free FAs (mostly C16:0 and C18:1) are *de novo* synthesised in the plastid and this pathway determines the length of carbon chain (up to 18 carbons) and the levels of saturated FAs in seed oil (Nikolau et al. 2003) (**Figure 1**, reaction 1 to 5). The resulting FAs in the plastid react with coenzyme A to form the acyl-CoA (acyl-CoA pool) (**Figure 1**, reaction 6 to 9) which is the substrate of TAG assembly. The acyl-CoA, e.g. C18:1-CoA, is then somehow transported into ER where C18:1-CoA can be either directly assembled to TAG or elongated to very-long-chain-fatty-acid (VLCFA, i.e. a fatty acid with more than 18 carbon atoms).

VLCFAs result from the endoplasmic reticulum (ER)-associated acyl-CoA elongase activity, which is catalyzed by multi-enzymatic fatty acid elongase (FAE) complexes (Joubès et al. 2008). The elongation of the carbon chain of FAs is performed by elongation reactions, generating an acyl chain extension (two carbons at a time). Each elongation reaction consists of four sequential steps: (i) a condensation of malonyl-CoA with an acyl-CoA catalysed by a fatty acid elongase (β-ketoacyl-CoA synthase, KCS/FAE1), (ii) the resulting 3–ketoacyl-CoA is then reduced by 3–ketoacyl-CoA reductase (KCR) into 3–hydroxyacyl-CoA, (iii) which is subsequently dehydrated to a trans-2,3-enoyl-CoA by a β-hydroxyacyl-CoA dehydratase (HCD), and (iv) the trans-2,3-enoyl-CoA is reduced by the trans-2,3-enoyl-CoA reductase (ECR), which yields an acyl-CoA elongated by two carbons (ECR) (**Figure 1**, reaction 10 and 11). The resulting VLCFA and C18:1 are incorporated into glycerol-3-phosphate (G-3-P) to form triacylglycerol (TAG) via four sequential reactions, the so called *de novo* TAG assembly (also known as the Kennedy pathway (**Figure 1**, reaction 12 to 15) (Weiss et al. 1960, Kennedy 1961, Barron and Stumpf 1962).

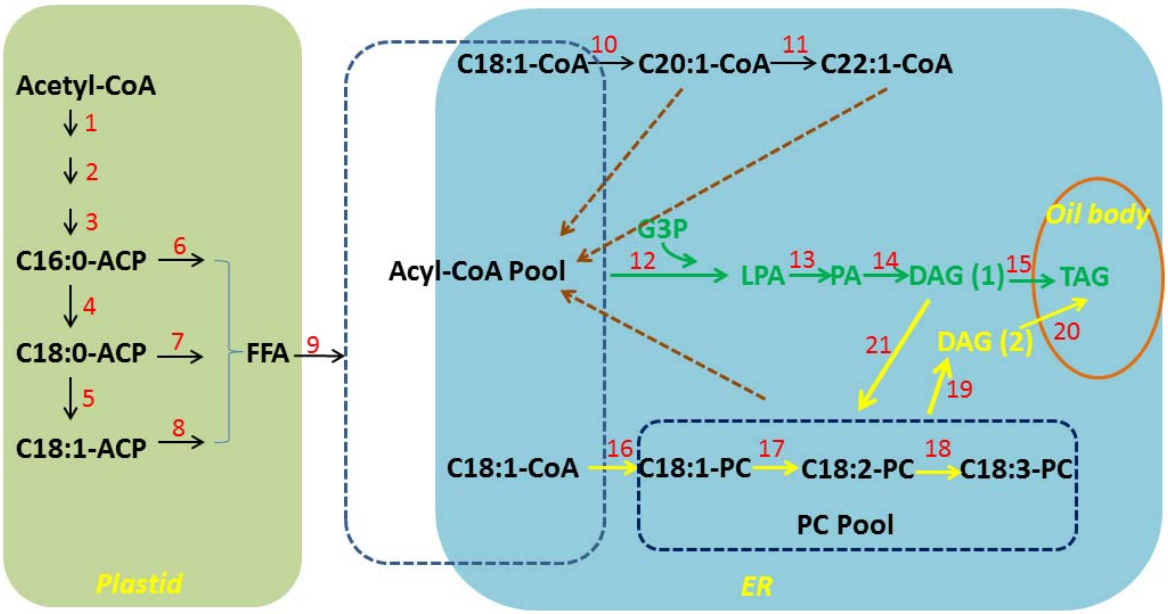


Figure 1. The overview of major reactions in seed oil biosynthesis. Reactions of *de novo* synthesis and modification of Fatty acids are depicted as black arrows. Acyl transfer reactions are dashed lines. Green lines are *de novo* TAG synthesis; yellow lines are PC-derived DAG synthesis. The DAG(1) is de novo synthesized DAG and DAG(2) is PC-derived DAG. Abbreviations: **ACP**, acyl carrier protein; **FFA**, free fatty acid; **G3P**, glycerol-3-phosphate; **LPA**, lyso-phosphatidic acid; **PA**, phosphatidic acid; **DAG**, diacylglycerol; **TAG**, triacylglycerol; **PC**, phosphatidylcholine. Enzymes in reactions: 1) ACCase, acetyl-CoA carboxylase; 2) **MT**, malonyl CoA-ACP transferase; 3) **FAS**, fatty acid synthase; 4) **KASII**, ketoacyl-ACP synthase II, 5) **SAD**, Stearoyl-ACP desaturase; 6) and 7) **FATB**, acyl-ACP thioesterase B; 8) **FATA**, acyl-ACP thioesterase A; 9) **LACS**, long chain acyl-CoA synthetase; 10) and 11) **FAE**, fatty acid elongase complex; 12) **GPAT**, acyl-CoA:G3P acyltransferase; 13) **LPAAT**, acyl-CoA:LPA acyltransferase; 14) **PAP**, PA phosphatase; 15) **DGAT**, acyl-CoA:DAG acyltransferase; 16) **LPCAT**, acyl-CoA:LPC acyltransferase; 17) **FAD2**, omega-6 fatty acid desaturase, 18) **FAD3**, omega-3 fatty acid desaturase; 19) **PLC**, phospholipase C; 20) **PDAT**, phospholipid:diacylglycerol acyltransferase; 21) **PDCT**, PC:DAG cholinephosphotransferase.

In addition to the *de novo* TAG assembly, FAs fluxes within the seed could occur through a more complex pathway, which involves the membrane phosphatidylcholine (PC) as a central intermediate in the flux of substrates into TAG. In this pathway, the FAs are esterified into the PC (**Figure 1**, reaction 16) and undergo modifications such as desaturation (**Figure 1**, reaction 17 and 18) and hydroxylation (Sperling et al. 1993, Van De Loo et al. 1995). The resulting FA-PCs are eventually converted to diacylglycerol (DAG) and TAG (**Figure 1**, reaction 19 and 20). Therefore, acyl flux into and out of PC determines the levels of PC-

modified FAs in TAG such as the level of polyunsaturated fatty acids (PUFA) (Wallis and Watts 2002).

Crambe abyssinica

Crambe (*Crambe abyssinica* Hochst. ex Fries, 6x=2n=90) is an annual herb of the Brassicaceae/Cruciferae family, related to rape and mustard. Crambe is considered to originate from Ethiopia and adapted in the Mediterranean region (Knights 2003, Rogério et al. 2013). The height of crambe plant varies from 0.5-2 m with white flowers (Falasca, 2010) (**Figure 2**). Crambe, a self-pollinated plant, has a short cropping cycle, ranging from 90 to 100 days (Falasca et al. 2010, Santos et al. 2013). Flowering occurs approximately 35 days after sowing (Rogério et al. 2013), and lasts approximately three weeks, followed by a period of seed maturation of about 30 days. The mature flowers produce small seed capsules (pods) of ca 5 mm in diameter, containing a single and spherical seed (**Figure 2**).

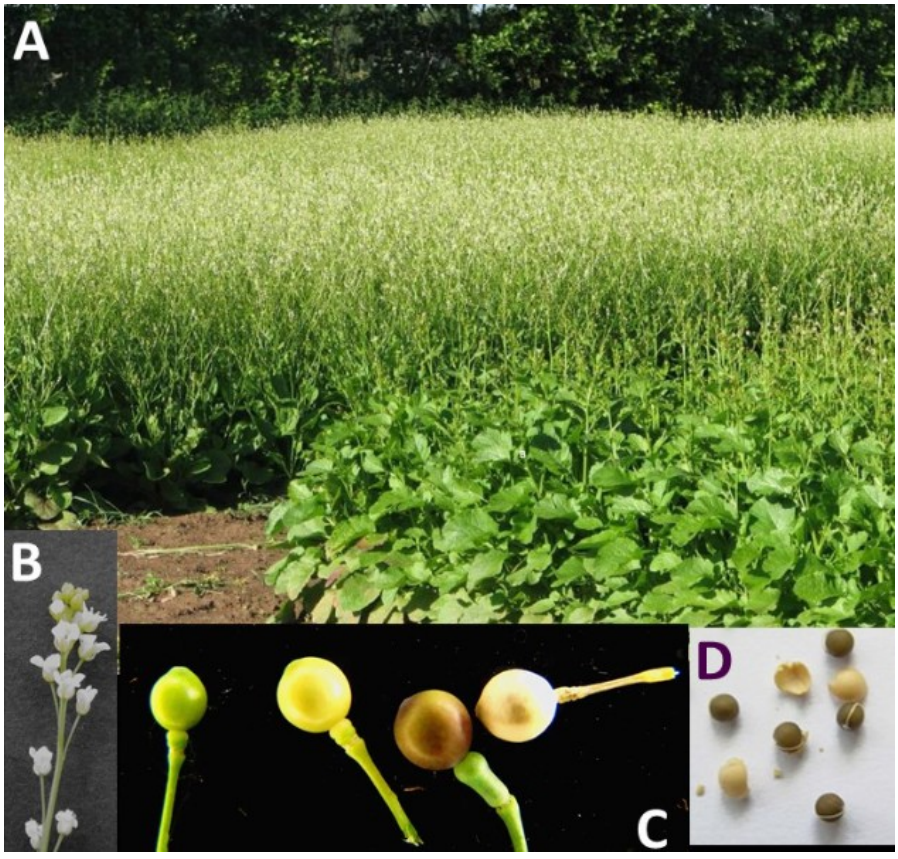


Figure 2. *C. abyssinica* plants and tissues. A) flowering plants; B) white crambe flowers; C) developing seeds; D) dehulled seeds.

Crambe has a great tolerance to drought, pests and diseases, and it can be cultivated around the world with low input requirements when compared to other oilseed crops. Crambe can

grow as spring crop in Europe or as a winter crop in the Mediterranean region. Nowadays, crambe is cultivated in many countries for production and research.

Little is known about the agricultural cultivation of crambe in the past. The cultivation of crambe and its wild relatives can only be traced back to the 1930s in the former Sovjet Union (USSR). The crambe varieties were improved since their introduction to other European countries like The Netherlands, Poland and Sweden (Mastebroek et al. 1994). At the beginning of the 21st century, *C. abyssinica* was cultivated in America, Canada, several European countries, Pakistan, India and China (Fontana et al. 1998, Wang et al. 2000, Lalas et al. 2012). Basically, crambe is still under development as an agricultural crop so that there is only limited yield data available. Crambe seed yields widely ranged from 1,000 to 1,600 kg per hectare (kg/ha) in Russia compared to 450 to 2,500 kg/ha in USA. In European countries, crambe yielded up to 3,300 kg/ha with an average of 2,300 kg/ha (Zelmer and McVetty 2009).

Oil composition in crambe seed and its uses

Interests in crambe arise from crambe seed-oil and the by-product seed meal, which is a residual after oil extraction (Downey 1971, Nieschlag and Wolff 1971). Crambe seed contains about 35 % of oil, and the major fatty acid components range from 16 to 24 carbon atoms in chain length. Crambe oil is rich in erucic acid (C22:1, over 60%) and oleic acid (C18:1, 12%), and it also contains a large fraction of polyunsaturated fatty acid (PUFA), i.e. linoleic acid (C18:2) and linolenic acid (C18:3) (**Table 1**). Because of the high content of C22:1, crambe oil is not suitable for use in food applications and it has thus been considered to be a candidate industrial crop (Carlsson, 2006). An established market for high-erucic acid oils is erucamide, a slip agent critical to the production of polyolefin films and polyethylene such as bread wrappers, shopping and garbage bags, shrink wraps, and plastic sheeting (Friedt and Lühs 1998, Lalas et al. 2012). By chemical splitting using for example ozonolysis, erucic acid yields two primary products, brassylic acid (C13) and pelargonic acid (C9) (Lalas et al. 2012). Brassylic acid can be used in the manufacture of polyester, plasticizers, alkyd resins, lubricants, rubber additives, surface-active agents, nylon, and other polymers (Santos et al. 2013). Pelargonic acid finds use in the field of plasticizers, alkyd resins, vinyl stabilisers, hydrotropic salts, pharmaceuticals, synthetic flavours, and flotation agents (Lalas et al. 2012).

A high C22:1 content is unfavourable in food oil but preferred in industrial oil. Currently, high C22:1 plant oil is mainly provided by high-erucic acid rapeseed (HEAR) (Lessman 1990), which contains up to 54% of C22:1. However, production of HEAR raises a risk that

food type oilseed rape varieties (low-erucic rapeseed) become contaminated by the industrial oil type of HEAR because the two types are from the same species and cross-fertilise. So, HEAR must be grown at appropriate isolation distances from food oilseed rape varieties. In contrast, crambe and other food oil crops are not interfertile in nature so that this risk does not exist when cultivating crambe as a source of C22:1. Therefore, many countries (Austria, Belgium, Denmark, France, Germany, Italy, Portugal, the UK, and Argentina) have shown a scientific interest in crambe (Lessman 1990, Laghetti et al. 1995, Klaus 1999, Knights 2003, Falasca et al. 2010, Zanetti et al. 2010).

Meal of crambe seed

Crambe seeds contain 40 % of oil and 60 % of meal, so after the process of oil extraction a considerable amount of residues is left (‘seed cake’; ‘press cake’; ‘meal’), which is an excellent protein source for livestock (VanEtten et al. 1965, Yong-Gang et al. 1993). Dehulled crambe meal contains a substantial amount of protein and a well-balanced amino acid profile showing a good nutritional quality (Anderson et al. 1993). In crambe meal, the concentration of most of the essential amino acids are comparable or even higher to that of rapeseed (Lessman 1990). However, like other *Brassicaceae* crops, crambe meal contains glucosinolates (8 % to 10 %), even at a higher concentration than traditional rapeseed (3 % to 7 %). The predominant glucosinolate in crambe seeds, epi-progoitrine, is a strong anti-nutritional factor. The glucosinolate fraction is readily hydrolysed to isothiocyanates by myrosinases, which together with glucosinolates themselfes may impart unpalatability or toxicity to seed meals (VanEtten et al. 1965, Lessman 1990). So, glucosinolates limit the potential use of crambe meal (Anderson et al. 1993). Although some approaches, e.g. heat-toasting, chemical and microwave treatments are able to decrease the content of glucosinolates considerably, the treated meal is still unsuitable to monogastric animals because some toxic products from the hydrolysis of glucosinolates remain (Yong-Gang et al. 1993, Mastebroek et al. 1994, Yong-Gang et al. 1994). Therefore, a limited amount of crambe meal is allowed to feed animals, for example in the diet of cattle the inclusion of crambe meal should amount to less than 4.2 % (FDA).

Efforts for crambe breeding

Although since the 1950s breeding efforts resulted in crambe varieties for the commercial production of high C22:1 oil, crambe has not yet become established as a crop. The primary breeding objectives for crambe include a higher seed yield, with a higher oil content, a higher

fraction of C22:1, a lower glucosinolate content in the seed and more tolerance to diseases. The selection in crambe started in Russia and Sweden in the late 1950s. Subsequently, several cultivars (bred on the basis of mass selection) were released from the 1960s to 1985 in the US (Mastebroek et al. 1994).To date, only few crambe cultivars (less than 10) are available on the market (Rudloff and Wang 2011).

Successful plant breeding depends first of all on the level of genetic variation that is present in the starting material. The challenge in breeding with regard to the improvement of crambe is the lack of adequate genetic variation for important agronomic traits (Lessman and Meier 1972). Therefore, interspecific hybridization or mutagenesis was recommended to produce new crambe cultivars with desirable traits (Lessman and Meier 1972). Thus, Campbell (1986) introgressed wild crambe germplasm into the crambe cultivar 'Indy' to produce two new cultivars and crambe lines, with a greater tolerance to diseases.

However, breeders have consistently reported that *C. abyssinica* has insufficient variation in its important traits to expect substantial progress (Lessman 1975, Mastebroek et al. 1994). As an alternative to traditional breeding, genetic engineering has been used to improve oil quality in crambe. Through regulating the expression of endogenous genes or/and introducing genes of interest (heterologous genes; from other species), crambe lines with higher erucic acid and/or oleic acid content in seeds were established (Li et al. 2012; Chapter 3 in this thesis). Also, efforts are underway to produce valuable specialties such as wax esters in crambe oil (W. Qi, Wageningen UR Plant Breeding, WUR; unpublished results). However it is difficult and expensive to market such genetically modified crambe lines, due to the strict legislations for genetically-modified (GM) organisms, especially in Europe.

EMS mutagenesis to increase genetic diversity for crop improvement

The apparent absence of sufficient natural variation in crambe (together with a continuous reduction in diversity due to the recurrent selection for some specific traits) has eventually become a bottleneck for crambe improvement using traditional breeding. In such a case, the genetic diversity of the germplasm can be enriched with various non-GM techniques. A rapid method to introduce mutations in genomes is chemical mutagenesis which utilises chemical mutagens such as ethyl methane sulphonate (EMS) (Sikora et al. 2011). EMS treatment induces the formation of random point mutations, predominantly C-to T and G- to A transitions, throughout the genome. In this way a wide range of genomic variation is created

including allelic series of variants, containing functional knock-outs or variants with altered functionality (Wells et al. 2014).

Traditional mutagenesis has been widely used as a forward genetic strategy to create variation in plants followed by the identification of selectable phenotypes. A reverse genetics screening method was developed (TILLING, Targeting Induced Local Lesion IN Genomes) to discover induced mutations at the DNA level, which is suitable for functional analysis in most plants (McCallum et al. 2000). Routinely the procedure consists of three steps: i) Development of a mutagenizing population using EMS or other mutagens, ii) DNA isolation and pooling, iii) discovery of mutations in genes. The details of each step are summarized by Till (2006, 2007).

Any gene of interest can be the target for mutagenesis and a mutant population can be used as a resource for crop improvement on various traits. Mutagenesis introduces allelic series of mutations including nonsense mutations and missense mutations which may result in variable phenotypes (Perry et al. 2009). Interesting lines with specific mutations in target genes were identified in different species including soybean (Cooper et al. 2008), maize (Till et al. 2004), *Brassica rapa* (Stephenson et al. 2010), *Brassica napus* (Wang et al. 2008), bread wheat and durum wheat (Slade et al. 2005). These successes indicate that nowadays mutation breeding is generally applicable to genomes whether small or large, diploid or even polyploid (e.g. like crambe which is allohexaploid).

Furthermore, in polyploid species genomic redundancy is causing tolerance to high mutation densities and to the negative effects introduced by mutagenesis (Tsai et al. 2013). However, the high sequence identity between homoeoloci in polyploid genomes provides difficulties for PCR-based targeted mutation screening methods that require the development of PCR--primers for the production of amplicons from a unique locus, so that the application in polyploid species is complicated by this problem. Since the introduction of next generation sequencing technologies (NGS) mutation breeding has been moved to a new era of -mutation detection by sequencing. It has become realistic to identify polymorphisms in any gene of interest through large-scale, high-throughput sequencing using NGS platforms. In this way existing natural polymorphisms or induced mutations in plant populations can be identified with high efficiency and throughput (Gilchrist and Haughn 2010). Using NGS, such as the Illumina platform, Tsai (2011) described mutation detection experiments using EMS-mutant populations of diploid rice and tetraploid durum wheat, and stated in the same publication that they also have used the protocol for screening in tomato (*Lycopersicum esculentum*),

Arabidopsis, and camelina (*Camelina sativa*). EMS mutagenesis can be also used as a tool in forward genetics (from phenotypes to the associated gene discovery). Conventional genetic mapping requires a mapping population that is established by crossing of two diverged accessions. Alternatively, a mapping population can be also established by crossing a genotype with an induced mutant that is from the same genotype. By use of this population, candidate genes can be identified using a so-called Mutmap constructed by the mutagen-induced changes as segregating markers (Abe et al. 2012, Hartwig et al. 2012, Zhu et al. 2012).

Thesis outline

Since *C. abyssinica* is considered to be an ideal crop for industrial oil, further improving the value of crambe oil is of importance. The value of crambe oil can be improved by increasing the C22:1 content or reducing polyunsaturated fatty acids (PUFA). In the separation of C22:1 fatty acids with distillation two fractions arise: a fraction of all C18 fatty acids is released in distillation, forming the so-called 'top fatty acids', and a fraction with all C22:1 (and longer FA). The value of the top fatty acids for industrial use can be highly increased when it would consist of only C18:1. However, currently the level of PUFA in the top fatty acids can be as high as 50 % (starting from 16 % in the seed oil). The gene *FAD2* plays a critical role in PUFA biosynthesis, so natural allelic variations and mutations in *FAD2* are able to result in "high oleic acid, low PUFA" phenotypes. The aims of the research described in this thesis were: first, to isolate the *FAD2* gene family members from *C. abyssinica* and analyse their functions and second, to develop a platform for high-throughput mutation detection with next generation sequencing (NGS), third, to use the platform to discover EMS induced mutations in *FAD2* genes in a crambe mutant population and to characterise the oil properties of the identified mutants.

In allohexaploid crambe seven copies of *FAD2* (*CaFAD2-A, -B, -C1, -C2*, *-C3, D and E*) were isolated. Three genes (*CaFAD2-C1, -C2* and *-C3*) were expressed in multiple tissues including root, seedling, leaf, flower, flower bud and developing seeds. The other four genes coded for truncated proteins and were not expressed in any of these tissues. Among six selected crambe breeding lines, a low level of natural variation was shown in *CaFAD2* expression which could not explain the variation in oil composition (**Chapter 2**). In **Chapter 3**, the function of *CaFAD2* was confirmed using RNAi to silence the endogenous *CaFAD2* genes in crambe. In *CaFAD2*-RNAi transgenic lines, *CaFAD2* genes were silenced to various

degrees and this silencing resulted in significant increases in oleic acid content and reduction in PUFA content in seed oil compared to the control. Based on correlation analysis, we concluded that *CaFAD2-C3* is a promising candidate gene determining the PUFA levels in seed oil.

To improve crambe oil by a non-GM approach, a mutant population (12,480 lines) was established from the EMS-treated seeds of *C. abyssinica* cv. 'Galactica'. Furthermore, the mutants in *CaFAD2* were detected by 454-amplicon sequencing (**Chapter 4**) and Illumina sequencing (**Chapter 5**). With 454-sequencing, ten mutants were detected among 1,860 lines. The mutations were detected in four *CaFAD2* genes (*CaFAD2-A*, *-C1*, *-C2* and *-C3*), and for five mutations, including four missense mutants and one silent mutation, the individual plants carrying the homozygous mutation were identified. However, these mutants showed the same phenotype in oil composition as the wild-type (**Chapter 4**). With Illumina sequencing, in total 11,100 plants were screened for mutations in a conserved part of the sequence known to be important for FAD2 function. Subsequently, five novel mutations (three nonsense and two missense) were found in *CaFAD2-C*. Of these five mutations, the mutants corresponding to four mutations (two nonsense in respective *CaFAD2-C1* and *-C3*, two missense in respective *CaFAD2-C1* and *-C2*) were identified. The heterozygous and homozygous mutants were identified for all the mutations with exception of the nonsense mutation in *CaFAD2-C1* for which only a heterozygous mutant was identified. The oil compositions in these mutants were examined. It was found that the nonsense homozygotes of *CaFAD2-C3* showed an increased content of C18:1 (+25 %) and a reduced content of PUFA (−25 %) compared to the segregated wild-type. Compared to the nonsense heterozygote of *CaFAD2-C3*, the nonsense heterozygote of *CaFAD2-C1* showed a significant difference in the C18:1(2 % decrease) and C22:1 (7 % decrease) contents. The missense homozygote of *CaFAD2-C1* showed an alteration in C18:1 and PUFA (with a 5 % increase and an 8 % decrease relative to the wild-type). The remaining missense mutant for *CaFAD2-C2* showed almost wild-type typical oil composition (**Chapter 5**).

In the final chapter (**Chapter 6**), the results obtained from the previous chapters are discussed together with their impact on crambe breeding in the future.

References

Abe, A., S. Kosugi, K. Yoshida, S. Natsume, H. Takagi, H. Kanzaki, H. Matsumura, K. Yoshida, C. Mitsuoka, and M. Tamiru. 2012. Genome sequencing reveals agronomically important loci in rice using MutMap. Nature Biotechnology **30**:174-178.

Ackman, R. 1990. Canola fatty acids—an ideal mixture for health, nutrition, and food use. Pages 81-98 Canola and Rapeseed. Springer.

Anderson, V. L., W. D. Slanger, S. L. Boyles, and P. T. Berg. 1993. Crambe meal is equivalent to soybean meal for backgrounding and finishing beef steers. Journal of animal science **71**:2608-2613.

Barron, E. J. and P. Stumpf. 1962. Fat metabolism in higher plants XIX. The biosynthesis of triglycerides by avocado-mesocarp enzymes. Biochimica et biophysica acta **60**:329-337.

Bates, P. D., S. Stymne, and J. Ohlrogge. 2013. Biochemical pathways in seed oil synthesis. Current Opinion in Plant Biology **16**:358-364.

Budin, J. T., W. M. Breene, and D. H. Putnam. 1995. Some compositional properties of camelina (Camelina sativa L. Crantz) seeds and oils. Journal of the American Oil Chemists' Society **72**:309-315.

Campbell, T., J. Crock, J. Williams, A. Hang, R. Sigafus, A. Schneiter, E. McClain, C. Graves, D. Woolley, and R. Kleiman. 1986. Registration of C-22, C-29, C-37 Crambee Germplasm. Crop Science **26**:1088-1089.

Carlsson, A. S. 2009. Plant oils as feedstock alternatives to petroleum–A short survey of potential oil crop platforms. Biochimie **91**:665-670.

Chapman, K. D., J. M. Dyer, and R. T. Mullen. 2012. Biogenesis and functions of lipid droplets in plants Thematic Review Series: Lipid Droplet Synthesis and Metabolism: from Yeast to Man. Journal of lipid research **53**:215-226.

Chapman, K. D. and J. B. Ohlrogge. 2012. Compartmentation of triacylglycerol accumulation in plants. Journal of Biological Chemistry **287**:2288-2294.

Cooper, J. L., B. J. Till, R. G. Laport, M. C. Darlow, J. M. Kleffner, A. Jamai, T. El-Mellouki, S. Liu, R. Ritchie, and N. Nielsen. 2008. TILLING to detect induced mutations in soybean. BMC Plant Biology **8**:9.

Downey, R. K. 1971. Agricultural and genetic potentials of cruciferous oilseed crops. Journal of the American Oil Chemists Society **48**:718-722.

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

Falasca, S. L., N. Flores, M. C. Lamas, S. M. Carballo, and A. Anschau. 2010. Crambe abyssinica: An almost unknown crop with a promissory future to produce biodiesel in Argentina. international journal of hydrogen energy **35**:5808-5812.

Fontana, F., L. Lazzeri, L. Malaguti, and S. Galletti. 1998. Agronomic characterization of some Crambe abyssinica genotypes in a locality of the Po Valley. European Journal of Agronomy **9**:117-126.

Friedt, W. and W. Lühs. 1998. Recent developments and perspectives of industrial rapeseed breeding. Lipid/Fett **100**:219-226.

Gilchrist, E. and G. Haughn. 2010. Reverse genetics techniques: Engineering loss and gain of gene function in plants. Briefings in Functional Genomics and Proteomics **9**:103-110.

Gunstone, F. and J. Harwood. 2007. Occurrence and characterization of oils and fats. The lipid handbook:37-141.

Hartwig, B., G. V. James, K. Konrad, K. Schneeberger, and F. Turck. 2012. Fast Isogenic mapping-by-sequencing of ethyl methanesulfonate-induced mutant bulks. Plant Physiology **160**:591-600.

Joubès, J., S. Raffaele, B. Bourdenx, C. Garcia, J. Laroche-Traineau, P. Moreau, F. Domergue, and R. Lessire. 2008. The VLCFA elongase gene family in Arabidopsis thaliana: phylogenetic analysis, 3D modelling and expression profiling. Plant molecular biology **67**:547-566.

Kennedy, E. 1961. Biosynthesis of complex lipids. Page 934 *in* Federation proceedings.

Klaus, M. 1999. Investigations into cultivation technique of crambe in Germany. Agro Food Industry Hi-Tech **10**:35-38.

Knights, S. 2003. Crambe: A North Dakotan case study. Pages 0-4 *in* Solutions for a better environment: Proceedings of the 11th Australian Agronomy Conference, Geelong, Victoria, Australia, 2-6 February 2003. Australian Society of Agronomy Inc., Victorian Institute for Dryland Agriculture.

Kumar, A. and S. Sharma. 2008. An evaluation of multipurpose oil seed crop for industrial uses (Jatropha curcas L.): A review. Industrial Crops and Products **28**:1-10.

Kumar, S., M. Mangal, A. Dhawan, and N. Singh. 2012. Biotechnological advances in jojoba [Simmondsia chinensis (Link) Schneider]: recent developments and prospects for further research. Plant biotechnology reports **6**:97-106.

Laghetti, G., A. R. Piergiovanni, and P. Perrino. 1995. Yield and oil quality in selected lines of Crambe abyssinica Hochst. ex R.E. Fries and C. hispanica L. grown in Italy. Industrial Crops and Products **4**:203-212.

Lalas, S., O. Gortzi, V. Athanasiadis, E. Dourtoglou, and V. Dourtoglou. 2012. Full characterisation of Crambe abyssinica Hochst. seed oil. JAOCS, Journal of the American Oil Chemists' Society **89**:2253-2258.

Lazzeri, L., F. De Mattei, F. Bucelli, and S. Palmieri. 1997. Crambe oil - A potential new hydraulic oil and quenchant. Industrial Lubrication and Tribology **49**:71-77.

Lessman, K. and V. Meier. 1972. Agronomic evaluation of crambe as a source of oil. Crop Science **12**:224-227.

Lessman, K. J. 1975. Variation in crambe, Crambe abyssinica hochst. Journal of the American Oil Chemists Society **52**:386-389.

Lessman, K. J. 1990. Crambe: A new industrial crop in limbo. J. Janick and lE. Simon (eds.), Advances in new crops. Timber Press, Portland, OR:217-222.

Li, X., E. N. van Loo, J. Gruber, J. Fan, R. Guan, M. Frentzen, S. Stymne, and L. H. Zhu. 2012. Development of ultra-high erucic acid oil in the industrial oil crop Crambe abyssinica. Plant Biotechnology Journal **10**:862-870.

Liu, Q., R. M. Siloto, R. Lehner, S. J. Stone, and R. J. Weselake. 2012. Acyl-CoA: diacylglycerol acyltransferase: molecular biology, biochemistry and biotechnology. Progress in Lipid Research **51**:350-377.
Mastebroek, H. D., S. C. Wallenburg, and L. J. M. van Soest. 1994. Variation for agronomic characteristics in crambe (Crambe abyssinica Hochst. ex Fries). Industrial Crops and Products **2**:129-136.
McCallum, C. M., L. Comai, E. A. Greene, and S. Henikoff. 2000. Targeted screening for induced mutations. Nature Biotechnology **18**:455-457.
Miwa, T. K. 1971. Jojoba oil wax esters and derived fatty acids and alcohols: gas chromatographic analyses. Journal of the American Oil Chemists' Society **48**:259-264.
Nieschlag, H. and I. Wolff. 1971. Industrial uses of high erucic oils. Journal of the American Oil Chemists' Society **48**:723-727.
Nikolau, B. J., J. B. Ohlrogge, and E. S. Wurtele. 2003. Plant biotin-containing carboxylases. Archives of Biochemistry and Biophysics **414**:211-222.
Ogunniyi, D. 2006. Castor oil: A vital industrial raw material. Bioresource Technology **97**:1086-1091.
Perry, J., A. Brachmann, T. Welham, A. Binder, M. Charpentier, M. Groth, K. Haage, K. Markmann, T. L. Wang, and M. Parniske. 2009. TILLING in Lotus japonicus identified large allelic series for symbiosis genes and revealed a bias in functionally defective ethyl methanesulfonate alleles toward glycine replacements. Plant Physiology **151**:1281-1291.
Rivarola, M., J. T. Foster, A. P. Chan, A. L. Williams, D. W. Rice, X. Liu, A. Melake-Berhan, H. H. Creasy, D. Puiu, M. J. Rosovitz, H. M. Khouri, S. M. Beckstrom-Sternberg, G. J. Allan, P. Keim, J. Ravel, and P. D. Rabinowicz. 2011. Castor Bean Organelle genome sequencing and worldwide genetic diversity analysis. PLoS ONE **6**.
Rogério, F., T. R. B. da Silva, J. I. dos Santos, and J. P. Poletine. 2013. Phosphorus fertilization influences grain yield and oil content in crambe. Industrial Crops and Products **41**:266-268.
Rudloff, E. and Y. Wang. 2011. Crambe. Pages 97-116 Wild Crop Relatives: Genomic and Breeding Resources.
Santos, J. I. D., T. R. B. da Silva, F. Rogério, R. F. Santos, and D. Secco. 2013. Yield response in crambe to potassium fertilizer. Industrial Crops and Products **43**:297-300.
Sikora, P., A. Chawade, M. Larsson, J. Olsson, and O. Olsson. 2011. Mutagenesis as a tool in plant genetics, functional genomics, and breeding. International Journal of Plant Genomics **2011**.
Slade, A. J., S. I. Fuerstenberg, D. Loeffler, M. N. Steine, and D. Facciotti. 2005. A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. Nature Biotechnology **23**:75-81.
Sperling, P., M. Linscheid, S. Stöcker, H. Mühlbach, and E. Heinz. 1993. In vivo desaturation of cis-delta 9-monounsaturated to cis-delta 9, 12-diunsaturated alkenylether glycerolipids. Journal of Biological Chemistry **268**:26935-26940.
Stephenson, P., D. Baker, T. Girin, A. Perez, S. Amoah, G. J. King, and L. Østergaard. 2010. A rich TILLING resource for studying gene function in Brassica rapa. BMC Plant Biology **10**.
Suzuki, R., R. Noguchi, T. Ota, M. Abe, K. Miyashita, and T. Kawada. 2001. Cytotoxic effect of conjugated trienoic fatty acids on mouse tumor and human monocytic leukemia cells. Lipids **36**:477-482.
Till, B. J., L. Comai, and S. Henikoff. 2007. TILLING and EcoTILLING for crop improvement. Genomics-Assisted Crop Improvement:333-349.
Till, B. J., S. H. Reynolds, C. Weil, N. Springer, C. Burtner, K. Young, E. Bowers, C. A. Codomo, L. C. Enns, A. R. Odden, E. A. Greene, L. Comai, and S. Henikoff. 2004. Discovery of induced point mutations in maize genes by TILLING. BMC Plant Biology **4**.
Till, B. J., T. Zerr, L. Comai, and S. Henikoff. 2006. A protocol for TILLING and Ecotilling in plants and animals. Nature Protocols **1**:2465-2477.
Tsai, H., T. Howell, R. Nitcher, V. Missirian, B. Watson, K. J. Ngo, M. Lieberman, J. Fass, C. Uauy, R. K. Tran, A. A. Khan, V. Filkov, T. H. Tai, J. Dubcovsky, and L. Comai. 2011. Discovery of rare mutations in populations: Tilling by sequencing. Plant Physiology **156**:1257-1268.
Tsai, H., V. Missirian, K. J. Ngo, R. K. Tran, S. R. Chan, V. Sundaresan, and L. Comai. 2013. Production of a High-Efficiency TILLING Population through Polyploidization. Plant Physiology **161**:1604-1614.
Van De Loo, F. J., P. Broun, S. Turner, and C. Somerville. 1995. An oleate 12-hydroxylase from Ricinus communis L. is a fatty acyl desaturase homolog. Proceedings of the National Academy of Sciences **92**:6743-6747.
VanEtten, C. H., M. E. Daxenbichler, J. E. Peters, I. A. Wolff, and A. N. Booth. 1965. Seed meal from Crambe abyssinica. Journal of Agricultural and Food Chemistry **13**:24-27.
Vanhercke, T., C. C. Wood, S. Stymne, S. P. Singh, and A. G. Green. 2013. Metabolic engineering of plant oils and waxes for use as industrial feedstocks. Plant Biotechnology Journal **11**:197-210.
Wallis, J. G. and J. L. Watts. 2002. Polyunsaturated fatty acid synthesis: what will they think of next? Trends in biochemical sciences **27**:467-473.
Wang, N., Y. Wang, F. Tian, G. J. King, C. Zhang, Y. Long, L. Shi, and J. Meng. 2008. A functional genomics resource for Brassica napus: Development of an EMS mutagenized population and discovery of FAE1 point mutations by TILLING. New Phytologist **180**:751-765.
Wang, Y. P., J. S. Tang, C. Q. Chu, and J. Tian. 2000. A preliminary study on the introduction and cultivation of Crambe abyssinica in China, an oil plant for industrial uses. Industrial Crops and Products **12**:47-52.
Weiss, S. B., E. P. Kennedy, and J. Y. Kiyasu. 1960. The enzymatic synthesis of triglycerides. Journal of Biological Chemistry **235**:40-44.
Wells, R., M. Trick, E. Soumpourou, L. Clissold, C. Morgan, P. Werner, C. Gibbard, M. Clarke, R. Jennaway, and I. Bancroft. 2014. The control of seed oil polyunsaturate content in the polyploid crop species Brassica napus. Molecular Breeding **33**:349-362.
Yang, J., W. Guai, and M. Yang. 2013. Biodiesel production from a novel raw material Tung oil. Pages 406-410 Advanced Materials Research.
Yong-Gang, L., A. Steg, and V. A. Hindle. 1993. Crambe meal: a review of nutrition, toxicity and effect of treatments. Animal Feed Science and Technology **41**:133-147.

Yong-Gang, L., A. Steg, B. Smits, and S. Tamminga. 1994. Crambe meal: removal of glucosinolates by heating with additives and water extraction. Animal Feed Science and Technology **48**:273-287.
Zanetti, F., T. Vamerali, S. Bona, and G. Mosca. 2010. Can we "cultivate" erucic acid in southern Europe? Italian Journal of Agronomy **1**:3-10.
Zelmer, C. D. and P. B. McVetty. 2009. Industrial Products. Page 343 *in* S. K. Gupta, editor. Biology and Breeding of Crucifers. CRC Press.
Zhu, Y., H.-g. Mang, Q. Sun, J. Qian, A. Hipps, and J. Hua. 2012. Gene discovery using mutagen-induced polymorphisms and deep sequencing: application to plant disease resistance. Genetics **192**:139-146.

Chapter 2

Isolation and characterization of the omega-6 fatty acid desaturase (*FAD2*) gene family in the allohexaploid oil seed crop *Crambe abyssinica* Hochst

Jihua Cheng[1,2], E.M.J. Salentijn[1], Bangquan Huang[2], F.A. Krens[1], A.C. Dechesne[1], R.G.F. Visser[1] and E.N. van Loo[1*]

[1] Wageningen UR Plant Breeding, P.O. Box 16, 6700 AA Wageningen, The Netherlands

[2] College of Life Science, Hubei University, P. R. China

Published in Molecular Breeding (2013) 32:517-531

Abstract

Crambe (*Crambe abyssinica* Hochst ex. R. E. Fries) is an ideal crop for industrial oil production because of a high erucic acid content (C22:1, ~ 60%) in its seed oil. The value of crambe oil can be improved by increasing C22:1 or reducing polyunsaturated fatty acids (PUFA). The gene *FAD2* plays a critical role in PUFA biosynthesis. To identify targets for breeding, *FAD2* in crambe was characterized for copy number and expression profile.

In allohexaploid crambe seven copies of *FAD2* were detected. Three genes (*CaFAD2-C1, -C2* and *-C3*) were expressed in multiple tissues including root, seedling, leaf, flower, bud and developing seeds. In developing seeds, their expression was up-regulated with *CaFAD2-C3* being expressed predominantly with a peak at 20 days after pollination. This gene thus is a promising candidate gene for determining the PUFA levels in seed oil. Four other *FAD2* genes were considered to be "pseudo-genes" as they harbour internal stop codons and were not expressed. Among six crambe varieties with consistent variation in oil composition, no nucleotide polymorphisms were found in the *CaFAD2-C1, -C2 and -C3* genes. In seeds at 30 days after pollination, statistically significant expression level polymorphisms for only one gene, *CaFAD2-C2*, was found among the varieties. However, although significantly different, the difference in expression was small and did not explain the variation in oil composition.

Given the absence of genetic variation in *CaFAD2* genes in crambe breeding lines this calls for a molecular breeding approach, whereby mutations are chemically induced to increase the genetic variation.

Keywords: Allohexaploid, *Brassicaceae*, *Crambe abyssinica*, Fatty acid desaturase, Expression variation, 454 sequencing

Introduction

Crambe (*Crambe abyssinica* Hochst ex. R. E. Fries, 2n = 6x = 90) is an oil crop belonging to the *Brassicaceae* plant family. The seed oil from crambe contains a high percentage of erucic acid (55-60%) (Mastebroek et al. 1994). Such oils have many industrial applications. For instance, erucic acid can be converted into erucamide that is used as anti-slip agent in plastics (Leppik and White 1975; Falasca et al. 2010). The high erucic acid seed oils of crambe and other *Brassicaceae* crops like rapeseed can serve as biobased feedstocks, replacing petroleum, for the production of erucic acid derived products. Crambe seeds contain 35% of oil (60% when excluding the pod) and oil yields of up to 1 t ha^{-1} can be reached (Mastebroek et al. 1994; Carlsson et al. 2007). This high yield and the fact that crambe is not able to cross with other food oil crops in nature, make crambe an ideal crop for the production of industrial oil types (Wang and Luo 1998; Zanetti et al. 2006). As industrial oil, the value of crambe oil can be improved by producing oil with higher content erucic acid (C22:1) and oleic acid (C18:1) or a lower amount of the polyunsaturated fatty acids (PUFA), linoleic acid (C18:2) and linolenic acid (C18:3).

The biosynthesis pathway leading to fatty acids in seed oil is well studied. In short, the fatty acids destined for incorporation into seed oil, mainly oleic acid (C18:1), are *de novo* synthesized in plastids. The C18:1 is transported to the cytosol and activated to C18:1-CoA, and then incorporated in the endoplasmic reticulum. In the cytoplasm, C18:1 may undergo either elongation to C22:1 by a β-ketoacyl-CoA synthase (KCS, also known as fatty acid elongase 1, FAE1) or desaturation to linoleic acid (C18:2) and subsequently to linolenic acid (C18:3) by the action of respectively the membrane bound, microsomal enzymes delta-12-fatty acid desaturase (FAD2) [omega-6 desaturase; EC 1.14.19.6] and delta-15- fatty acid desaturase (FAD3) [omega-3 desaturase; EC 1.14.19.-] (Reviewed by Lu et al. 2011). Next, the fatty acids are stored as triacylglycerol in endoplasmic reticulum derived oil bodies (Ohlrogge 1995; Slabas et al. 2001). It has been estimated that a 10% increase in the erucic fatty acid content in, for example, rapeseed would reduce the processing costs by half (Jadhav et al. 2005). Also, a reduction of the level of PUFA will result in more pure oil and reduce the extraction costs. The enzyme directing C18:1 to the PUFA fraction is FAD2. The *FAD2* gene is topic of many studies in oil crops and the model plant *Arabidopsis thaliana* which also belongs to the *Brassicaceae*. In *Arabidopsis thaliana* a single *FAD2* gene is present which is expressed throughout the plant (Okuley et al. 1994). Unlike in *Arabidopsis*, four *FAD2* genes were mapped to four different linkage groups in the tetraploid *Brassica napus* genome

(AACC), two genes per sub-genome (Scheffler et al. 1997; Schierholt et al. 2000). Of the two functional genes on the A-genome one gene has a major effect and the other gene has a minor effect on oleic acid composition in *Brassica napus*, whereas there is no evidence to support that the two *FAD2* genes on the C-genome are functional (Hu et al. 2006). In camelina, another polyploid oilseed crop in the *Brassicaceae* family, three *FAD2* copies were evident (Hutcheon et al. 2010; Kang et al. 2011). Five *FAD2* genes have been reported in soybean (*FAD2-1A*, *-1B* and *FAD2-2A*, *-2B*, *-2C*). Of these only two (*FAD2-1A* and *-1B*) are expressed specifically in seeds and have been considered to be the major loci responsible for C18:1 content of the seed oil. *FAD2-2A* was predicted to be a pseudogene because of a 100 bp deletion in the coding sequence (Heppard et al. 1996; Tang et al. 2005; Schlueter et al. 2007).

Based on the function of FAD2, desaturation of oleic acid (C18:1) to linoleic acid (C18:2), silencing of the *FAD2* gene is expected to lead to increased levels of C18:1 and because of the higher substrate concentration for FAE, maybe even increased levels of C22:1. Therefore, lots of studies were carried out with the aim to change the oil composition by influencing *FAD2* expression. It has been reported that in *Arabidopsis* C18:1 increased from 17% to 53% after silencing *FAD2* with hairpin RNA (Stoutjesdijk et al. 2002) and with the same strategy C18:1 content in cotton seeds increased from 15% to 77% (Liu et al. 2002). In case of *Brassica carinata* which contains a high percentage of C22:1 (35%) in its oil, both C18:1 and C22:1 content significantly increased when silencing *FAD2* with both antisense and co-suppression (Jadhav et al. 2005). When a zero erucic acid line of *Brassica oleracea* was employed to aim at a high C18:1 content by antisense suppression of *FAD2*, as expected only an increase in C18:1 content was observed (Sivaraman et al. 2004). For *C. abyssinica*, although silencing FAD2 genes led to increase in the C18:1 content by 10%, this gene silencing did not result in increase in C22:1(Li et al. 2012).To obtain a higher C22:1 content it is more efficient to combine genes involved in both C22:1 and C18:1 synthesis (e.g. over-expression of *FAE1* and *FAD2*-RNAi) (Sasongko and Möllers 2005; Mietkiewska et al. 2008). These studies proved that *FAD2* is using C18:1 as substrate in plants, while the relationship between *FAD2* expression and C22:1 content might depend on the flux of C18:1 to C22:1, so FAE1 activity. However, to date not much knowledge is available about metabolic fluxes in lipid metabolism (Slabas et al. 2001).

Mining valuable natural genetic variation in *FAD2* or selecting induced mutants of *FAD2* is a promising way to breed for lines with a changed oil composition. QTL mapping and whole genome scans, identified *FAD2* gene variants that were associated with a high C18:1 content.

Some of these changes were due to nucleotides substitutions within conserved nucleotide regions and affected the protein activity (Tanhuanpää et al. 1998; Beló et al. 2008). Out of the two *FAD2* genes that are expressed in soybean, a mutation in *FAD2-1B* alone was sufficient to increase the C18:1 content from 20.5% to 30.6%. The C18:1 level was elevated further to over 82% when this mutant was crossed with a mutant with a deletion in the other expressed gene, *FAD2-1A* (39.4% of C18:1) (Pham et al. 2010; Pham et al. 2011).

Breeding of the allohexaploid *Crambe abyssinica* mainly depended on a traditional approach (reviewed by Rudloff and Wang 2011) and not much is known about the *FAD2* genes of crambe. At present only fourteen *FAD2* sequences derived from *C. abyssinica* and other crambe species are available at public gene banks. Here, the *FAD2* genes in *Crambe abyssinica* cv. 'Galactica' are sequenced, mined for natural genetic variation by high throughput 454 amplicon sequencing, the copy number is determined and the expression profiles of the different *FAD2* genes are shown. Moreover, the study reveals information regarding natural differences in *FAD2* expression level during seed development in relation to oil composition in breeding materials.

Materials and methods

Plant materials

Crambe abyssinica cv. 'Galactica' was used for *FAD2* gene cloning and copy number determination. Six crambe varieties, including four breeding lines (Elst2007-03, Elst2007-04, Elst2007-15, PRI9104-71) and two cultivars 'Galactica' and 'Nebula' (**Table 1**) which performed various/differently for fatty acid composition in the field (Wageningen, 2007) were chosen to investigate the relationship between *CaFAD2* natural variation, the expression profile of *CaFAD2* genes and oil composition. All materials were grown in a glasshouse at 22°C with 16 h of daylight. Young leaves were collected for DNA isolation. Individual seeds in each variety were labelled and five to ten seeds were harvested every few days after the onset of pollination (DAP). For RNA extraction, the tissues including leaves, flowers and three developmental stages of seeds (10DAP, 20DAP, 30DAP) were harvested from all materials, and the extra tissues in cv. 'Galactica' including seedlings, roots, buds, and four extra developmental stages of seeds (7DAP, 14DAP, 21DAP, 28DAP) were harvested.

Cloning of *FAD2* genes

FAD2 genes were cloned from cv. 'Galactica' by conventional means. A degenerated primer pair for cloning of *FAD2* (**Supplementary Table S1**) was developed based on the *FAD2* sequence of *Arabidopsis* (GI: 145338403) in regions that are conserved among *FAD2* genes of *Brassicaceae* (see results). With these primers, an *FAD2* amplicon was amplified in two subsequent PCRs. The first PCR was performed in a 25 µl reaction volume containing 200 ng genomic DNA, 1×Taq buffer, 1 unit of Taq DNA polymerase (Goldstar, Eurogentec) plus 0.1 unit of *Pfu* DNA polymerase (Fermentas), 1.5 mM $MgCl_2$ (Fermentas, USA), 0.4 mM dNTPs, 0.4 mM of each of the primer (FAD2-Fw1, FAD2-Rv1). The cycling conditions for PCR were 95°C for 5 min, followed by 35 cycles at 94°C for 30 s, 55°C for 1 min, and 72°C for 2 min, with a final step of 72°C for 10 min. An aliquot of 0.5 µl of the first PCR product was used as template in a second PCR, using the same conditions. The product was purified from agarose gel, cloned into pGEM-T easy vector (Promega, USA) and sequenced.

The sequences from genomic cloning were assembled at 90% sequence homology with Seqman (Lasergene 9 software, USA), and the resulting five contigs were designated as references for the different *FAD2* genes of crambe (*CaFAD2-C1*, *CaFAD2-C2*, *CaFAD2-C3*, *CaFAD2*-A, *CaFAD2-B*). Based on these reference sequences, nested primers were developed to clone flanking sequences of the coding region, and 3-prime untranslated regions (3'-UTRs) of three different FAD2 genes from cv. 'Galactica' by combining genome walking with 3'-RACE. For genome walking, four genomic DNA libraries were prepared in an overnight digestion reaction at 37°C, each containing 500 ng of genomic DNA, 1×restriction buffer and 10 unit of restriction enzyme (Dra I, EcoR V, Sca I and Ssp I respectively, Fermentas), 50 pmol of ligation adaptor, 200 µM of ATP and 5 unit of T4 Ligase (Fermentas) in a reaction volume of 50 µl. Of each DNA library, 1 µl aliquots were used as template for a first PCR. The first PCR was performed in a reaction volume of 25 µl containing 1×Taq buffer, 0.4 unit of Hotstar DNA polymerase (Qiagen, Germany) plus 0.04 unit of *Pfu* DNA polymerase (Fermentas, USA), 0.2 µM of adaptor primer (Adpr1) and gene specific primer (Fw-GSP1, Rv-GSP1 and Rv2-GSP1 for downstream and upstream sequence cloning respectively). The PCR cycling conditions were 95°C for 15 min followed by 25 cycles at 94°C for 30 s, ramping to 50°C for 30 s with 1°C increments per second, then ramp to 72°C for 2 min with 1°C increments per second, a final step of 72°C for 10min. The product of the first PCR was 50×diluted as template for a nested PCR performed with a nested gene specific primer (Fw-GSP2, Rv-GSP2, Rv2-GSP2 and FAD2A-Rv-GSP2 respectively) and a nested adaptor primer (Adpr2) for 30 cycles under the same thermal conditions. The adaptor primer (Adpr2) was

combined with the different gene specific primers depending on the enzymes, and the combinations are listed in **Supplementary Table S1**.The fragments of the nested PCR were cloned into pGEM-T easy vector and sequenced. The 3' end including the stop codon and 3'-UTR of the different *FAD2* genes was cloned by 3'-RACE according to the manufacturer's instruction (Invitrogen, USA). The sequences of adaptor and primers for genome walking and 3'-RACE are listed in **Supplementary Table S1**.

Sequences upstream of the start codon and downstream the stop codon, harboured useful sequence variation unique to five specific *FAD2* genes, and were used to develop primer pairs (Flanking primers, in **Supplementary Table S1**) to clone *FAD2* genes from the different *C. abyssinica* varieties.

To confirm the number of *FAD2* genes in crambe cv. 'Galactica' several FAD2 amplicons were sequenced in depth with the 454 amplicon sequencing technology (Roche, Switzerland). Primers were developed on the five cloned crambe *FAD2* sequences, *CaFAD2-C1*, *CaFAD2-C2*, *CaFAD2-C3*, *CaFAD2*-A and *CaFAD2-B*. FAD2 amplicons for 454 sequencing were amplified with five primer combinations (see **Supplementary Table S1**). PCR reaction contained 100 ng of genomic DNA, 200 µM of dNTPs, 0.25 µM of each of the primers, 1 unit of Phusion Taq DNA polymerase and 1×Phusion buffer (NEB, USA) in 20 µl. An initial denaturation cycle at 98°C for 1 min was followed by 30 cycles at 98°C for 10 s, 58°C for 10 s and 72°C for 40 s, with a final step of 72°C for 5 min. The PCR products amplified with the primer combinations were mixed and then sequenced from two sides using the 454 amplicon sequencing technology at Greenomics, Wageningen UR.

All five unique full lengths *FAD2* sequences and two additional unique *FAD2* fragments are deposited in the GenBank; accession numbers: JX964741, JX964742, JX964743, JX964744, JX964745, JX964746 and JX964747.

Phylogenetic analysis

The deduced amino acid sequence of the *FAD2* sequences of *C. abyssinica* were aligned with known *FAD2* sequences from several other crops and a phylogenetic tree was constructed using the neighbour-joining method (MEGA4, USA) (Tamura et al. 2007).

Southern analysis

A probe which was specific to all detected crambe *FAD2* genes was designed on a conserved part of the gene and radioactively labelled with radioactive [^{32}P]ATP via random prime labelling (Random Prime Labelling kit, Gibco). The FAD2 probe was prepared with the primer pair FAD2-probe-Fw, FAD2-probe-Rv (**Supplementary Table S1**). Genomic DNA was isolated from leaves of *in-vitro* propagated crambe plantlets according to a method described by Aldrich and Cullis (1993) but with 1% (w/v) polyvinylpyrrolidone-10 in the DNA extraction buffer. A total of 20 to 30 µg of genomic DNA extracted from cv. 'Galactica' leaves was digested with EcoR I, Xba I and Dra I respectively, fractionated on 0.8% agarose gel and transferred to Hybond N+ membrane(Amersham Biosciences, UK) according to the manufacturer's recommendations. The membrane was hybridized at 65°C overnight with 20 ng of [^{32}P]ATP-labelled FAD2 probe in 1% SDS, 1M NaCl, 10% dextran sulphate and washed 1 time 30 minutes with 1×concentrated saline-sodium citrate (SSC) buffer, 0.1% SDS at 65 ^{o}C and 2 times 30 minutes with 0.1×SSC buffer, 0.1% SDS at 65^{o}C. The DNA gel blots were exposed to a phosphorimager screen and subsequently scanned into a Bioimager device (Fujix BAS2000).

RT-PCR

To investigate the expression profiles of the different *FAD2* genes, specific primers (**Supplementary Table S1**) were developed based on single nucleotide polymorphisms (SNPs) that can distinguish the different crambe *FAD2* sequences. Total RNA was isolated from various tissues of cv. 'Galactica' with RNeasy Plant Mini Kits (Qiagen, Germany) according to the manufacturer's instructions. The isolated RNA was treated with RNase-free TURBO DNase (Ambion, USA) to remove genomic DNA contamination. First-strand cDNA was synthesized in 20 µl from 1 µg of total RNA with iScript™ cDNA Synthesis Kit (Bio-rad, USA), in parallel 1 µg of RNA of each sample was treated in the same way but without adding reverse transcriptase, as negative controls (RT-). The cDNA and RT- samples were 20×diluted and used as templates for RT-PCR. RT-PCR was performed at 95°C for 3 min followed by 28 cycles at 95°C for 30 s, 58°C for 30 s, 72°C for 45 s, then a final step at 72°C for 5 min. The PCR reaction contained 1 µl of template, 1×Taq buffer, 1 unit of Taq DNA polymerase (Dreamtaq, Fermentas), 0.4 mM dNTPs, and 0.5 mM of each of the primers. Genomic DNA (100 ng) and DNA of plasmids (10^{-3} ng) containing the respective *FAD2* gene were involved as positive controls for RT-PCR. The 18S ribosomal RNA gene (*18S rRNA*, GI: 342673780) was used as reference gene.

Real-time quantitative PCR and analysis

The efficiency of real time PCR primers (listed in **Supplementary Table S1**) was determined with a serial dilution of plasmid DNA (10^{-1} ng, 10^{-2} ng, 10^{-3} ng, 10^{-4} ng, 10^{-5} ng) containing a specific *FAD2* gene, and efficiency calculation was performed as described previously (Hu et al. 2009). Total RNA isolation and cDNA synthesis including a negative control (RT-) were described above. The cDNA and control samples were 20-fold diluted and used as template for a PCR reaction containing 2 μl template, 5 μl SYBR Green Super Mix (Bio-rad, USA), 1 μl of each of the forward and reverse primers (3 μM). Cycling conditions were 1 cycle at 95°C for 3 min followed by 30 cycles at 95°C for 10 s, 60°C for 1 min, then a final melt step from 65°C to 95°C ramp with 0.5 °C increments per cycle to monitor specificity. For each line and each time point, three replicates were prepared from independent plants, and PCR reactions were performed in triplicate for each replicate. Means of three repeats were used for the calculation of normalized expression by reference gene using the comparative CT method (User bulletin no. 2, ABI PRISM 7700 Sequence Detection System, December 1997; Perkin-Elmer, Applied Bio-systems). Means of calculated normalized expression level in three replicates represented the gene expression level of lines. The β-actin 2 gene (GI: 20465834) was used as reference gene. One-way ANOVA (SPSS, USA) was run to test differences on the normalized expression levels among the lines at significance level $P = 0.01$.

Oil composition and statistical analysis

For field-grown materials, the fatty acid (FA) fraction was extracted from 0.5 g of ripe crambe seeds and compounds of fatty acid methyl esters (FAMEs) were analysed and identified using gas chromatograph-mass spectrometry (GC-MS, column/DB-23, Agilent). The seeds (0.5 g) were crushed in a plastic tube with screw cap and 24 ml of hexane and 3.2 ml of KOH (5M)/methanol were added, followed by vortexing and heating samples at 60°C for 6 min. The samples were then allowed to cool down to room temperature before centrifugation at 3000 rpm for 5 min, the upper layer was used for GC-MS. The extraction (1 μl) was injected into GC-MS with split ratio 1:20 and the condition of oven was 180°C for 10 min, ramp to 240°C for 7 min with 7.5°C increments per min. The mass range (*m*/z) scanned from 10 to 800 amu. Solvent delay was 4 min and gain factor was 0.5.

For glasshouse-grown materials, three plants per variety were sampled and 60 seeds per plant were collected for FA extraction. The FA was extracted in triplicates from every 20 seeds.

The procedure for oil extraction followed the extraction steps for the field-grown materials, however with adding 6 ml of hexane and 0.8 ml of KOH (5M)/methanol. The extraction (1 µl) was analysed using GC (column/DB-23)-FID (flame ionization detector).The GC conditions were the same with that of GC-MS and the FID temperature was 350°C. FAMEs compounds were identified by calibrating with a known standard. ANOVA and correlation analysis were carried out at significance level $P = 0.01$ (SPSS, USA).

Results

Cloning of *FAD2* genes from *Crambe abyssinica*

Using degenerated primers that were designed on conserved regions among *FAD2* genes of *Brassicaceae* plant species a fragment of 880 base pairs was amplified from *C. abyssinica* cv. 'Galactica' and sequenced after cloning. The obtained sequences (number of sequences N=181) clustered into five groups based on 90% sequence identity, and were designated *CaFAD2-A*, *-B*, *-C1*, *-C2* and *-C3* based on their properties (see below). *CaFAD2-C3* was only represented by a single sequence. Genome walking and 3'-RACE resulted in the full length *FAD2* coding sequences of five genes. *CaFAD2-C1*, *CaFAD2-C2* and *CaFAD2-C3* share 95% to 97% identity among each other whereas the nucleotide identity of *CaFAD2-A* and *-B* compared to *CaFAD2-C1*, *-C2* and *-C3* ranged from 89% to 91%. Furthermore, it was shown that, *CaFAD2-A* and *CaFAD2-B* both have a deletion of respectively seven base pairs and one base pair (see **Supplementary Figure S1**) and harbour an internal stop codon (**Supplementary Figure S2**). So, *CaFAD2-A* and *CaFAD2-B* encode for truncated FAD2 proteins and are regarded as "pseudogenes". The crambe *FAD2* sequences show 81% to 99% nucleotide identity to *FAD2* sequences from several *Crambe* species (*C. abyssinica*, *C. hispanica*, *C. kralikii* and *C. maritime*) deposited in GenBank database and 82% to 92% identity to other *Brassicaceae* species such as *Arabidopsis* (84% to 89%), *Brassica* (86% to 92%) and *Camelina* (82% to 88%).

The deduced amino acid sequences (383 amino acids) of the genes with full open reading frames, CaFAD2-C1, -C2 and -C3 share a high degree of identity (91% to 99%). Similar to other *FAD2* genes, CaFAD2-C1, -C2 and -C3 harbour three conserved histidine-rich motifs that are required for the binding of two iron molecules which is essential for the catalytic function of the FAD2 enzyme (Okuley et al. 1994; Shanklin et al. 1994) (**Supplementary Figure S2**).

FAD2 genes from plants are "intronless" genes in the coding sequences but instead harbour an intron of ~1.2 kb that is located upstream of the coding sequence (Liu et al. 2001; Pirtle et al. 2001; Kim et al. 2006; Kang et al. 2011). We cloned the sequences flanking the coding regions of *CaFAD2* genes. The sequences upstream of the start codon, which are assumed to be introns, were diverse (36% to 88% nucleotide identity for the regions covering all five *CaFAD2* genes). Similar to the introns of other *FAD2* genes, an intron splice site (AG) is located in four base pairs in front of the start codon of *CaFAD2* genes (**Supplementary Figure S3A**). In contrast, the 3'UTR sequences of the respective genes share higher levels of nucleotide identity that ranged from 77% to 95% (see **Supplementary Figure S3B**).

Copy number of the *FAD2* gene-family in *C. abyssinica*

Using a conserved *CaFAD2* fragment (310 base pairs) as hybridization probe varying numbers of hybridizing bands were detected depending on the restriction enzymes used. Seven hybridizing fragments were detected in Xba I-digested genomic DNA, whereas only six hybridizing fragments were detected in both Dra I- and EcoR I digests (**Fig. 1**). Since the probe is not cut by these latter two enzymes the number of hybridizing restriction fragments corresponds to the number of *FAD2* homologs present in *C. abyssinica* cv.'Galactica'.

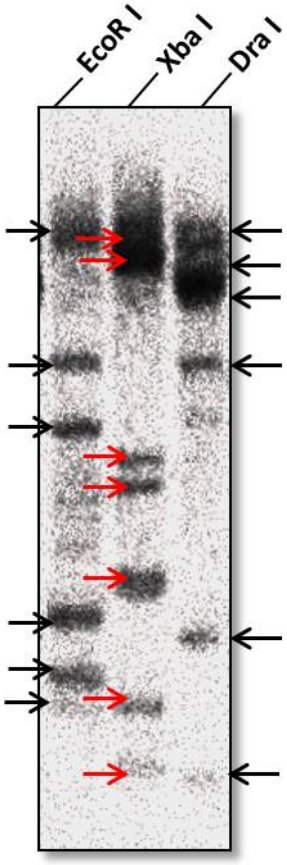


Fig. 1 Copy number determination of FAD2 in C. abyssinica by Southern blotting. Genomic DNA was digested with Dra I, EcoR I and Xba I respectively and then hybridized with [32P]ATP labelled-FAD2 probe. The arrows indicate the bands detected

Deep sequencing of the *FAD2* gene family

Thus far we determined the sequences of five *FAD2* genes, whereas six to seven hybridizing restriction fragments were observed in *C. abyssinica* cv. 'Galactica'. To explore the genetic variation in the *FAD2* gene family of *C. abyssinica* in-depth, and to determine sequences of all six to seven genes, 454-amplicon sequencing was performed. For this, five primer combinations were used to amplify five different *FAD2* amplicons and the locations of amplicons in the *CaFAD2* gene were shown in **Supplementary Figure S5**. The five amplicons were sequenced by 454 amplicon sequencing, and over 90,000 sequences were generated and analysed. All sequences were assembled with the five known *CaFAD2* sequences (*CaFAD2-C1*, *-C2*, *-C3*, *-A*, *-B*). Numerous sequences representing all five *CaFAD2* genes were amplified by amplicon 1- and amplicon 3- primers. The primers for amplicon-2 were designed to amplify *CaFAD2-C1*, *-C2* and *-C3*, but still 15% of the sequence reads were from *CaFAD2-A* gene (**Supplementary Table S2**). The primers for amplicon-4 and amplicon-5 favoured the amplification of respectively *CaFAD2-C1* and *-C2*, and *CaFAD2-C2* and *-C3* (**Supplementary Table S2**). Moreover, 5% of the sequences of amplicon-3 represented a novel *FAD2* sequence (*CaFAD2-E*) whereas in both, amplicon-2 and amplicon-5, two other novel *FAD2* sequences were found (respectively *CaFAD2-D* and *CaFAD2-F*) (**Supplementary Table S2**). All sequences were aligned and it was found that *CaFAD2-E* and *CaFAD2-F* shared an overlapping region of 258 base pairs, indicating that most likely; they represent a single variant (see **Supplementary Figure S5** and **S6**). *CaFAD2-D* and *CaFAD2-E/F* harbour an internal stop codon and encode for truncated FAD2 proteins (**Supplementary Figure S5** and **S6**). So, after deep 454-amplicon sequencing it can be concluded that the *FAD2* gene family of *C. abyssinica* cv. 'Galactica' consists of seven family members, *CaFAD2-A*, *-B, -C1*, *-C2*, *-C3*, , *-D* and *-E/F* and that only the *CaFAD2-C1*, *-C2* and *-C3* genes are encoding the full *FAD2* open reading frame.

A Neighbour-Joining (NJ)-analysis, using a short sequence of 142 base pairs that covered all seven *CaFAD2* genes, showed that the *CaFAD2* genes formed two groups. *CaFAD2-D*, grouped together with *CaFAD2-C3*, *-C2*, *-C1* whereas the other novel *FAD2* fragment *CaFAD2-E/F* show a higher homology to *CaFAD2-A* and *-B* (**Supplementary Figure S7**). The functional *CaFAD2* sequences that encode full-length open reading frames (*CaFAD2-C1*,*-C2* and *-C3*) were aligned to known full-length *FAD2* sequences from other seed oil crops, and a phylogenetic NJ-tree was constructed at the amino acids level. The resulting NJ-tree showed that the functional *CaFAD2* sequences derived from *C. abyssinica* cv. 'Galactica'

grouped together with the *FAD2* sequences from other *Brassicaceae* and are most closely related to *FAD2* sequences from *B. napus, B. juncea, B. rapa* and *B. carinata* (**Supplementary Figure S8**).

Expression profiling of *CaFAD2* genes

Three *FAD2* genes in *C. abyssinica*, *CaFAD2-C1*, *-C2* and *-C3*, are encoding the full *FAD2* open reading frame whereas the other genes encode for truncated versions of FAD2. Nothing is known about the expression profile of the different genes in crambe and most importantly, which of the genes are expressed in correlation with the biosynthesis of seed oil during seed development. Therefore, the expression profiles of the *CaFAD2* genes were studied in cv. 'Galactica', in vegetative tissues (seedlings, leaves and roots) and reproductive tissues (buds, flowers and four developmental stages of seeds).

SNPs in the sequences allowed the design of primers that are specific for the respective *FAD2* genes, *CaFAD2-C1*, *-C2*, *-C3*, *-A* and *-B* (**Supplementary Figure S9**). RT-PCR amplification with these primers showed that *CaFAD2-C1*, *-C2* and *-C3*, that are encoding the full *FAD2* open reading frame, are expressed in all tissues tested whereas *CaFAD2-A* and *-B*, that harbour internal stop codons, are probably not functional throughout the plants (**Supplementary Figure S9**).

Next, using Real-time qPCR, the relative expression levels of three functional *CaFAD2* genes (*-C1*, *-C2* and *-C3*) were quantified in leaves (**Fig. 2A**) and during seed development (flowers, 10 DAP, 20 DAP and 30 DAP) (**Fig. 2B**). For this, the expression levels of the three *CaFAD2* genes were normalized to the expression of a reference gene (*β-actin2*), which has been found stably expressed in *C. abyssinica* under arsenate stress and constitutively expressed in *B. napus* cultivars (Hu et al. 2009; Paulose et al. 2010).

In leaves, the *FAD2* genes were expressed at levels ranging from 2% to 25% of the reference gene expression (**Fig. 2A**). In leaves *CaFAD2-C2* showed the highest expression which was respectively 3 times and 15 times higher than the expression of *CaFAD2-C3* and *CaFAD2-C1* (**Fig. 2A**).

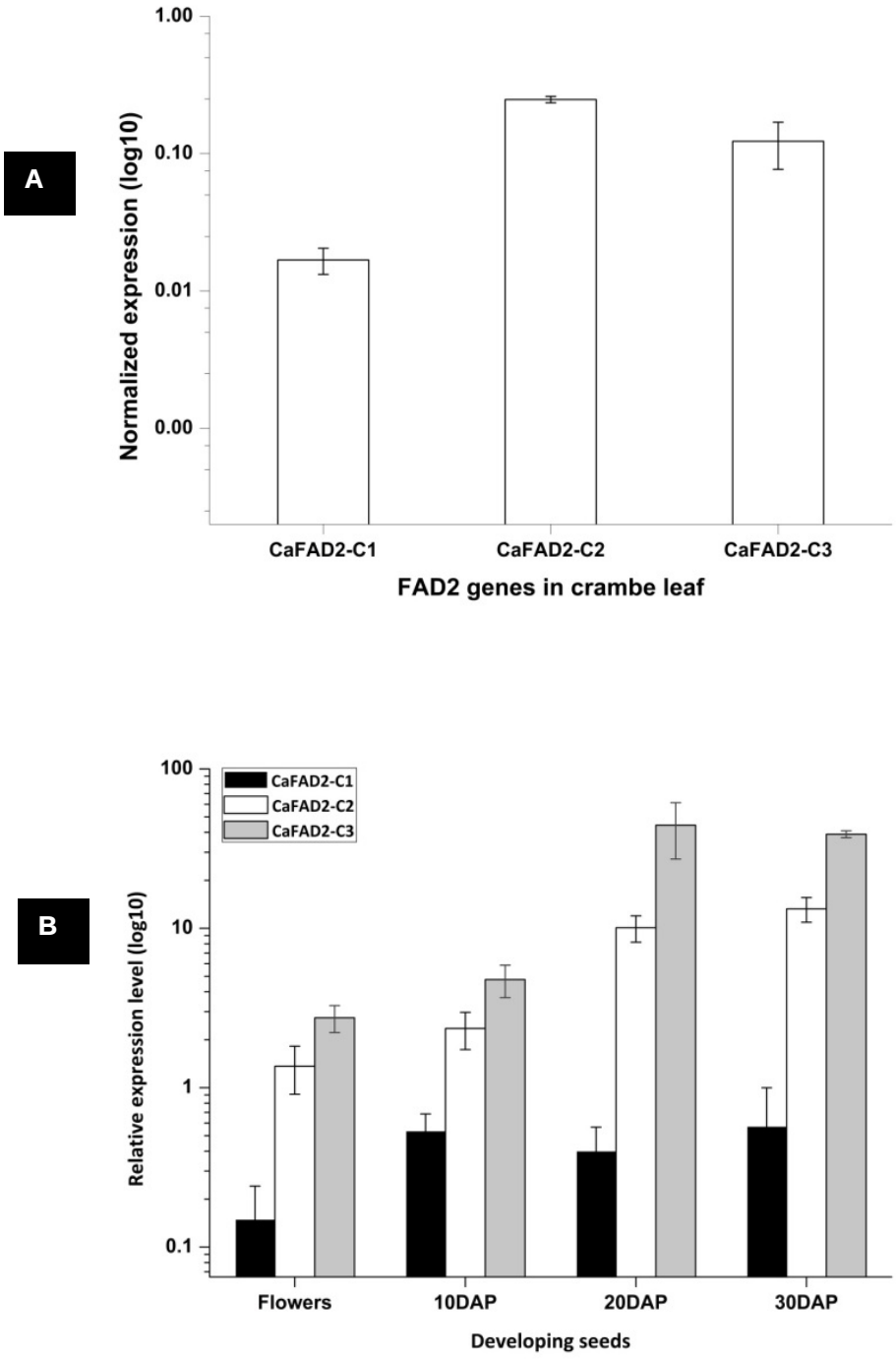


Fig. 2 Expression level of three *CaFAD2* genes in crambe cv. 'Galactica'.
(A) Real-time qPCR expression level of *CaFAD2-C1 to -C3* in leaf tissue. The expression was normalized with reference gene (*β-actin2*, GI: 20465834). (B) Relative expression level of *CaFAD2-C1 to -C3* in flowers and developing seeds. The data are normalized to *β*-actin2 gene expression and for comparison calculated relative to the expression of *CaFAD2-C2* in leaf tissue (*CaFAD2-C2* in leaves, relative to *β-actin2*= 0.25). DAP=Days after pollination

The expression of the three *CaFAD2* genes was compared in flowers and developing seeds tissues (using *CaFAD2-C2* expression in leaves as reference for comparison, **Fig. 2B**). It showed that in all seed developmental stages *CaFAD2-C3* is the most actively expressed gene with an expression that is 3 to 45 times higher compared to the reference (*CaFAD2-C2* in leaves) and with a peak of expression at 20 DAP (**Fig. 2B**). *CaFAD2-C2* shows also an ascending level of expression during seed development, ranging from 1.5 to 15 times its expression in leaves, with an expression peak at 30 DAP. Of all three genes, *CaFAD2-C1* shows the lowest expression level in developing seeds (0.25 to 1 time lower than the expression of *CraFAD-C2* in leaves) (**Fig. 2B**). Overall, the expression of *CaFAD2* genes is up-regulated from low to high level during seed development. *CaFAD2-C3* expression rose to the peak at 20 DAP when its expression was ~ 100-fold higher than that of *CaFAD2-C1* and ~ 4-fold higher than that of *CaFAD2-C2* at 20 DAP, and then declined slightly to an expression that was ~ 70-fold higher than that of *CaFAD2-C1* and ~ 3-fold higher than that of *CaFAD2-C2* at 30 DAP (**Fig. 2B**).

Characterization of *CaFAD2* in six varieties of *C. abyssinica*.

Classical breeding has yielded various breeding-lines and cultivars of *C. abyssinica*. The composition and yield of the seed oil are important criteria for selection. Perhaps selection for oil characteristics imposed differences in the *CaFAD2* expression profile and/or caused the selection of specific DNA polymorphisms in the *CaFAD2* genes. To study this, we sequenced the homologs of the functional *CaFAD2* genes (*C1*, *-C2* and *-C3*) and studied the expression of these *CaFAD2* genes in six crambe varieties, including four selected breeding lines (Elst2007-03; Elst2007-04; Elst2007-15; PRI9104-71) and two cultivars, 'Nebula' and 'Galactica'. These crambe varieties were chosen because the C18:1 content in the seed oil differed significantly, when grown in the field (**Table 1**).

CaFAD2-C1, *-C2* and *-C3* fragments from each of the crambe varieties were amplified and sequenced directly with Sanger sequencing. The alignment of the coding sequences show that the sequences obtained from the four breeding lines and cv. 'Nebula' are 100% identical to *CaFAD2-C1*, *-C2* and *-C3* from cv. 'Galactica' (data not shown). This indicates that the genetic variation in these breeding materials is limited, especially regarding the *CaFAD2* genes.

Table 1. Fatty acid composition in the seed oil of six crambe varieties grown in the field (Wageningen, 2007) **and glasshouse (**2010).

Lines		Field-grown materials								Glasshouse-grown materials					
		C18:1[a] (%)	C18:2[a] (%)	C18:3 (%)	C22:1[a] (%)	PUFA[a] (%)	Others[a] (%)	Oil (%)	Seed Yield (Kg/ha)	C18:1(%)	C18:2 (%)	C18:3 (%)	C22:1 (%)	PUFA (%)	Others (%)
Elst2007-03		18.53	9.11	6.77	60.06	15.9	5.5	42.31	3410	14.41	7.17	6.42	61.80	13.59	10.19
Elst2007-04		18.32	9.19	6.87	60.17	16.1	5.5	42.22	3535	14.42	7.34	6.27	61.68	13.60	10.27
Elst2007-15		19.55	9.68	5.55	58.20	15.2	7.0	41.31	3652	16.64	8.14	5.16	60.12	13.26	9.96
PRI9104-71		18.50	9.03	6.85	60.21	15.9	5.4	40.75	3761	14.31	7.77	6.02	61.49	13.79	10.39
'Nebula'		19.25	9.67	5.56	59.17	15.2	6.4	40.39	3448	15.99	7.92	4.80	60.13	12.72	10.85
'Galactica'		18.72	9.64	5.84	60.31	15.5	5.5	39.03	3787	15.30	7.58	5.46	61.57	13.47	9.64
	LSD	0.27	0.21	0.28	0.48	/	/	1.90	337	0.55	0.71	0.21	1.19	0.68	0.71
	***p* value**	< 0.001	< 0.001	< 0.001	< 0.001	/	/	0.002	/	< 0.001	0.088	< 0.001	0.037	0.057	0.051

In the field, each variety was planted in two blocks. For each block, the oil fraction of 0.5 g of seeds was measured in duplicate. For the glasshouse grown materials, three individual plants were analysed per variety. Twenty seeds were measured for each individual plant in triplicate. Data represent the average of three independent plants per variety. LSD=least significant difference. The *p* values indicate difference among the varieties. "/"=not determined.

[a]: Significant difference ($P = 0.01$, by t-test) between the field-grown and glasshouse-grown materials.

To compare the expression levels of *FAD2* genes among these six varieties, the expression of three active *CaFAD2* genes was normalized to the expression of a reference gene (*β-actin2*) for all varieties. It was shown that the normalized expression profiles of *CaFAD2-C1, -C2* and *-C3* are similar in the developing seeds of all varieties (**Fig. 3A** to **3C**). For the expression levels, the expression of *CaFAD2-C1*, *-C2* and *-C3* showed no significant differences during seed development at significance level $P = 0.01$, and only the expression level of *CaFAD2-C2* showed significant difference in the seeds at 30 DAP ($P = 0.012$) (**Fig. 3A** to **3C**).

GC and correlation analysis

The oils of ripe seeds of all selected *C. abyssinica* varieties, grown in the glasshouse, were profiled for fatty acid compositions using gas chromatography (GC). The content of oleic acid (C18:1), linoleic acid (C18:2) and linolenic acid (C18:3) were quantified (**Table 1**). Among the crambe varieties, minor but statistically significant variations in the content of both C18:1 and C18:3 were found (respectively, $F(5, 12) = 29.626$, $P < 0.01$ and $F(5, 12) = 87.396$, $P < 0.01$). 'Nebula' and Elst2007-15 show the lowest C18:3 content which is significantly lower than Elst2007-03, Elst2007-04 and PRI9104-71 ($P < 0.01$); furthermore, of all varieties, Elst2007-15 shows the highest content of C18:1 ($P < 0.01$) (**Table 1).** Additionally, the C18:3 content in the oil of cv. 'Galactica' is slightly higher (+1%, $P < 0.01$) than in cv. 'Nebula' but for C18:1 no significant differences are observed between cv. 'Galactica' and 'Nebula' (**Table 1**). A Pearson correlation was run to determine the relationships between *CaFAD2-C2* gene expression in the seeds at 30 DAP and the content of C18:1 and C18:3 in ripe seeds. The results show that for the six tested crambe varieties grown under glasshouse conditions, there is no significant correlation between the expression of *CaFAD2-C2* gene and the content of C18:1 and C18:3 in the seed oil.

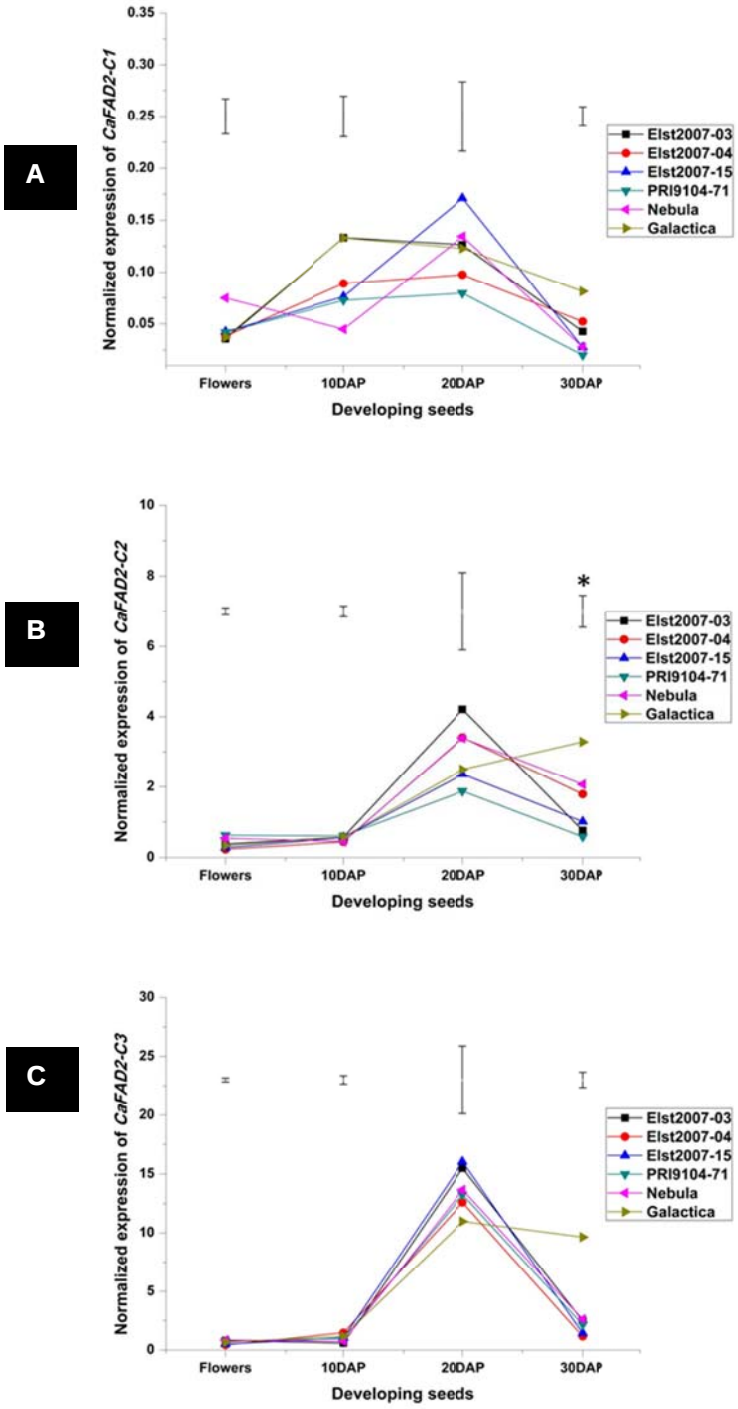


Fig. 3 Expression level of CaFAD2 genes in developing seeds of six crambe varieties.

Normalized Real-time qPCR expression of CaFAD2-C1 (A), CaFAD2-C2 (B) and CaFAD2-C3 (C) in developing seeds of six crambe varieties in flowers and at 10-, 20- and 30 DAP. The bars correspond to Fisher's LSD values (P = 0.01) for comparison of the expression level of genes among the varieties at one point. The symbol "*" indicates that expression level is significantly different (P < 0.05) at the time point among the six varieties

Comparison and relationship of oil composition in the field-grown and glasshouse-grown crambe varieties

Under the field-conditions there was much more significant variation among the varieties for oil composition. As shown in **Table 1**, the contents of C18:1, C18:2, C18:3 and C22:1, are significantly different among these varieties when grown in the field. In contrast, the glasshouse-grown materials only differed in C18:1 and C18:3 contents. In addition, significant and positive correlations were found in these varieties between the two environments for the contents of C18:1 ($r = 0.972$, $P < 0.01$), C18:3 ($r = 0.940$, $P < 0.01$) and C22:1 ($r = 0.910$, $P < 0.05$), and such correlations suggest genetic stability of seed oil composition in these varieties. Compared to the glasshouse-grown materials, the field-grown materials perform differently in the major compounds with exception of C18:3 (**Table 1**).

Discussion

Molecular characterization of the *FAD2* gene-family

Initially, we cloned five *FAD2* genes from crambe whereas seven copies were detected by Southern analysis. Usually, upstream or downstream sequences of *FAD2* coding regions diverge more than coding regions and are useful to design homolog-specific probes to determine the copy number of single *FAD2* homolog (Martínez-Rivas et al. 2001). However, the obtained flanking sequences of three *FAD2* coding regions (*CaFAD2-C1* to *-C3*) in our study were 71% to 88% homologous, and not suited to design homolog-specific probes (**Supplementary Figure S3**). The introduction of next generation sequencing technologies provides means to acquire knowledge on sequence variation within a gene family and to determine the exact copy number of genes (Varshney et al. 2009). This approach is commonly applied in simply organized genomes, but more difficult to use in polyploidy genomes (Griffin et al. 2011).

Aiming to mine *FAD2* sequence variation in crambe, we sequenced five different *FAD2* amplicons in depth with 454 sequencing. In total, seven *FAD2* variants were detected in *C. abyssinica* cv. 'Galactica'. Depending on primer combinations, different *FAD2* variants were detected; moreover, there was no primer combination which detected all variants. This result shows that the specificity of primers for certain *FAD2* sequences due to polymorphisms in the primer recognition sites is a limiting factor, hence the use of more (degenerated) primer combinations was needed to eliminate the bias towards certain sequence variants. Taken

together with the same number of *FAD2* copies detected by Southern analysis, it was concluded that seven different *FAD2* genes are present in the homozygous, allo-hexaploid genome of *C. abyssinica* cv. 'Galactica'. The deep-sequencing for crambe *FAD2* genes might be also helpful for establishing a hypothesis on how these seven *FAD2* genes are organized in the hexaploid crambe genome. In amplicon 1, the proportions of sequence reads of five crambe *FAD2* genes (*CaFAD2-A*, *-B*, *-C1*, *-C2* and *-C3*) obtained from 50 lines did not differ substantially (**Supplementary Table S2**), which suggests that these five genes might be five loci because otherwise segregation should occur in the highly inbred cultivar 'Galactica'. In amplicon 2 (for which the primers aimed at PCR-ing the group of *CaFAD2-C1*, *-C2* and *-C3*), an additional gene *CaFAD2-D* was found which was so highly homologous to *CaFAD2-C3* that *CaFAD2-D* gene might be a rare allele of the C3-locus, as it was also represented with a much lower proportion in the sequence reads of amplicon 2. Similarly, *CaFAD2-E/F* might be a rare allele of the CaFAD2-A locus (**Supplementary Table S2** and **Figure S7**).

In the case of allotetraploid *B. napus* (AACC) it was well studied that the active *FAD2* genes are located in two homoeologous regions with four sequence variants across sub-genomes (HSVs) (Scheffler et al. 1997). To prove how the seven *FAD2* genes are located and organized in the allo-hexaploid crambe genome, whole genome sequencing or genetic mapping is required.

FAD2 expression profile

The expression pattern of *FAD2* family of plant species in previous reports can be roughly classified into three classes: i) a single *FAD2* gene in a diploid genome expressed throughout the plant, i.e. *FAD2* in *Arabidopsis* and rice (Okuley et al. 1994; Shi et al. 2012); ii) multiple *FAD2* genes in a polyploid genome all of which are expressed, i.e. *Camelina sativa* (Hutcheon et al. 2010); iii) multiple *FAD2* genes in a polyploid genome of which some genes are expressed constitutively and the others are not expressed or expressed in specific tissues, i.e. *B. napus*, sunflower and soybean (Martínez-Rivas et al. 2001; Hu et al. 2006; Pham et al. 2011).

The case of crambe matches most to situation iii): of the seven *FAD2* genes of crambe, four were proven non-functional because they are not expressed and harboured premature stop codons. Only three highly homologous (95% to 97%) *FAD2* genes (*CaFAD2-C1*,*-C2* and *-C3*) are expressed constitutively in all tested tissues. Although *CaFAD2-C3* peaked during seed development, the expression of none of the genes was confined to a specific tissue (**Fig. 2** and

Supplementary Figure S9). The expression of *CaFAD2-C1*, *-C2* and *-C3* shows a similar pattern of regulation in developing seeds, and this pattern coincides with the course of accumulation of C18 fatty acids. During seed development, C18 fatty acids are mainly accumulated within 10 to 20 days after pollination (Gurr et al. 1972), when expression levels of all three *FAD2* genes in crambe increase dramatically and reach their peak. The expression of one particular gene, *CaFAD2-C3*, was significantly up-regulated (2 to 110 times higher) during seed development, unlike in camelina where the three *FAD2* genes have similar expression levels (Hutcheon et al. 2010). This might suggest that *CaFAD2-C3* plays a critical role in oil biosynthesis in crambe seeds. The fact that two genes (*CaFAD2-A* and *-B*) were not functional explained why no 3'UTR sequence of them was cloned with 3'-RACE. Similar to *CaFAD2-A* and *-B*, no 3'UTR sequence had been cloned from *CaFAD2-D* and *CaFAD2-E/F*, this supported the idea that also they are not functional. Several studies have shown unequal transcription from the homoeologous loci i.e. in hexaploid wheat (Salentijn et al. 2009) and tetraploid *B. napus* (Chen et al. 2011). Polyploidization and the modifications that follow through time may affect gene expression in allopolyploids. Also, epigenetic factors, such as methylation, small RNAs, chromatin remodelling, and alternative splicing, all may have an impact on gene expression in allopolyploids (reviewed in Soltis et al. 2010; Buggs et al. 2012).

Genetic variation in breeding lines

Both, sequence variations and expression level polymorphisms (ELPs) have been shown to contribute to phenotypic variation (Van Poecke et al. 2007). We chose six breeding varieties of crambe which differed in fatty acids composition in the field, and hypothesized that the variation in sequences and ELPs of *FAD2* affected the fatty acids composition when grown in the glasshouse. Whereas in the oil from field grown materials the differences in the major compounds (C18:1, C18:2, C18:3, C22:1) were observed among the crambe lines, under controlled glasshouse conditions only differences in C18:1 and C18:3 were observed among the lines. It is known that *FAD2* genes play a role in resistance to cold and other abiotic stresses in different plant species (i.e. Makarenko et al. 2011; Yuan et al. 2012; Shi et al. 2012). Oliva et al. (2006) showed that the contents of oleic acid, linoleic acid and linolenic acid in soybean oil were strongly correlated with average temperature during the final 30 days of the reproductive period. Even though lower temperature generally causes higher unsaturated fatty acid content in all genotypes, the magnitude of the temperature effect varied widely among genotypes. While all crambe lines experienced similar external conditions in the field, fluctuations in temperature (°C to °C) may have augmented the variation in oil

composition among the different lines. Among the breeding lines no sequence variation was found for *CaFAD2-C1* to *-C3* and significant difference on expression level of *CaFAD2-C2* gene in seeds (30 DAP) was observed with little to no effect (ΔC18:2 < 1%) on the fatty acid composition of the oil. Taken together with the fact that the proportion of C18 fatty acids in crambe oil declined from 12 DAP onwards (Gurr et al. 1972), the expression variation in 30 DAP seeds would not extremely affect C18 fatty acids composition. To answer the question if ELPs in the FAD2 genes would influence the oil composition, studies are required to be carried out for crambe lines, in which a higher level of *FAD2* expression variation takes place on earlier stages of seed development together with higher and consistent levels of variation in oil composition.

The low genetic sequence variation among the breeding varieties and the flexibility of oil composition in response to external factors raises the question whether the breeding of crambe lines with large, stable, stress-independent reduction of PUFA in oil composition is feasible by conventional breeding. The purpose of mutation breeding is to introduce SNPs into genomes. For instance, in wheat, another hexaploid crop, it was possible to manipulate phenotype using reverse genetics, i.e. targeted induced local lesion in genome (TILLING) (Slade et al. 2005). Since high throughput sequencing is widely applied in TILLING, one challenge is to discriminate mutations and variations. Here we deeply mined *FAD2* variants of crambe, identified the *FAD2* genes expressed in crambe seeds and paved the way for identifying mutations in this gene.

Supplementary information

Table S1. Primers list in this study

Category	Primers combination	Primers sequences (5'-3')	Size (bp)	Remark
FAD2 degenerated primers	FAD2-Fw1; FAD2-Rv1	AA(A/C)GC(A/T)(G/C)AATCCCT(A/C)GCTCTTT; C(A/G)TACCACGGTGTTCCATC		
FAD2 probe	FAD2-probe-Fw; FAD2-probe-Rv	CGCCACCATTCCAACACTG; CGAAGCAGCGTAACGGTAGA	310	
Genome walking	Adaptor	ACTCGATTCTCAACCCGAAAGTATAGATCCCA		
	Adpr1; Rv-GSP1	GTTTACTCGATTCTCAACCCGAAAG; AGGAGAAGTAAGGGACGAGGAGGAA		Enzyme: EcoR V;
	Adpr2; Rv-GSP2	CAACCCGAAAGTATAGATCCCA; AGTAAGAGAGAGGGTGAGGGAGGAG		Enzyme: EcoR V;
	Adpr1; Rv2-GSP1	GTTTACTCGATTCTCAACCCGAAAG; GACCATAACAAACGGCGAGGATACCAG		Enzyme: Dra I; EcoR V
	Adpr2; Rv2-GSP2	CAACCCGAAAGTATAGATCCCA; GTAGATGGGAGCGTTAGGGTGGAAATG		Enzyme: EcoR V
	Adpr2; FAD2A-Rv-GSP2	CAACCCGAAAGTATAGATCCCA; CGTGGTGGCCACATTCATGCAGAGG		Enzyme: Dra I
	Adpr1; Fw-GSP1	GTTTACTCGATTCTCAACCCGAAAG; CTACGATTCATCTGAGTGGGATTGGTT		Enzyme: Dra I; EcoR V; Ssp I
	Adpr2; Fw-GSP2	CAACCCGAAAGTATAGATCCCA; TGGCTACTGTTGACAGAGACTATGG		Enzyme: Dra I; EcoR V; Ssp I
3'-RACE	3'Primer; Fw2-GSP1	GCTGTCAACGATACGCTACGTAACG; TTCCGCTTCTGATAGTCAACGGGTT		
	3'Nested Primer; Fw-GSP2_C1-2	CGCTACGTAACGGCATGACAGTG; CACGCATCCTTCATTGCCTCACTAC	700	
	3'Nested Primer; Fw2-GSP2_C3	CGCTACGTAACGGCATGACAGTG; TTCATCTGAGTGGGATTGGTTAACG	700	
Flanking primers	upFAD2-A; dnFAD2-A	GTCTCCACCTTTTGACTCTT; ATAACTCTGTTCTTTCACCATC	1254	
	upFAD2-B; dnFAD2-B	TCATATCCACATGCCCCTAT; ACCAGTGGCTTTGACGA	1572	
	upFAD2-C1; dnFAD2-C1	GCTATCGTTTATTTATTTTTCTTTC; AGCGATGAGAAGAACAATACAGAGA	1572	
	upFAD2-C2; dnFAD2-C2	CTCAACGCTATCGTTTATTTCTTTC; AGATAAAAGTAGCTTCACAGG	1660	
	upFAD2-C3; dnFAD2-C3	CCACGTACTATCCATTTTTGAAAGT; GGCACAACACAATGGATACTT	1393	
RT-PCR	G1F3; G1R4	TCCAACACTGGCTCCCTCGAG; AAGTAAGTGATCAAAACGGGA	383	CaFAD2-A specific
	G2F3; G2R5	TCCTGCTTCAACTACATCT; TGCGTTATAATGCGGCATT	800	CaFAD2-B specific
	C1F3; C1R3	CAAGGCTGTGTCCTAACA; AACACTTCGTCTCTTTCAACC	200	CaFAD2-C1 specific
	C2F3; C2R4	TGCCAAGGCTGTGTCCTAGGT; GCTAAGTACAAAGGCCAGCAA	310	CaFAD2-C2 specific
	C3F2; C3R2	CAACACTGGATCATTGGAA; GTAATGAGGCAATGAAGGATGA	411	CaFAD2-C3 specific
	18S F2; 18S R2	CGTGCCCTTTACGCCCAGTCATT; AGGTAGCTTGTCTCGCCCAGGAG	360	
Real time PCR	qRT-C1F5; qRT-C1R5	CAAGGCTGTGTCCTAAGA; CGTCGTCTCTTTCAAGCGGA	200	CaFAD2-C1 specific, Efficiency: 1.003
	qRT-C2F2; qRT-CF2	CGCCATTCCAACACTGGCTCC; GCGAAGCCGTCGTAAGGTCT	191	CaFAD2-C2 specific, Efficiency: 1.001
	qRT-C3F7; qRT-C3F7	CAAGGCTGTGTCCTAAGC; AACACTTCATCTCTTTCCAAT	200	CaFAD2-C3 specific, Efficiency: 0.923
	Actin2-F1; Actin2-R1	ATTCAGATGCCCAGAAGT; CTCATACGGTCAGCGATA	180	Efficiency: 0.823
454 amplicon	454-AM1-Fw; 454-AM1-Rv	CTCCTCCCTCACCCTCTCTC; ACGTGCGTGTCCGTGATGT	724	Amplicon1
	454-AM2-Fw; 454-AM2-Rv	TTCTACTACATCGCCACCAAC; ATTGGTAGTCGCTGAAAGCAT	153	Amplicon2
	454-AM3-Fw; 454-AM3-Rv	TCACCCTCTCTCTTACTTAGCTT; AGACGTTCACGGTCGTTG	437	Amplicon3
	454-AM4-Fw; 454-AM4-Rv	AGAAGAAATCAGACATCAAGTGGTAC; GAAGACCTTGTTCAACATCCCATA	448	Amplicon4
	454-AM5-Fw; 454-AM5-Rv	GCTCAATTCCTCGCTCTTTC; GTCCAAGAGGGTTGTTGAGG	392	Amplicon5

Table S2. Deep 454-amplicon sequencing of the *FAD2* gene family from *C. abyssinica* cv. 'Galactica'. Five different amplicons were analysed for genetic variants of the *FAD2* gene.

Gene	Amplicon1	Amplicon2	Amplicon3	Amplicon4	Amplicon5	
CaFAD2-A	20%	/	15%	1%	15%	
CaFAD2-B	25%	1%	16%	0%	1%	
CaFAD2-C1	20%	30%	22%	57%	0%	
CaFAD2-C2	22%	32%	23%	40%	40%	
CaFAD2-C3	13%	27%	19%	1%	33%	
CaFAD2-D	/	11%	/	/	/	
CaFAD2-E/F	/	/	5%	/	11%	
Nseq total	43267	12883	15395	13562	5577	90684
Genotypes	> 50 lines	5 lines	5 lines	5 lines	5 lines	
Per line	865	2577	3079	2712	1115	

Nseq total= total number of the sequence. %= frequency of a *FAD2* variant in *FAD2*-gene family among the sequences of an amplicon. The symbol "/" =no sequences matched to certain *FAD2* variant.

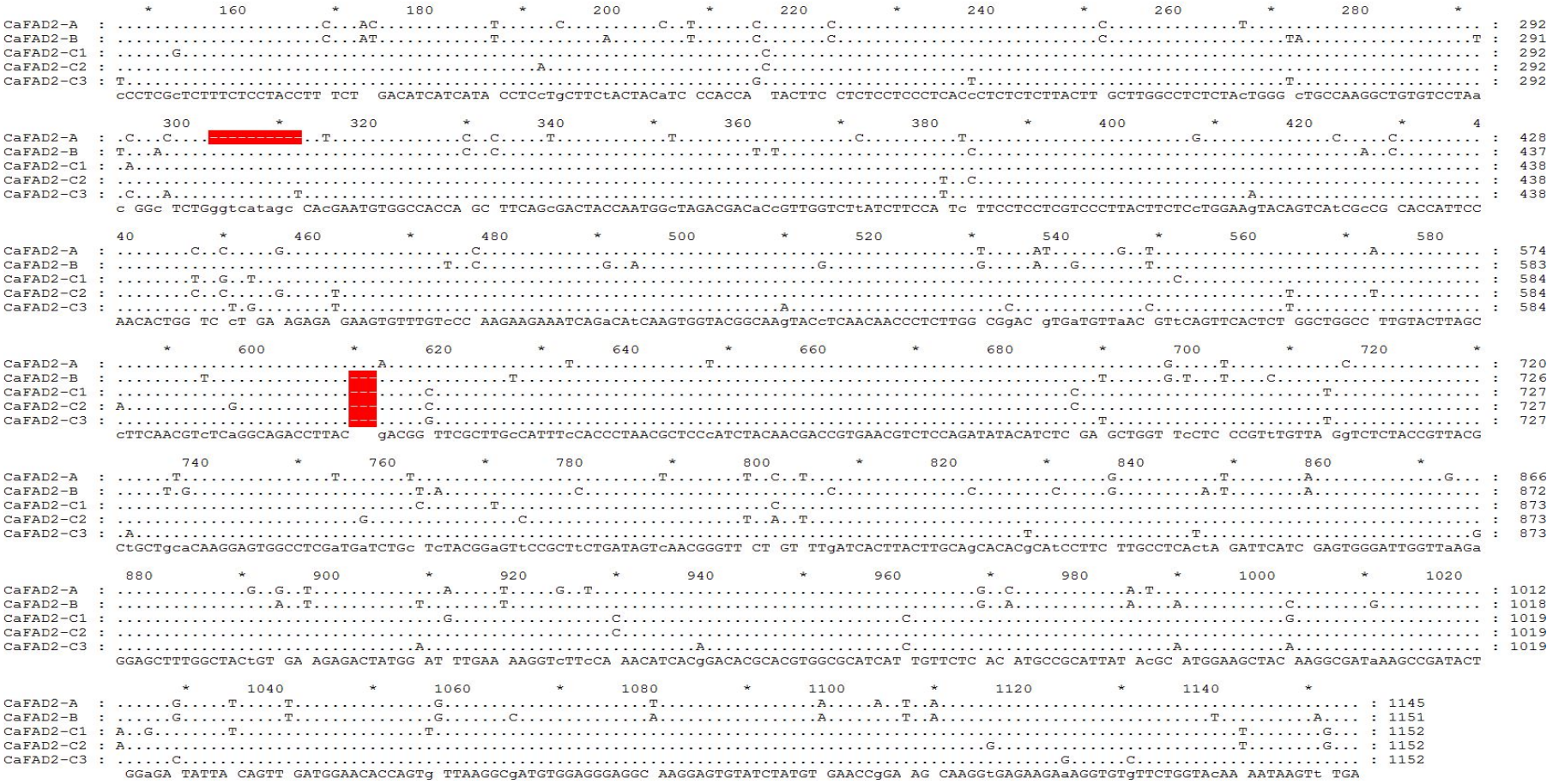


Figure S1. The nucleotide alignment of full length Crambe *FAD2* (*CaFAD2-A, -B, -C1, -C2* and *-C3*). The dots indicate identical nucleotides, and the deletions are highlighted. The sizes of *CaFAD2-A* and *-B* are seven base pairs and one base pair shorter than the others (1152 bp) respectively. *CaFAD2-C1*, *-C2* and *-C3* share 95%-97% identity to each other. *CaFAD2-A* and *-B* share 90% identity to each other and 88%-90% to the others.

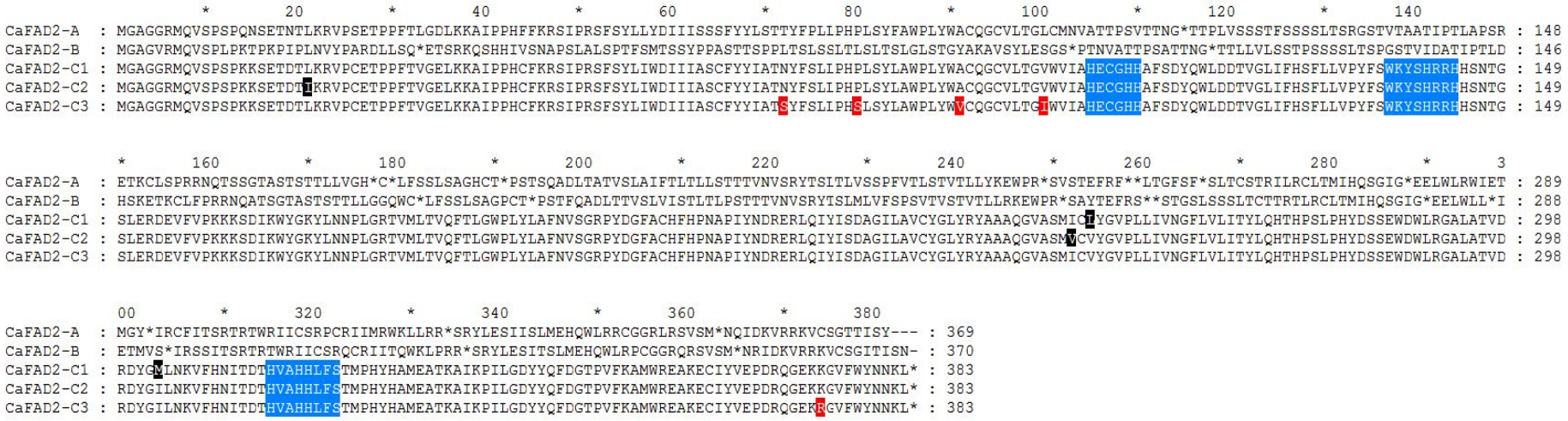


Figure S2. Deduced amino acids alignment of the FAD2 genes in *C. abyssinica* cv. 'Galactica'. CaFAD2-A and -B encode for truncated FAD2 proteins because of harbouring internal stop codons (indicated by *). CaFAD2-C1, -C2 and -C3 encode for functional FAD2 proteins which share 98% to 99% identity. CaFAD2-C1 differs from CaFAD2-C2 in four residues (highlighted in black colour), and five residues (highlighted in red colour) are specific to CaFAD2-C3. The three conserved histidine-rich domains are highlighted in blue colour.

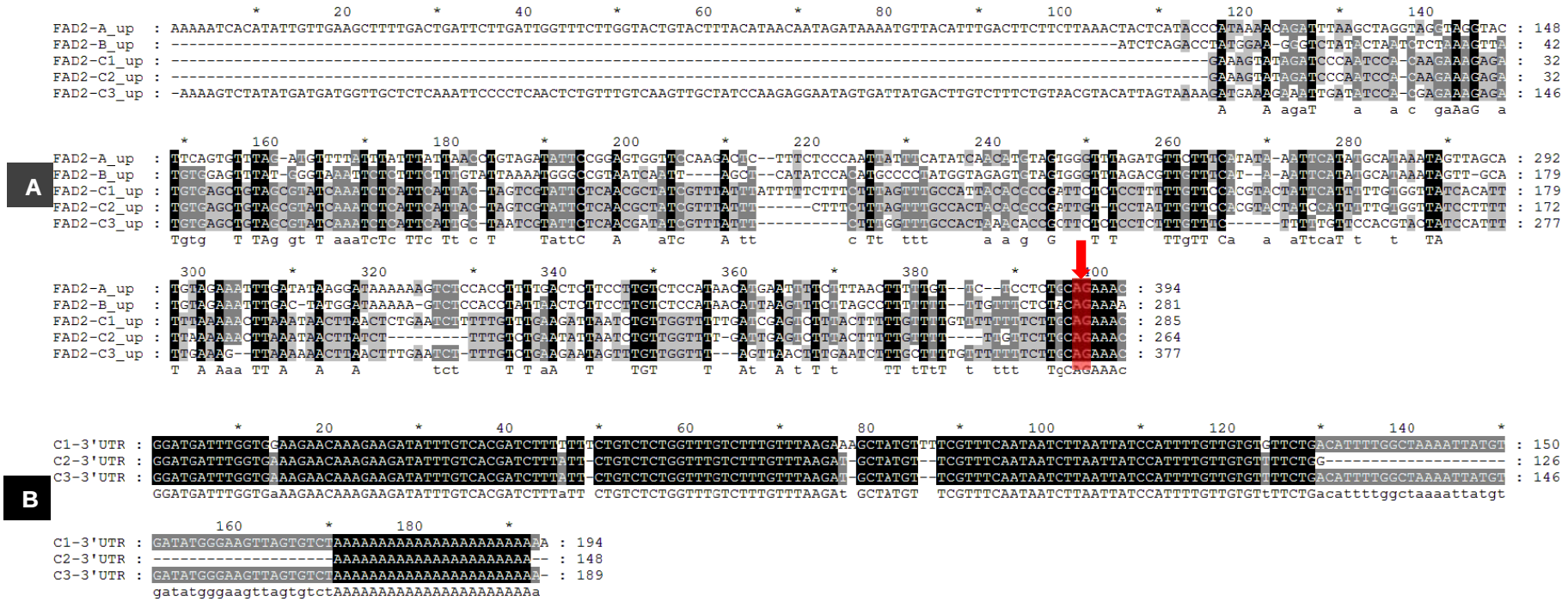


Figure S3. The nucleotide alignments of partial intron sequences (A) and 3'-UTR sequences (B) of crambe *FAD2*. The arrow indicates an intron splice site (AG)

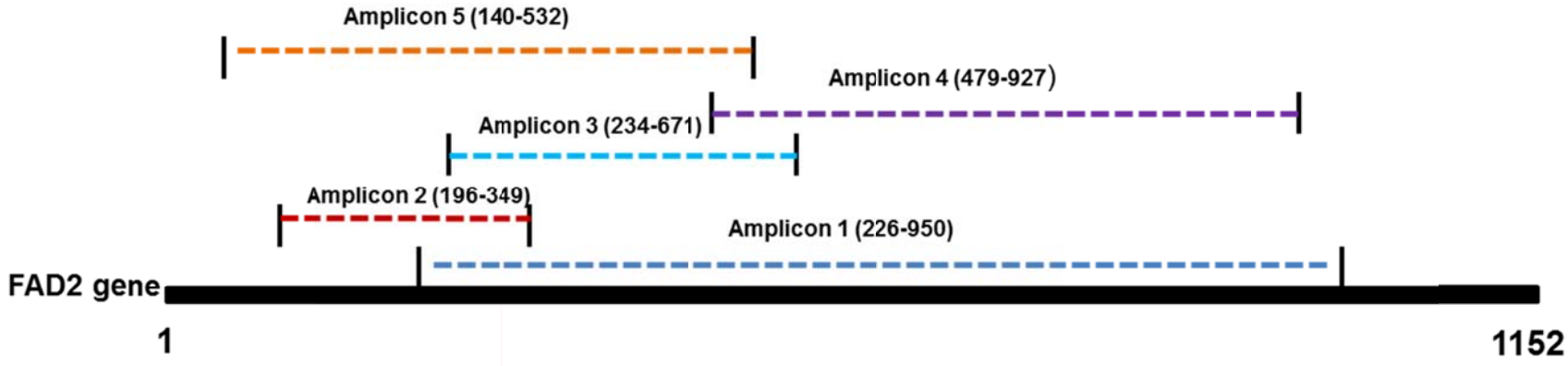


Figure S4. Schematic diagram on the locations of five amplicons in crambe *FAD2* gene. Amplicon1 (226-950) = amplicon-1 starts from the position 226 to position 950.

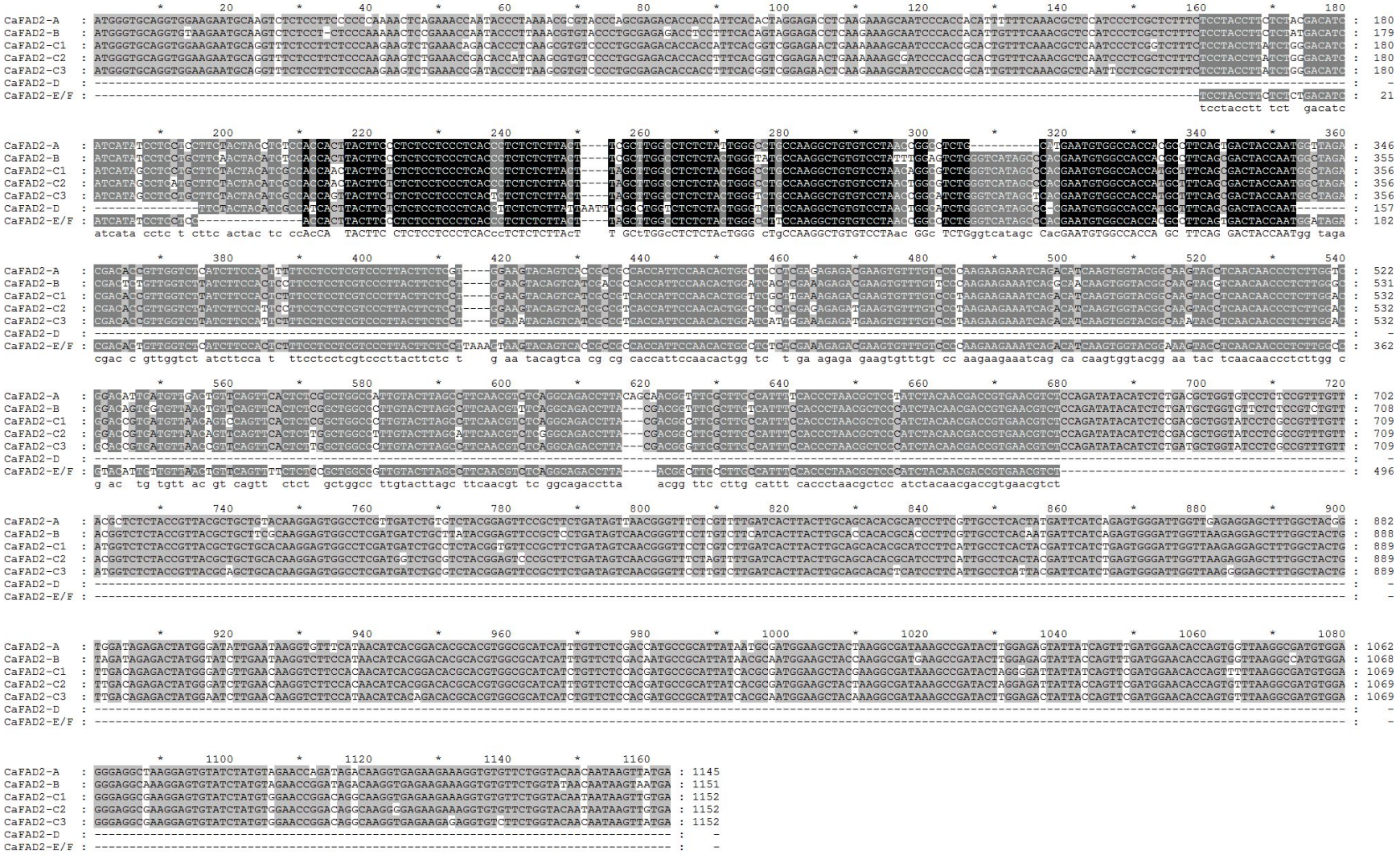


Figure S5. The nucleotide alignment of seven *FAD2* genes in *C. abyssinica*. *CaFAD2-E* and *CaFAD2-F* are considered to be one gene CaFAD2-E/F (see results). The symbols of "-" indicate unavailable sequences.

Figure S6. The amino acids alignment of FAD2 family in *C. abyssinica*. The amino acid sequences were deduced from DNA sequences of eight *FAD2* variants. *CaFAD2-A*, *-B*, *-D*, *-E* and *-F* harbour internal stop codons (indicated by *) and encode for truncated FAD2 proteins. The amino acids shown in grey blocks are polymorphic residues among CaFAD2-C1, -C2 and -C3.

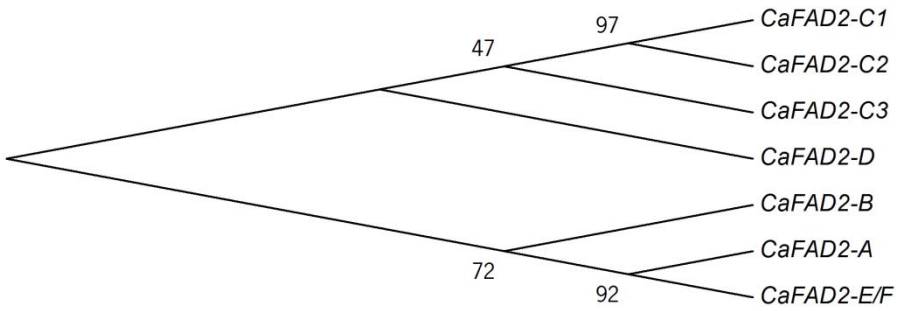


Figure S7. Genetic variation of the *FAD2* family in *C. abyssinica*. An amplicon (142 bp) covering all *CaFAD2* variants was used to test bootstrap of phylogeny with Neighbour-Joining method (MEGA4, USA).

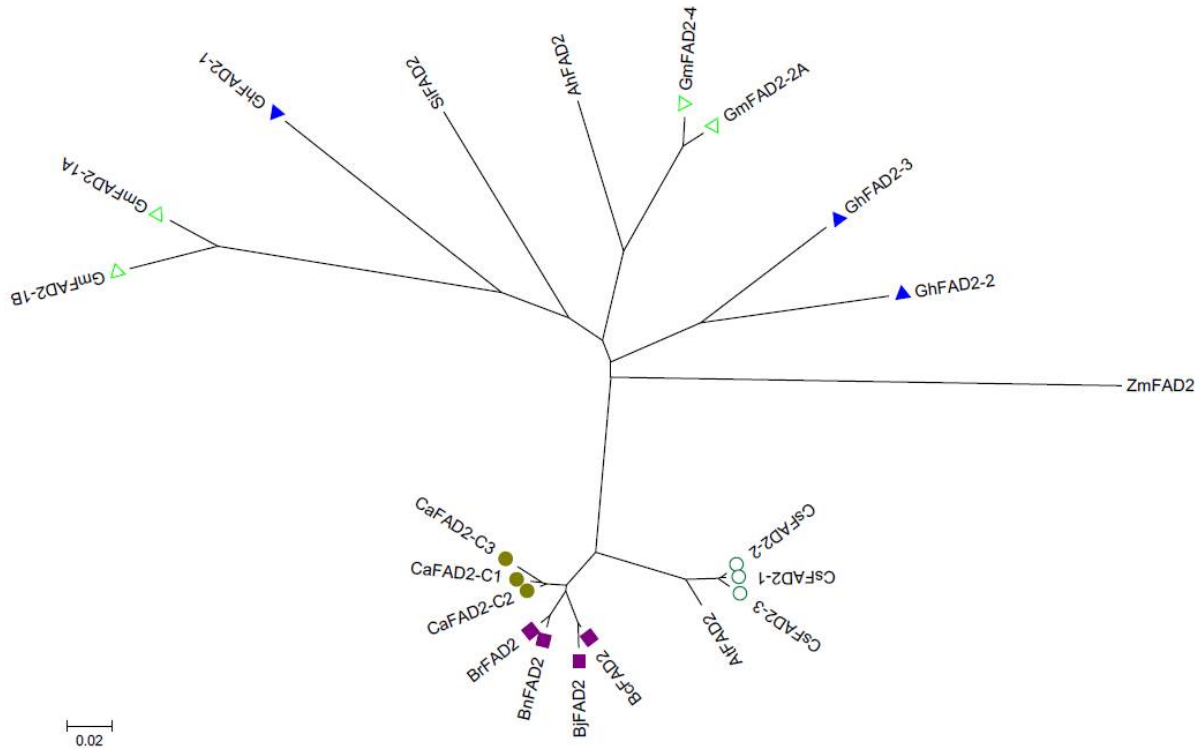


Figure S8. Phylogenetic analysis of deduced amino acids of *FAD2* in plants. A phylogenetic tree was constructed with Neighbour-Joining method (MEGA4). Accession numbers of different *FAD2* in the analysis: *Arabidopsis thaliana* (*AtFAD2*, L26296.1); *Arachis hypogaea* (*AhFAD2*, FJ768732.1); *Brassica carinata* (*BcFAD2*, AF124360.2); *Brassica juncea* (*BjFAD2*, gi|1212780); *Brassica napus* (*BnFAD2*, AF243045.1); *Brassica rapa* (*BrFAD2*, gi|20520623); *Camelina sativa* (*CsFAD2-1*, HQ008320.1; *CsFAD2-2*, HQ008321.1; *CsFAD2-3*, HQ008322.1); *Crambe abyssinica* (*CaFAD2-C1*, *CaFAD2-C2* and *CaFAD2-C3* in this study); *Gossypium hirsutum* (*GhFAD2-1*, gi|2578032; *GhFAD2-2*, gi|4490390; *GhFAD2-3*, AF331163.1); *Glycine max* (*GmFAD2-1A*, L43920.1; *GmFAD2-1B*, gi|59956943; *GmFAD2-2A*, gi|59956945; *GmFAD2-4*, gi|351725574); *Sesamum indicum* (*SiFAD2*, AF192486.1); *Zea mays* (*ZmFAD2*, EU965315.1)

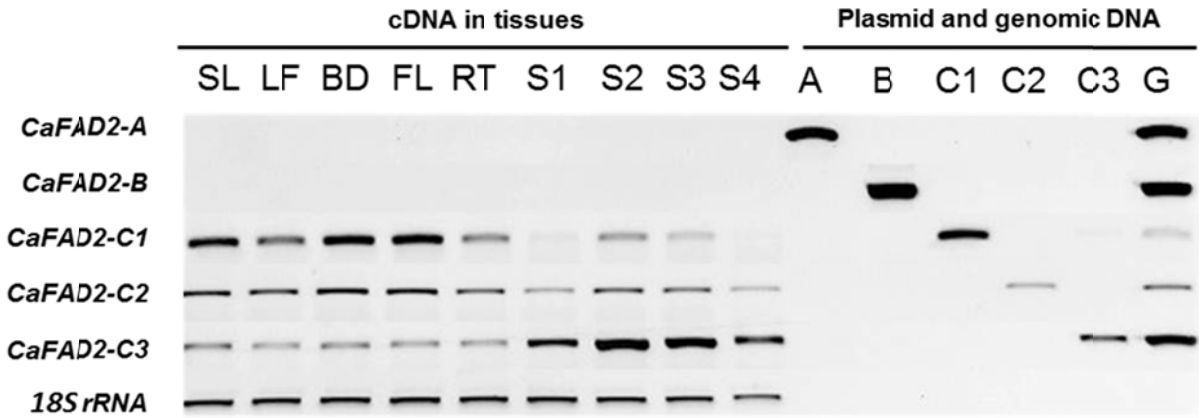


Figure S9. Expression profile of *CaFAD2* genes in vegetative and reproductive tissues of *C. abyssinica* cv. 'Galactica'. The expression of *CaFAD2-A, -B*, *-C1* to *-C3* was measured using RT-PCR in different tissues. SL=seedling, LF=leaf, BD=flower bud, FL=flower, RT=root, S1=7 DAP seeds, S2=14 DAP seed, S3=21 DAP seed, S4= 28 DAP seed. DAP=Days after pollination. Lane A to C3 are positive controls with plasmid clones of the different variants; G= Genomic DNA from *C. abyssinica* used as a template for PCR with the set of 5 primer combinations for the detection of five FAD2 genes. RT-PCR-product of 18S ribosomal RNA (*18S rRNA*) was used as a reference for the amount of cDNA (reference)

.

References

Aldrich J, Cullis CA (1993) RAPD analysis in flax: optimization of yield and reproducibility using klen Taq 1 DNA polymerase, chelex 100, and gel purification of genomic DNA. Plant Molecular Biology Reporter 11 (2):128-141

Beló A, Zheng P, Luck S, Shen B, Meyer DJ, Li B, Tingey S, Rafalski A (2008) Whole genome scan detects an allelic variant of fad2 associated with increased oleic acid levels in maize. Molecular Genetics and Genomics 279 (1):1-10

Buggs RJA, Renny-Byfield S, Chester M, Jordon-Thaden IE, Viccini LF, Chamala S, Leitch AR, Schnable PS, Bradley Barbazuk W, Soltis PS, Soltis DE (2012) Next-generation sequencing and genome evolution in allopolyploids. American Journal of Botany 99 (2):372-382

Carlsson AS, Clayton D, Salentijn E, Toonen M, Stymne S, Dyer WJ, Bowles D (2007). Oil Crop Platforms for Industrial Uses. CPL Press, Newbury, York

Chen X, Truksa M, Snyder CL, El-Mezawy A, Shah S, Weselake RJ (2011) Three homologous genes encoding sn-glycerol-3-phosphate acyltransferase 4 exhibit different expression patterns and functional divergence in brassica napus. Plant Physiology 155 (2):851-865

Falasca S, Flores N, Lamas M, Carballo SM, Anschau A (2010) Crambe abyssinica: An almost unknown crop with a promissory future to produce biodiesel in Argentina. International Journal of Hydrogen Energy 35 (11):5808-5812

Griffin PC, Robin C, Hoffmann AA (2011) A next-generation sequencing method for overcoming the multiple gene copy problem in polyploid phylogenetics, applied to Poa grasses. BMC Biology 9. doi:10.1186/1741-7007-9-19

Gurr MI, Blades J, Appleby RS (1972) Studies on seed-oil triglycerides. The composition of Crambé abyssinica triglycerides during seed maturation. European Journal of Biochemistry 29 (2):362-368

Heppard EP, Kinney AJ, Stecca KL, Miao GH (1996) Developmental and growth temperature regulation of two different microsomal [omega]-6 desaturase genes in soybeans. Plant Physiology 110 (1):311-319

Hu X, Sullivan-Gilbert M, Gupta M, Thompson SA (2006) Mapping of the loci controlling oleic and linolenic acid contents and development of fad2 and fad3 allele-specific markers in canola (Brassica napus L.). TAG Theoretical and Applied Genetics 113 (3):497-507

Hu Y, Wu G, Cao Y, Wu Y, Xiao L, Li X, Lu C (2009) Breeding response of transcript profiling in developing seeds of Brassica napus. BMC Molecular Biology 10 (1):49

Hutcheon C, Ditt R, Beilstein M, Comai L, Schroeder J, Goldstein E, Shewmaker C, Nguyen T, De Rocher J, Kiser J (2010) Polyploid genome of Camelina sativa revealed by isolation of fatty acid synthesis genes. BMC Plant Biology 10 (1):233

Jadhav A, Katavic V, Marillia EF, Michael Giblin E, Barton DL, Kumar A, Sonntag C, Babic V, Keller WA, Taylor DC (2005) Increased levels of erucic acid in Brassica carinata by co-suppression and antisense repression of the endogenous FAD2 gene. Metabolic Engineering 7 (3):215-220

Kang J, Snapp AR, Lu C (2011) Identification of three genes encoding microsomal oleate desaturases (FAD2) from the oilseed crop Camelina sativa. Plant Physiology and Biochemistry 49 (2):223-229

Kim MJ, Kim H, Shin JS, Chung CH, Ohlrogge JB, Suh MC (2006) Seed-specific expression of sesame microsomal oleic acid desaturase is controlled by combinatorial properties between negative cis-regulatory elements in the SeFAD2 promoter and enhancers in the 5'-UTR intron. Molecular Genetics and Genomics 276 (4):351-368

Leppik EE, White GA (1975) Preliminary assessment of Crambe germplasm resources. Euphytica 24 (3):681-689

Li X, van Loo EN, Gruber J, Fan J, Guan R, Frentzen M, Stymne S, Zhu LH (2012) Development of ultra - high erucic acid oil in the industrial oil crop Crambe abyssinica. Plant biotechnology journal 10 (7):862-870

Liu Q, Brubaker CL, Green AG, Marshall DR, Sharp PJ, Singh SP (2001) Evolution of the FAD2-1 fatty acid desaturase 5'UTR intron and the molecular systematics of Gossypium (Malvaceae). American Journal of Botany 88 (1):92-102

Liu Q, Singh SP, Green AG (2002) High-stearic and high-oleic cottonseed oils produced by hairpin RNA-mediated post-transcriptional gene silencing. Plant Physiology 129 (4):1732

Lu C, Napier JA, Clemente TE, Cahoon EB (2011) New frontiers in oilseed biotechnology: meeting the global demand for vegetable oils for food, feed, biofuel, and industrial applications. Current Opinion in Biotechnology 22 (2):252-259

Makarenko SP, Dudareva LV, Katyshev AI, Konenkina TA, Nazarova AV, Rudikovskaya EG, Sokolova NA, Chernikova VV, Konstantinov YM (2011) The effect of low temperatures on fatty acid composition of crops with different cold resistance. Biochemistry (Moscow) Supplement Series A: Membrane and Cell Biology 5 (1):64-69

Martínez-Rivas JM, Sperling P, Lühs W, Heinz E (2001) Spatial and temporal regulation of three different microsomal oleate desaturase genes (FAD2) from normal-type and high-oleic varieties of sunflower (Helianthus annuus L.). Molecular Breeding 8 (2):159-168

Mastebroek HD, Wallenburg SC, Van Soest LJM (1994) Variation for agronomic characteristics in crambe (Crambe abyssinica Hochst. ex Fries). Industrial Crops and Products 2 (2):129-136

Mietkiewska E, Hoffman TL, Brost JM, Giblin EM, Barton DL, Francis T, Zhang Y, Taylor DC (2008) Hairpin-RNA mediated silencing of endogenous FAD2 gene combined with heterologous expression of Crambe abyssinica FAE gene causes an increase in the level of erucic acid in transgenic Brassica carinata seeds. Molecular Breeding 22 (4):619-627

Ohlrogge J (1995) Lipid biosynthesis. Plant Cell 7 (7):957

Okuley J, Lightner J, Feldmann K, Yadav N, Lark E (1994) Arabidopsis FAD2 gene encodes the enzyme that is essential for polyunsaturated lipid synthesis. Plant Cell 6 (1):147-158

Oliva ML, Shannon JG, Sleper DA, Ellersieck MR, Cardinal AJ, Paris RL, Lee JD (2006) Stability of fatty acid profile in soybean genotypes with modified seed oil composition. Crop Science 46 (5):2069-2075
Paulose B, Kandasamy S, Dhankher OP (2010) Expression profiling of Crambe abyssinica under arsenate stress identifies genes and gene networks involved in arsenic metabolism and detoxification. BMC Plant Biology 10
Pham AT, Lee JD, Shannon JG, Bilyeu K (2010) Mutant alleles of FAD2-1A and FAD2-1B combine to produce soybeans with the high oleic acid seed oil trait. BMC Plant Biology 10 (1):195
Pham AT, Lee JD, Shannon JG, Bilyeu KD (2011) A novel FAD2-1 A allele in a soybean plant introduction offers an alternate means to produce soybean seed oil with 85% oleic acid content. TAG Theoretical and Applied Genetics:1-10
Pirtle IL, Kongcharoensuntorn W, Nampaisansuk M, Knesek JE, Chapman KD, Pirtle RM (2001) Molecular cloning and functional expression of the gene for a cotton Δ-12 fatty acid desaturase (FAD2). Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression 1522 (2):122-129
Rudloff E, Wang Y (2011) Crambe. Wild Crop Relatives: Genomic and Breeding Resources.
Salentijn EMJ, Goryunova SV, Bas N, van der Meer IM, van den Broeck HC, Bastien T, Gilissen LJ, Smulders MJ (2009) Tetraploid and hexaploid wheat varieties reveal large differences in expression of alpha-gliadins from homoeologous Gli-2 loci. BMC Genomics 10:48
Sasongko ND, Möllers C (2005) Toward increasing erucic acid content in oilseed rape (Brassica napus L.) through the combination with genes for high oleic acid. Journal of the American Oil Chemists' Society 82 (6):445-449
Scheffler JA, Sharpe AG, Schmidt H, Sperling P, Parkin IAP, Lühs W, Lydiate DJ, Heinz E (1997) Desaturase multigene families of Brassica napus arose through genome duplication. TAG Theoretical and Applied Genetics 94 (5):583-591
Schierholt A, Becker H, Ecke W (2000) Mapping a high oleic acid mutation in winter oilseed rape (Brassica napus L.). TAG Theoretical and Applied Genetics 101 (5):897-901
Schlueter JA, Vasylenko-Sanders IF, Deshpande S, Yi J, Siegfried M, Roe BA, Schlueter SD, Scheffler BE, Shoemaker RC (2007) The FAD2 Gene Family of Soybean: Insights into the Structural and Functional Divergence of a Paleopolyploid Genome. Crop Science 47(S1)
Shanklin J, Whittle E, Fox BG (1994) Eight histidine residues are catalytically essential in a membrane-associated iron enzyme, stearoyl-CoA desaturase, and are conserved in alkane hydroxylase and xylene monooxygenase. Biochemistry 33 (43):12787-12794
Shi J, Cao Y, Fan X, Li M, Wang Y, Ming F (2012) A rice microsomal delta-12 fatty acid desaturase can enhance resistance to cold stress in yeast and Oryza sativa. Molecular Breeding 29 (3):743-757
Sivaraman I, Arumugam N, Sodhi YS, Gupta V, Mukhopadhyay A, Pradhan AK, Burma PK, Pental D (2004) Development of high oleic and low linoleic acid transgenics in a zero erucic acid Brassica juncea L.(Indian mustard) line by antisense suppression of the fad2 gene. Molecular Breeding 13 (4):365-375
Slabas AR, Hanley Z, Schierer T, Rice D, Turnbull A, Rafferty J, Simon B, Brown A (2001) Acyltransferases and their role in the biosynthesis of lipids-opportunities for new oils. Journal of Plant Physiology 158 (4):505-513
Slade AJ, Fuerstenberg SI, Loeffler D, Steine MN, Facciotti D (2005) A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. Nature Biotechnology 23 (1):75-81
Soltis PS, Burleigh JG, Chanderbali AS, Yoo MJ, Soltis DE (2010) Gene and genome duplications in plants. Evolution after Gene Duplication:269-298
Stoutjesdijk PA, Singh SP, Liu Q, Hurlstone CJ, Waterhouse PA, Green AG (2002) hpRNA-mediated targeting of the Arabidopsis FAD2 gene gives highly efficient and stable silencing. Plant Physiology 129 (4):1723
Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24 (8):1596-1599
Tang GQ, Novitzky WP, Carol Griffin H, Huber SC, Dewey RE (2005) Oleate desaturase enzymes of soybean: evidence of regulation through differential stability and phosphorylation. The Plant Journal 44 (3):433-446
Tanhuanpää P, Vilkki J, Vihinen M (1998) Mapping and cloning of FAD2 gene to develop allele-specific PCR for oleic acid in spring turnip rape (Brassica rapa ssp. oleifera). Molecular Breeding 4 (6):543-550
Van Poecke RMP, Sato M, Lenarz-Wyatt L, Weisberg S, Katagiri F (2007) Natural variation in RPS2-mediated resistance among Arabidopsis accessions: Correlation between gene expression profiles and phenotypic responses. Plant Cell 19 (12):4046-4060
Varshney RK, Nayak SN, May GD, Jackson SA (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends in Biotechnology 27 (9):522-530
Wang Y, Luo P (1998) Intergeneric hybridization between Brassica species and Crambe abyssinica. Euphytica 101 (1):1-7
Yuan SW, Wu XL, Liu ZH, Luo HB, Huang RZ (2012) Abiotic Stresses and Phytohormones Regulate Expression of FAD2 Gene in Arabidopsis thaliana. Agricultural Sciences in China 11 (1):62-72
Zanetti F, Vamerali T, Bona S, Mosca G (2006) Can we "cultivate" erucic acid in southern Europe? Italian Journal of Agronomy 1 (1):3-10

Chapter 3

Functional analysis of the omega-6 fatty acid desaturase (*CaFAD2*) gene family of the oil seed crop *Crambe abyssinica*

Jihua Cheng[1,2] , Li-Hua Zhu[3], Elma M.J. Salentijn[1], Bangquan Huang[2], Jens Gruber[4], Annemarie C. Dechesne[1], Frans A. Krens[1], Weicong Qi[1,2], Richard G.F. Visser[1], Eibertus N. van Loo[1*]

[1] Wageningen UR Plant Breeding, P.O. Box 16, 6700 AA Wageningen, the Netherlands

[2] College of Life Science, Hubei University, People's Republic of China

[3] Plant Breeding and Biotechnology, Swedish University of Agricultural Science, Sweden

[4] Institute for Biology I-Botany, RWTH Aachen University, Aachen, Germany

[*] Corresponding author

Published in BMC Plant Biology 2013, 13:146

Abstract

Background

Crambe abyssinica produces high erucic acid (C22:1, 55-60 %) in the seed oil, which can be further increased by reduction of polyunsaturated fatty acid (PUFA) levels. The omega-6 fatty acid desaturase enzyme (FAD2) is known to be involved in PUFA biosynthesis. In crambe, three *CaFAD2* genes, *CaFAD2-C1, CaFAD2-C2* and *CaFAD2-C3* are expressed.

Results

The individual effect of each *CaFAD2* gene on oil composition was investigated through studying transgenic lines (*CaFAD2-RNAi*) for differential expression levels in relation to the composition of seed-oil. Six first generation transgenic plants (T_1) showed C18:1 increase (by 6% to 10.5 %) and PUFA reduction (by 8.6% to 10.2 %). The silencing effect in these T_1-plants ranged from the moderate silencing (40% to 50% reduction) of all three *CaFAD2* genes to strong silencing (95% reduction) of *CaFAD2-C3* alone. The progeny of two T_1-plants (WG4-4 and WG19-6) was further analysed. Four or five transgene insertions are characterized in the progeny (T_2) of WG19-6 in contrast to a single insertion in the T_2 progeny of WG4-4. For the individual T_2-plants of both families (WG19-6 and WG4-4), seed-specific silencing of *CaFAD2-C1* and *CaFAD2-C2* was observed in several individual T_2-plants but, on average in both families, the level of silencing of these genes was not significant. A significant reduction in expression level ($P < 0.01$) in both families was only observed for *CaFAD2-C3* together with significantly different C18:1 and PUFA levels in oil.

Conclusions

CaFAD2-C3 expression is highly correlated to levels of C18:1 (r = -0.78) and PUFA (r = 0.75), which suggests that *CaFAD2-C3* is the most important one for changing the oil composition of crambe.

Keywords: *Crambe abyssinica*, Fatty Acid Desaturase 2, Oil Crop, Oleic Acid, Polyunsaturated Fatty Acid, RNAi, Gene Expression

Background

Crambe (*Crambe abyssinica* Hochst ex. R. E. Fr.) belongs to the *Brassicaceae* plant family. The seed oil of crambe contains a high content of erucic acid (C22:1, 55% to 60%) and this oil thus has applications as industrial oil (Mastebroek et al., 1994). Erucic acid in form of erucamide, a major derivative from C22:1, can be used as slip agent in plastics, or lubricants, nylon and cosmetics (Mietkiewska et al., 2007). More recently, crambe oil is considered to be potential to produce biofuel (Falasca et al., 2010). The cultivation of crambe yields up to 1 t ha^{-1} of oil, comparative to that of high erucic acid rapeseed (HEAR). Furthermore, the processing costs for crambe oil extraction are in the same range of that for rapeseed (Carlsson et al., 2007; Cuperus et al., 1996; Mastebroek et al., 1994). The high yield and the fact that crambe is not able to cross with food oil crops in nature make crambe an ideal platform to produce industrial oils (Youping and Peng, 1998; Zanetti et al., 2006). In addition to high erucic acid (C22:1), crambe oil contains also certain amounts of oleic acid (C18:1), linoleic acid (C18:2) and linolenic acid (C18:3).

The value of crambe oil can be further improved not only by increasing the C22:1 content but also by reducing the content of polyunsaturated fatty acids (PUFA, C18:2 + C18:3). Firstly, an increase in C22:1 can reduce the purification cost of C22:1 from C18-fatty acids. It was estimated that 10% increase of C22:1 in oil would reduce the processing costs by half (Jadhav et al., 2005). Secondly, reduction in PUFA is beneficial for storage and extending the shelf life of the oil because PUFA are highly prone to oxidation during storage(Yang et al., 2012). Thirdly, high PUFA content causes higher viscosity for the oil (a disadvantage to biodiesel) (Ma and Hanna, 1999). Considerable efforts have been made to reduce PUFA in oil. A common chemical approach is to reduce the PUFA content or increase the C18:1 content through hydrogenation (Pham et al., 2012; Yang et al., 2012). However, this hydrogenating process is expensive and adds extra 2 to 3 cent per pound cost to the price of oil (Ohlrogge, 1994).

Molecular breeding approaches to change composition of seed oil are targeting important genes involved in the fatty acid biosynthesis pathways, which have been intensively studied and many genes involved have been characterized. In short, C18:1 is *de novo* synthesized in plastids and transported into the endoplasmic reticulum where the C18:1 is incorporated into phosphatidylcholine (PC) and may undergo desaturation to C18:2 and next to C18:3 by the actions of two microsomal enzymes; delta-12-fatty acid desaturase (FAD2) [omega-6

desaturase] and delta-15-fatty acid desaturase (FAD3) [omega-3 desaturase], respectively (Ohlrogge, 1995; Yang et al., 2012). Alternatively, C18:1 may undergo elongation to very long chain fatty acids (VLCFA, C22:1 for example) by the action of fatty acid elongation (FAE) complex. (Barret et al., 1998; Blacklock and Jaworski, 2002; Mietkiewska et al., 2007; Rossak et al., 2001; Salas et al., 2005). Mutant and genetic mapping studies showed that the enzyme FAD2 was found to be mainly responsible for C18:1 and PUFA content although FAD3 also contributes to a limited degree to these traits (Miquel, 1992; Yang et al., 2012). In addition, the acyl flux between the two pathways, the prokaryotic (plastidial) and eukaryotic (mainly in the ER) pathway, is influencing the C18:1 pool (Andre et al., 2012; Guan et al., 2012; Löhden and Frentzen, 1988)

Genetic modification (GM) aimed at regulating the *FAD2* expression has been applied to produce oils with higher C18:1 in various oil crops (Jadhav et al., 2005; Liu et al., 2002; Sivaraman et al., 2004). For example, by anti-sense suppression of *FAD2* in *Brassica juncea*, a transgenic line was obtained that produced oil with higher C18:1 (73%) and lower PUFA (8% of C18:2; 9% of C18:3) compared to the wild type (53% of C18:1; 24% of C18:2; 16% of C18:3) (Sivaraman et al., 2004). Similarly, a significant change in C18:1 and PUFA and even a C22:1 increase in the seed oil was observed when silencing *FAD2* with both co-suppression and anti-sense in *Brassica carinata* (Jadhav et al., 2005). Gene silencing by RNAi has been considered to be a particularly efficient way to obtain stable transgenic plants with the silenced target genes (Rathore et al., 2012; Wesley et al., 2001). RNAi-mediated silencing of *GhFAD2* in cotton enabled over 60% increase in C18:1 (Liu et al., 2002). Simultaneous RNAi-mediated silencing of *FAD2* and *FAE1* in *Brassica napus* caused not only significant increase in C18:1 (from 62% to 85%), but also reduction in C22:1 and PUFA (from 26% to 10% and from 0.87% to 0% respectively) (Peng et al., 2010). Recently, several efficient protocols for crambe transformation are available (Chhikara et al., 2012; Li et al., 2010) and RNAi has shown to be an effective gene knockdown tool for crambe where *CaFAD2* RNAi gene silencing resulted in increased C18:1 levels (from 14.5 % to 24.9%) (Li et al., 2012). Introduction of two heterologous genes, *LdLPAAT* and *BnFAE1*, in such *CaFAD2*-RNAi lines directed the oil biosynthesis towards the incorporation of C22:1 at the *sn*-2 position of triacylglycerol, thereby the C22:1 level increased from 60% in the wild type to 73% in the best transgenic crambe line (Li et al., 2012).

An obstacle to adaptation of such genetic modified (GM) crops is the lack of broad acceptance by a part of the community in many countries (Gómez-Galera et al., 2012). Furthermore, there are some cases where RNAi-mediated traits are not completely reliable on the long term in generating stable target gene suppression (Rathore et al., 2012). Currently, breeding of the allo-hexaploid crambe mainly relies on traditional approaches, however, the possibilities are restricted by the lack of genetic variation for important agronomic traits (Cheng et al., 2013; Lessman and Meier, 1972; Mastebroek et al., 1994). In such situations and particularly when genes controlling a phenotype are known, mutation breeding of induced or natural mutations, identified via "TILLING" (Targeting Induced Local Lesions IN Genomes) (McCallum et al., 2000), offers a reliable, stable, non-GM approach to obtain the desired oil quality in crambe. A drawback is that "TILLING" is still a challenging task in polyploid crops, where the multiple alleles are creating problems in identifying desirable genetic changes due to gene redundancy. Despite these problems successful cases of targeted mutagenesis have been reported in polyploid crops (Auld et al., 1992; Chen et al., 2012; Pham et al., 2010; Pham et al., 2011; Slade et al., 2012). For instance, targeted mutation breeding of natural or induced variation in the *FAD2* gene or mining natural variants has been used to develop crops with high C22:1 or less PUFA (Beló et al., 2008; Pham et al., 2011; Tanhuanpää et al., 1998). By combining mutations or natural variants of *FAD2* and *FAD3*, it was possible to produce oils with higher C18:1 and lower C18:3 in soybean and *Brassica napus* (Pham et al., 2012; Yang et al., 2012). In the allo-hexaploid genome of *Crambe abyssinica* cv. 'Galactica' seven *FAD2* genes are present, of which only three are transcriptionally active throughout plant development (*CaFAD2-C1, CaFAD2-C2* and *CaFAD2-C3*) (Cheng et al. 2013). The possibility of functional redundancy among the active *CaFAD2* family members may complicate the gain of crambe lines with desirable oil composition via mutation breeding.

The aim of the present study is to specify which of the *FAD2* genes in crambe is the key gene for increasing the C18:1 level, but reducing the PUFA content. *FAD2*-RNAi lines of crambe cv. 'Galactica' were studied for functional correlations between the individual *CaFAD2* family members, *CaFAD2-C1*, *CaFAD2-C2* and *CaFAD2-C3* and seed-oil composition. Seed-oil composition and gene expression studies were performed in two independent families of the second generation transgenic lines (T_2-plants). In addition, two other genes, *CaFAD3* and *CaFAE1* that are acting in close connection to *CaFAD2*, were also involved in the study.

Methods

Plant materials

C. abyssinica cv. 'Galactica' was previously transformed with an RNAi construct (pWatergate) (Li et al., 2012). This construct contains an inverted repeat (IR) of the *CaFAD2-C2* coding sequence (355 base pairs for each IR part, 97 % identical to *CaFAD2-C1* and 96 % identical to *CaFAD2-C3*) following a seed specific promoter (*Napin*) (**Additional file 1**). The plant used as control was transformed with an empty construct (pRCNG) which contains no genes involved in fatty acids biosynthesis (Schaart et al., 2004). Two T_0 lines were used to develop six T_1 plants/lines (WG4-3, WG4-4, WG4-5, WG19-4, WG19-5 and WG19-6), grown in the greenhouse with temperature of 22 °C and photoperiod of 16 h. From these T_1 plants, the developing seeds at 20 DAP (days after pollination) were collected for RNA isolation for gene expression analysis and ripe seeds were harvested for oil composition analysis. Seeds of two T_1-lines, WG4-4 and WG19-6, were used to produce T_2-plants. The young leaves of T_2-plants were collected for Southern analysis. The developing seeds (20 DAP) and ripe seeds of the T_2-plants were harvested for RNA isolation and oil composition analysis respectively.

Quantitative reverse transcription PCR (qRT-PCR) and data analysis

To examine the level of gene-silencing, the expression of the individual crambe FAD2 genes, *CaFAD2-C1*, *CaFAD2-C2* and *CaFAD2-C3*, was measured in six T_1 plants (WG4-3, WG4-4, WG4-5, WG19-4, WG19-5 and WG19-6) on bulks of five to ten developing seeds (20 DAP, days after pollination) and compared to their expression level in the control (plant transformed with an empty construct). The primers specific for the different crambe *FAD2* genes were developed based on their sequences in *C. abyssinica* cv. 'Galactica' (GenBank: JX964743, JX964744, JX964745), and the primers of *FAD3* and *FAE1* were developed on the sequences of *AtFAD3* (GenBank: 42570333) and crambe *FAE1* (GenBank: 60543786). The gene *β-actin* 2 (GenBank: 20465834) was used as a reference gene. All primers used in qRT-PCR are listed in **Additional file 2**. Total RNA was extracted from bulked seeds of T_1- and T_2 -plants respectively (5 to 10 seeds per T_1- plant and per T_2-plant, 20 DAP) with RNeasy Plant Mini Kits (Qiagen, Germany) according to the manufacturer's instructions. The isolated RNA was treated with RNase-free TURBO DNase (Ambion, USA) to remove residual genomic DNA. First-strand cDNA was synthesized in 20 µl from 1 µg of total RNA with iScript™ cDNA

Synthesis Kit (Bio-rad, USA), in parallel 1 µg of RNA of each sample was treated in the same way but without adding reverse transcriptase, as negative controls (RT-). The cDNA was 20 × diluted and used as templates for real-time PCR. The PCR reaction contains 2 µl templates, 5 µl SYBR Green Super Mix (Bio-rad, USA), and 1 µl of each of the forward and reverse primers (3 µM) in total 10 µl reaction. Cycling conditions were 1 cycle at 95 °C for 3 min followed by 30 cycles at 95 °C for 10 s, 60 °C for 1 min, then a final melt step from 65 °C to 95 °C ramp with 0.5 °C increments per cycle to monitor specificity. PCR reactions were performed in triplicate. The expression of each replicate was normalized by the reference gene, *β-actin* 2, which has shown to be stably expressed in both, crambe seedlings under arsenate stress and various *Brassica napus* cultivars (Hu et al., 2009; Paulose et al., 2010). The relative expression level of each replicate was calculated according to the comparative CT method (User bulletin no. 2, ABI PRISM 7700 Sequence Detection System, December 1997; Perkin-Elmer, Applied Biosystems). The mean of three replicates represents the relative expression level of a line.

Oil composition and correlation analysis

The fatty acid (FA) fraction was extracted from single crambe seeds, and fatty acid methyl esters (FAMEs) were analysed using gas chromatograph (GC, column/DB-23, Agilent). For each T_1-plant 20, and for each T_2-plant 10 individual ripe seeds were collected respectively per plant for fatty acid extraction. The pods of single seeds were removed and then crushed in a plastic tube with screw cap and 300 µl of hexane and 40 µl of KOH (5 M)/methanol were added, followed by vortexing and heating samples at 60 °C for 6 min. The samples were then allowed to cool down to room temperature before centrifugation at 3000 rpm for 5 min, the upper layer was used for GC. The extraction (1 µl) was injected into GC with split ratio 1:20 and the condition of oven was 180 °C for 10 min, ramp to 240 °C for 7 min with 7.5 °C increments per min. The mean percentages of FA compounds of each line were calculated from the average of 20 or 10 single seed values. The data of gene expression and oil composition in the seeds of T_2 generation were used for correlation analysis. These data were plotted and a Pearson correlation was run to determine correlation with R package (RDevelopment, 2008).

Southern analysis

Genomic DNA was isolated from young leaves of T_2 plants with the method described by Aldrich and Cullis (1993) but with 1 % (w/v) polyvinylpyrrolidone-10 in the DNA extraction

buffer. A probe (686 base pairs) was designed on the *nptII* gene in the WG construct and labelled with [^{32}P]ATP. The primers for nptII probe preparation are listed in **Additional file 2**.

For copy number determination, a total of 40 μg of DNA extracted was digested overnight with *Dra* I, an enzyme that cuts the T-DNA in a site outside the probe sequence, fractionated on 0.8 % agarose gel and transferred to Hybond N+ membrane (Amersham Biosciences, UK) according to the manufacturer's recommendations. The procedures of hybridization and visualization were performed as described by Cheng (2013).

Results

CaFAD2-RNAi lines

We have previously shown that the major expressed *CaFAD2* gene in developing crambe seeds is *CaFAD2-C3,* while the other two genes *FAD2-C1* and *CaFAD2-C2* are naturally expressed to much lower levels (4 and 100 times lower than *CaFAD2-C3* respectively) (Cheng et al., 2013). A DNA fragment of *CaFAD2-C2*, sharing 97% and 96% nucleotide identity with *CaFAD2-C1* and *CaFAD2-C3* respectively, was used to trigger RNAi-mediated silencing of the *CaFAD2* genes in crambe. Two primary transgenic plants (T_0 generation,WG4 and WG19) that showed significantly lower contents of C18:2 and C18:3, but higher level of C18:1 in the seed-oil were used in this study (Li et al., 2012). For these two independent transgenic lines, the effect of RNAi-silencing on the expression of the individual *CaFAD2* genes and the composition of the seed-oil were analysed for the seeds produced by the first generation (T_1) and the second generation (T_2). In the T_2 generation the transgene copy number was determined and the relative gene-expression of the individual genes was correlated to differences in seed-oil composition to determine the effect of silencing of the individual *CaFAD2* gene(s).

CaFAD2 gene-silencing and oil composition in T_1-generation.

Among six individual plants analysed, simultaneous but moderate silencing of all three *CaFAD2* genes (40% to 50% reduction) was found only in one plant, WG4-5. In two plants WG19-5 and WG19-6 no signs of gene-silencing were observed. In the remaining three plants (WG4-3, WG4-4 and WG19-4) no silencing was detected for *CaFAD2-C1* and *CaFAD2-C2* whereas *CaFAD2-C3* was silenced to different levels whereby a strong silencing of *CaFAD2-C3* (95% reduction) was detected in plant WG4-4 followed by a moderate level of *CaFAD2-C3* silencing (40% to 50% reduction) in WG4-3 and WG19-4 (**Figure 1**).

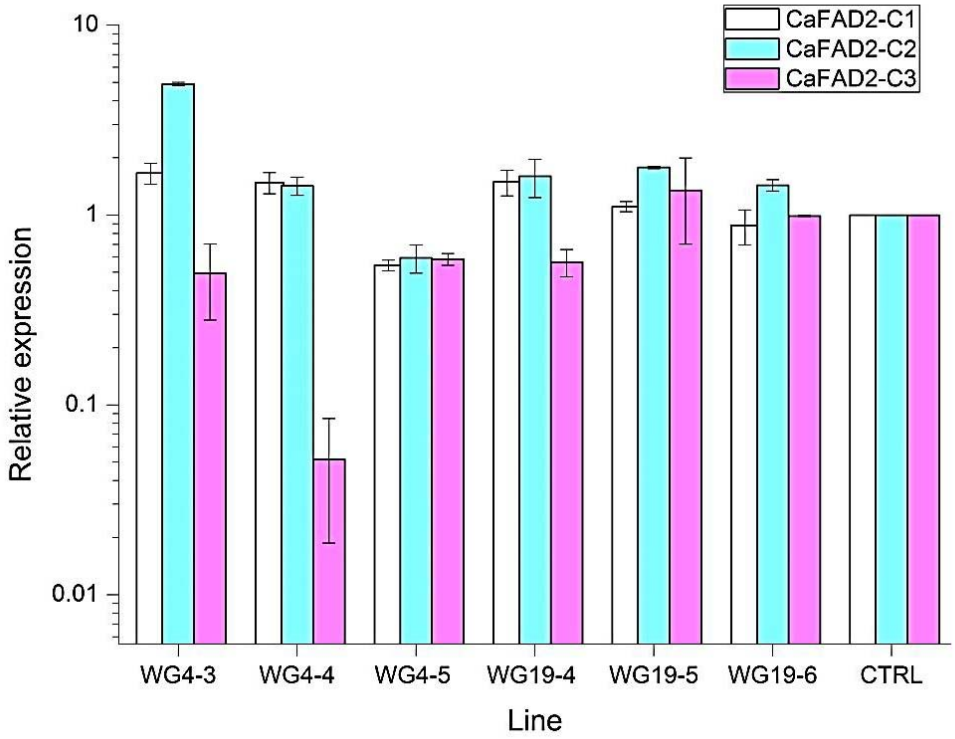


Figure 1. Relative expression levels of CaFAD2 gene family members in the seeds of T1 plants of C. abyssinica. Six T1 plants (WG4-3, WG4-4, WG4-5, WG19-4, WG19-5 and WG19-6) and a control (CTRL) were measured. Qantitification was performed by qPCR for bulk seeds (five to ten, 20 DAP) of each plant. The relative expression levels in the seeds of T1 plants were calculated with 2^-ΔΔCt method using β-ACT2 as a reference gene.

To measure the final effect of gene-silencing, the oil composition was determined for around 20 single ripe seeds per plant. Despite the fact that clear silencing was only detected in the developing seeds of one plant (WG4-4), the contents of C18:1, C18:2 and C18:3 in the seeds of all six plants were significantly different to that of the control (**Table 1**). The C18:1 content was found to be 6% to 10.5% above the control, and the content of PUFA (C18:2 + C18:3) was a concomitant 4.3% to 9.2% lower than the control. Regarding C22:1 content, the plants WG4-3, WG4-4 and WG19-6 scored significantly higher than the control by respectively 2.4% ($P < 0.01$), 3.9% ($P < 0.01$) and 3.1% ($P < 0.01$) whereas the remaining three plants showed no significant difference to the control (**Table 1**).

An explanation for the fact that the level of *CaFAD2* silencing and the oil composition are not clearly correlated in T_1 generation may reside in differences in zygosity level of the transgene

copies among the T_1-plants. As the expression profile was analysed on bulks of five to ten seeds, variation in transgene copy number and the presence of seeds that have segregated to the wild-type may mask the silencing effect. Indeed, among the fatty acid profiles of the individual seeds of WG4-5 and WG19-5 segregation to the wild-type oil composition was observed (**Additional file 3**).

Table 1. Oil composition (%) in seeds of T_1 lines of *C. abyssinica.*

ID	No. of seed	FA level	Oleic acid (C18:1)	Linoleic acid (C18:2)	Linolenic acid (C18:3)	Erucic acid (C22:1)
WG4-3	18	Mean ± SD [a]	20.1±2.4**	2.7±0.9**	3.0±0.8**	65.5±3.4**
		Max	26.3	4.7	5.4	69.3
		Min	16.8	1.6	1.9	54.6
WG4-4	20	Mean ± SD	20.0±1.1**	2.3±0.5**	3.0±0.5**	67±2.5**
		Max	21.9	3.3	4.1	70.8
		Min	18.0	1.5	2.3	60.7
WG4-5	19	Mean ± SD	21.9±3.5**	2.1±1.8**	2.3±0.9**	63.5±3.2
		Max	27.4	7.6	5.4	66.6
		Min	11.5	1.0	1.5	55.5
WG19-4	19	Mean ± SD	22.6±2.3**	1.6±0.2**	2.5±0.3**	60.7±5.2
		Max	29.5	2.1	3.1	65.9
		Min	20.4	1.3	1.8	46.1
WG19-5	20	Mean ± SD	21.4±3.2**	2.7±1.8**	3.3±0.7*	60±4.6
		Max	27.1	9.5	5.3	64.1
		Min	11.3	1.3	2.3	50.2
WG19-6	20	Mean ± SD	18.0±0.9*	2.0±0.4**	2.4±0.3**	66.2±1.0**
		Max	20.0	2.9	3.0	68.0
		Min	16.4	1.4	1.9	64.5
Control	19	Mean ± SD	12.0±0.5	8.2±0.4	5.1±0.2	63.1±0.8
		Max	12.8	9.0	5.6	64.4
		Min	10.9	7.5	4.7	61.6

The fatty acid composition in single seed was measured, and for each line several seeds were used for measurement. Kruskal-Wallis test was run to determine significant level of difference between the FAD2-RNAi lines and the control. a: SD = standard deviation. Significance levels: * $P < 0.05$ and ** $P < 0.01$

CaFAD2 silencing and oil composition in T_2-generation.

Two T_1-plants (WG4-4 and WG19-6) with stable and high C18:1 content were chosen to develop a second generation for an extensive study of the inheritance of the "high C18:1, low PUFA" oil-phenotype and to examine the correlation between silencing of the respective *CaFAD2* genes, the oil composition and the expression of two other genes involved in the seed oil biosynthesis (*CaFAD3* and *CaFAE1*).

Transgene copy-number

Seventeen T_2-plants (5 from WG4-4 and 12 from WG19-6, designated as family WG4-4 and WG19-6) were randomly selected and characterised for the number of transgene insertion by Southern analysis. Four or five transgene insertions were detected in the progeny of WG19-6 with a similar pattern. In contrast, a single insertion was detected in all five analysed progeny plants of WG4-4 (**Additional file 4**).

CaFAD2 gene-silencing

To test whether and to what extent RNAi-mediated gene-silencing was effective in the T_2-plants, the expression levels of *CaFAD2-C1*, *CaFAD2-C2* and *CaFAD2-C3* in developing seeds (five to ten bulked seeds per plant, 20 DAP) were quantified for all twenty-two T_2-plants by qRT-PCR. The relative expression levels were calculated relative to the average expression level of the respective genes in the control. Collectively, down-regulation of gene expression is clearly detectable in the T_2-plants of both T_2 families, WG4-4 and WG19-6, differential patterns of silencing are observed for all three *CaFAD2* genes (**Figure 2A**). The major *CaFAD2* gene expressed in crambe seed, *CaFAD2-C3,* is strongly silenced (> 50% reduction) in all seven T_2-plants of the WG4-4 family and in 14 out of 15 T_2-plants of the WG19-6 family. Only in a single plant, WG19-6-10, minor silencing of *CaFAD2-C3* is observed (~10% reduction) (**Figure 2A**).

In contrast to the constant and strong silencing of *CaFAD2-C3*, more variation in silencing is found for *CaFAD2-C2* and *CaFAD2-C1* in both families. In the single copy transgenic T_2-plants of the WG4-4 family, *CaFAD2-C2* showed strong silencing in almost half (3 out of 7) of the WG4-4 family (WG4-4-1, WG4-4-2 and WG4-4-4) whereas in the other T_2-plants of WG4-4 this gene was not silenced (**Figure 2A**). In the other family, WG19-6, 60% of the T_2-plants (9 out of 15) showed strong silencing of *CaFAD2-C2* whereas the rest showed zero- to moderate silencing of *CaFAD2-C2.* Also, *CaFAD2-C1* showed various silencing levels in

both T_2 families with strong silencing in three WG4-4 T_2 -plants (3 out of 7; 43%) and five WG19-6 T_2-plants (5 out of 15; 33%) (**Figure 2A**).

The average expression levels of the respective genes were calculated for each T_2 family in comparison with the control. In both families (WG4-4 and WG19-6), only the expression of *CaFAD2-C3* is constantly and significantly different to the average control level (**Figure 2B**). For *CaFAD2-C1* and *CaFAD2-C2,* the average expression levels in both families are not significantly different from the average expression level observed in the control (**Figure 2B**).

Inheritance of CaFAD2 silencing

As compared over two generations (T_1 and T_2), the level of silencing in WG19-6, the line which carries multiple transgene insertions, increased from zero silencing in the T_1 to ~40%, 60% and 70% reduction for respectively *CaFAD2-C1*, *-C2* and *-C3* in the T_2. Over two generations, T_1 and T_2, of the single copy transgene WG4-4 the expression patterns of the *CaFAD2* genes were similar. In both generations, only *CaFAD2-C3* was dominantly down-regulated in expression (**Figure 1 and 2B**).

Effect of CaFAD2 silencing on CaFAD3 and CaFAE1 expression

As a consequence of *CaFAD2* gene silencing, and accompanying changes in the substrate flow in the oil biosynthesis pathway, the expression of other genes acting in the pathway may change. Therefore, the expression of two genes, *FAD3,* involved in the conversion of C18:2 to C18:3, and *FAE1*, acting in the production of very long chain fatty acids (i.e. C22:1) by chain-elongation of C18:1, were studied in the two T_2 families (WG4-4 and WG19-6) and the control. Both genes showed differential expression among the individual plants within families, but on average no significant differences were observed between the two T_2 families and the control (**Figure 2B**). Notably, *CaFAD3* and *CaFAE1* showed a similar regulation pattern because both of them were up- or down-regulated within the same plant with exception of two plants, WG19-6-11 and WG19-6-12 (**Figure 2A**).

Oil composition

The observed changes in expression pattern of the target genes resulted in significant changes in seed oil composition in 21 T_2-plants studied (7 plants derived from WG4-4 and 14 from WG19-6). In accordance with the results found in the T_1 generation, the contents of C18:1, C18:2 and C18:3 showed significant differences to the control ($P < 0.01$). However, the high C22:1 content found in the oil of two T_1-plants (WG4-4 and WG19-6) was not observed throughout the entire T_2 offspring of these plants. The C22:1 content showed significant

differences to the control in only four T_2-plants (WG19-6-3, WG19-6-5, WG19-6-7, WG19-6-8) in which C22:1 levels were about 2% higher ($P < 0.05$) than the control level (C22:1% = 62.5%). Compared to the control (C18:1% = 12.9 %), the C18:1 contents in the oil of all the T_2-plants were clearly higher ($P < 0.01$) and ranged from 24.3% to 18.2%, and most of the T_2-plants (81%, 17 out of 21) showed a C18:1 content of more than 20% (**Figure 3A** and **3C**). Consequently, the PUFA (C18:2 + C18:3) contents in these plants were lower ($P < 0.01$) than the control (PUFA% = 13.8%) and ranged from 3.4% to 7.9%, and in 67% (14 out of 21) of the plants the PUFA content was below 5% (**Figure 3B** and **3D**).

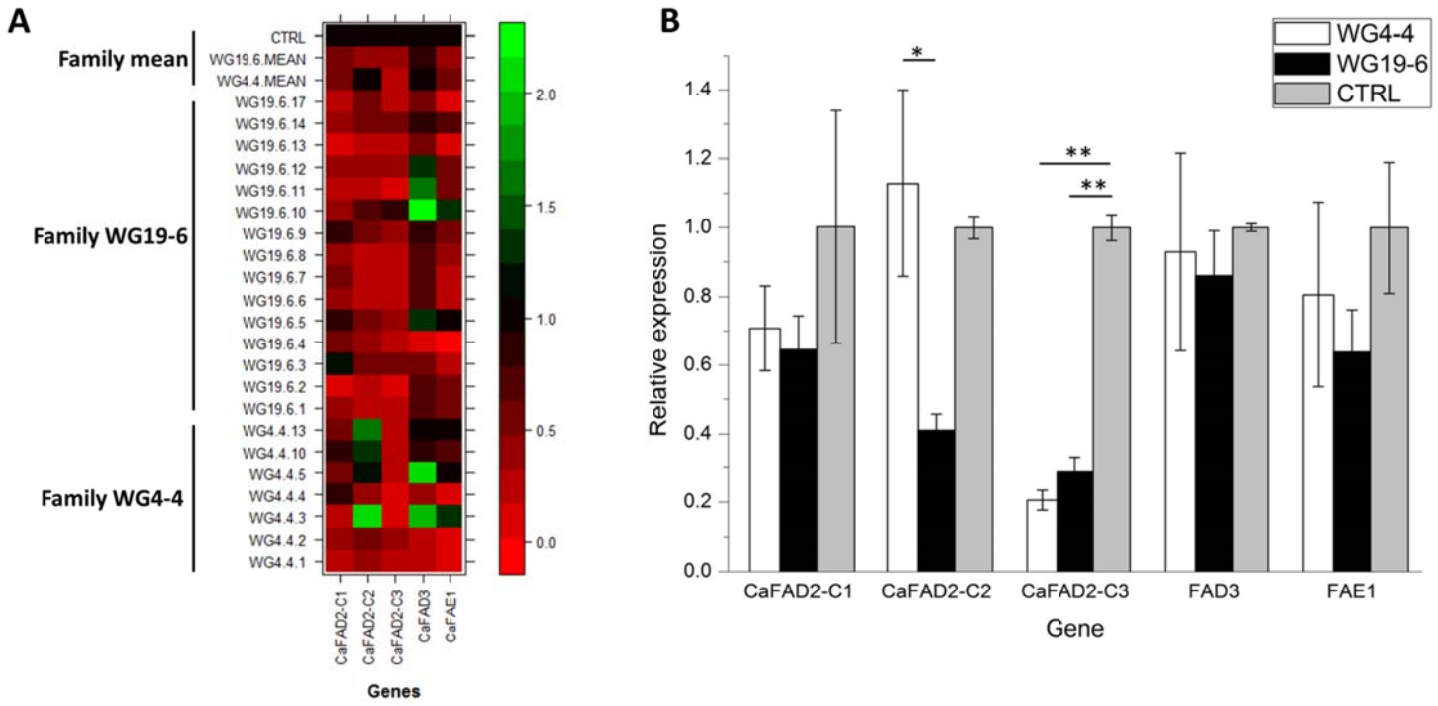


Figure 2. Relative expression levels of CaFAD2 gene family members, CaFAD3 and CaFAE1 in the seeds of T2 plants of C. abyssinica. A) Heatmap of the gene expression profiles in the seeds of T2 plants. The data indicate relative expression value of five genes (CaFAD2-C1, CaFAD2-C2, CaFAD2-C3, FAD3 and FAE1, in horizontal) in bulk seeds (five to ten) at 20 DAP compared to the control (CTRL). Measurements were carried out for 21 progenies of two families (WG4-4 and WG19-6) and two control plants. The colours represent up-regulation or down-regulation (green, up; red, down). The colour key represents the scale of expression regulation. B) Comparison of the relative expression levels among the two families and the control. The data indicate the averages of gene expression levels in the seeds of T2 plants. The comparison was run with non-parametric test Kruskal-Wallis. Only the significant differences were shown. *, $P < 0.05$; **, $P < 0.01$.

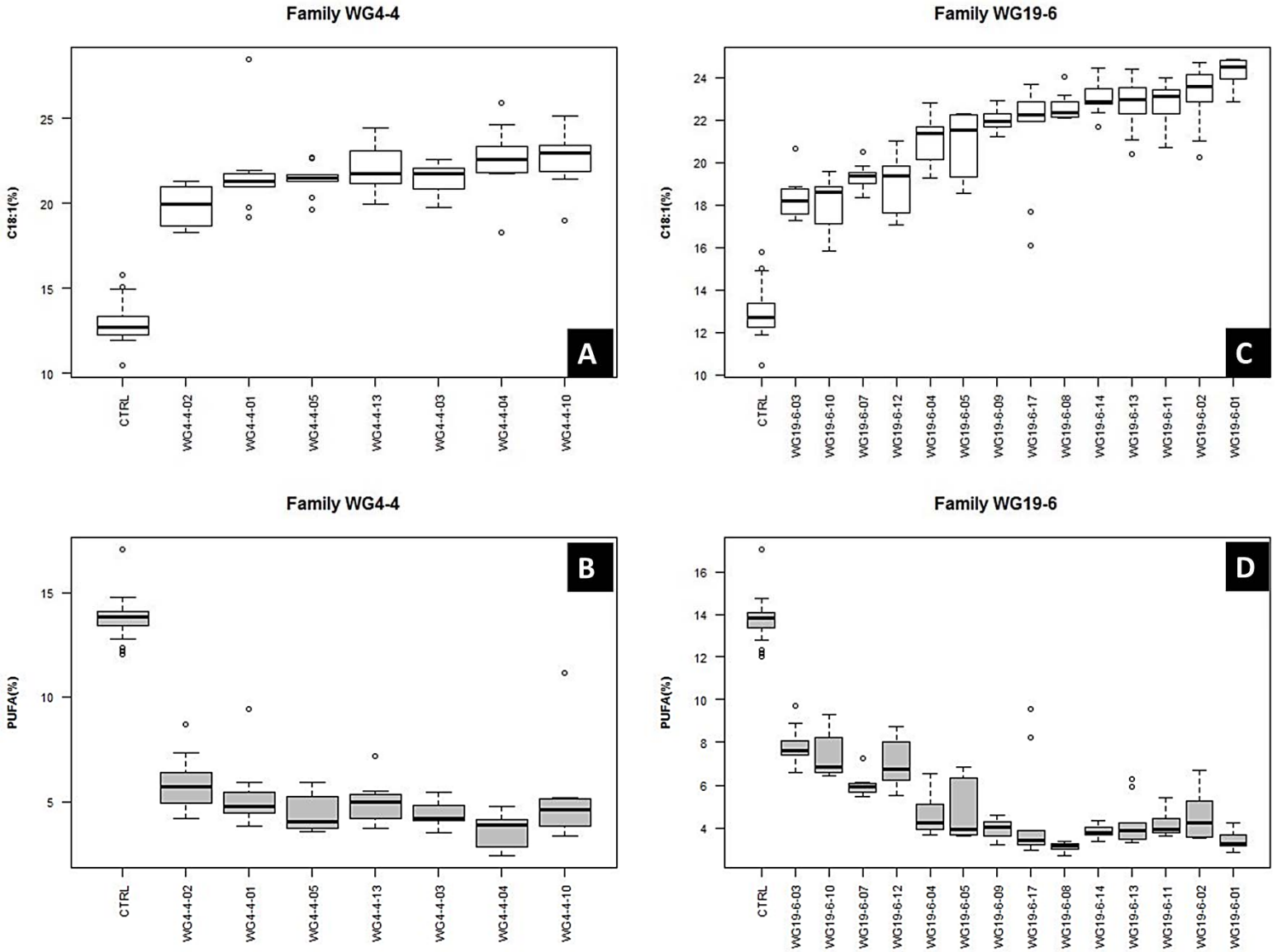


Figure 3. Boxplots of C18:1 (A) and PUFA (B) levels in the seeds of T2 plants of C. abyssinica. Each box indicates a T2 plant; ten single seeds per plant were measured.

Overall, seeds from the two T_2 families (WG4-4 and WG19-6) contained different C18:1 and PUFA contents compared to the control (higher or lower respectively) (**Figure 3**). Among both T_2 families, the content of C18:1, C18:2, C18:3, C22:1 and PUFA showed no significant difference (**Figure 4**). However, family WG19-6 showed more variation in both C18:1 and PUFA contents than family WG4-4 (**Additional file 5**).

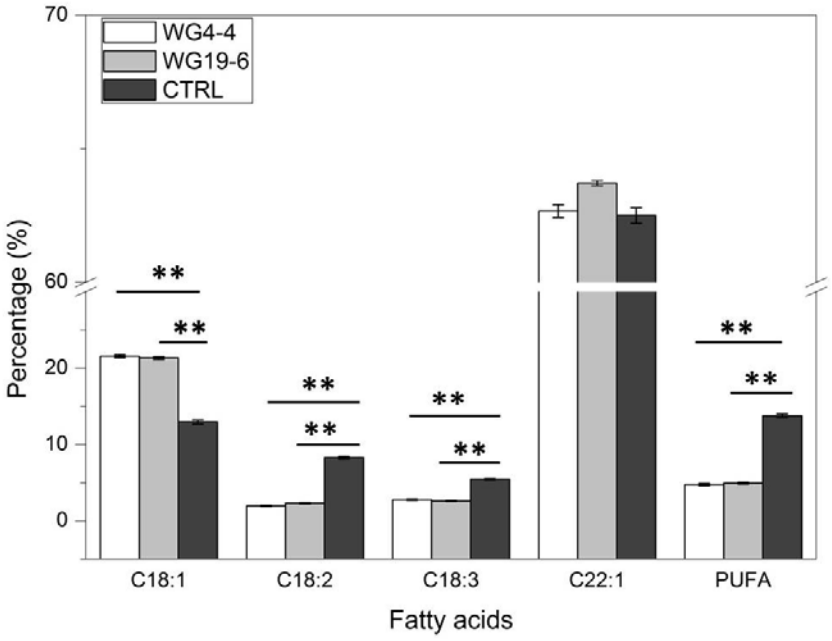


Figure 4. Comparison of the oil composition in the seeds of T2 family WG4-4, WG19-6 and control plants of C. abyssinica. The data represent the average of the two families and the control. For each plant, ten single seeds were measured. The bars indicate the comparisons with significant difference. **, $P < 0.01$. Error bar = standard error of mean.

*Correlation between gene expression and oil composition***.**

To investigate the relationship between differential expression levels of the target genes in the T_2 generation and the oil composition, correlation analysis was carried out. All the correlations and coefficients are plotted in **Figure 5**. . Among the individual T_2 plants of WG4-4 and WG19-6 that showed differential *CaFAD2* expression levels, some significant correlations were observed for the expression levels of *CaFAD2-C1* to *-C3*, *CaFAD3* and *CaFAE1*. Overall, the expression of *CaFAE1* is co-ordinately regulated with both, *CaFAD3* ($r = 0.92$, $P < 0.001$) and *CaFAD2-C2* ($r = 0.44$, $P < 0.05$) (**Figure 5**). These correlations are much stronger in the WG4-4 family ($r = 0.943$ and 0.957 respectively). Regarding the oil

composition, the contents of both C18:2 and C18:3 were negatively related (r = -0.97 and -0.95 respectively, $P < 0.001$) to the C18:1 content. However, the C22:1 content was not significantly related to the observed changes in other oil compounds ($P > 0.1$) (**Figure 5**). In accordance with its function in the oil biosynthesis pathway, the expression level of *CaFAD2-C3* is negatively related to the C18:1 content (r = -0.78, $P < 0.001$) and positively related to the contents of C18:2 (r = 0.77, $P < 0.01$) and C18:3 (r = 0.71, $P < 0.01$) (**Figure 5**). However, there were no correlations found between these fatty acids and the expression of other genes (**Figure 5**).

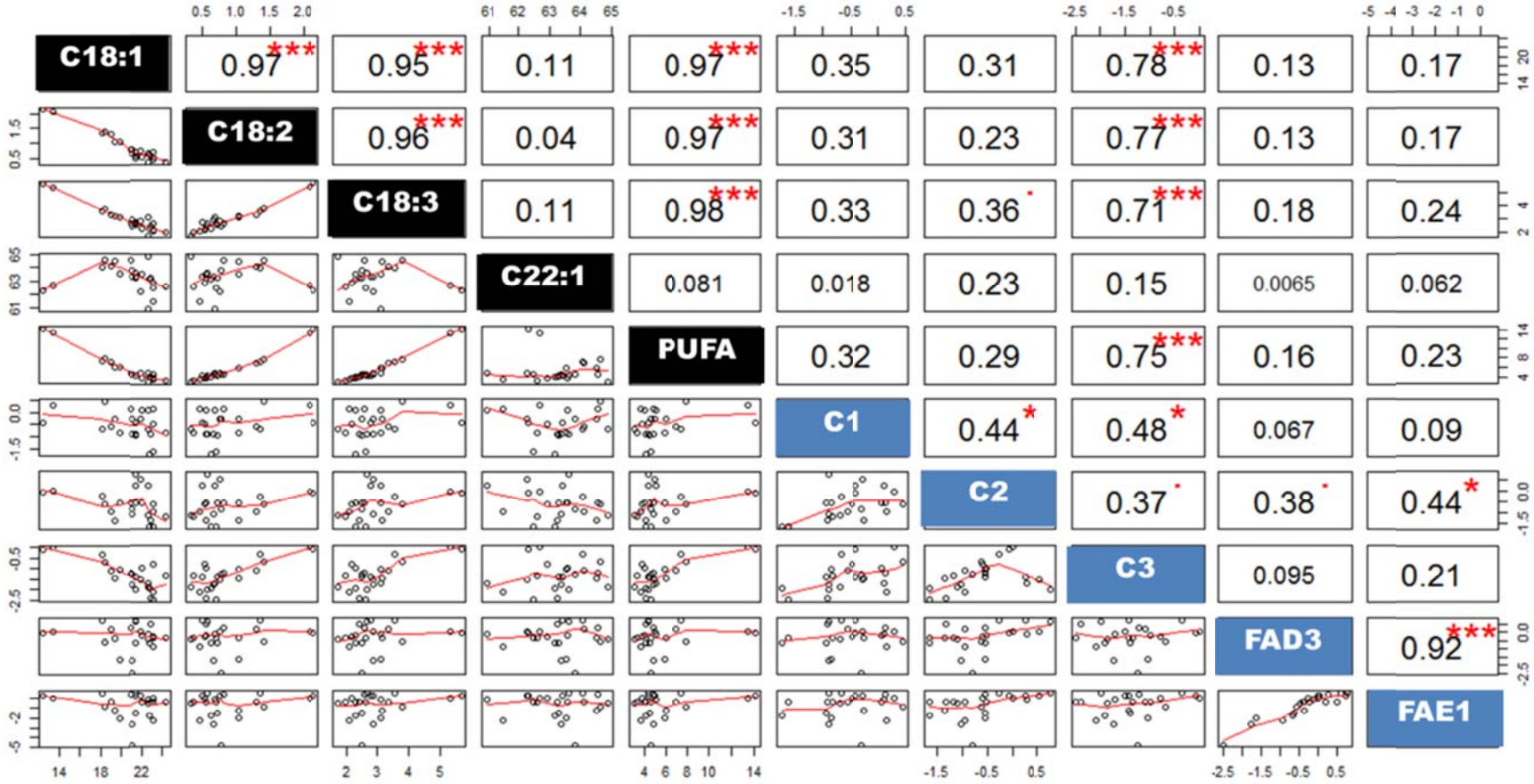


Figure 5. Correlation plotting between genes expression and oil composition of T2 plants of C. abyssinica. The expression levels and oil compositions were plotted on the lower part; and the values on the upper part represent the correlation coefficient "r". Significance levels: ".", P < 0.1; "*", P < 0.05; "**", P < 0.01; "***", P < 0.001.

Discussion

C. abyssinica, an under-utilized crop, has limited genetic variation and the improvement of crambe oil by traditional breeding has reached a bottleneck. Recently, using a genetic engineering approach, crambe lines with a desirable oil type have been obtained by silencing *CaFAD2* genes (Li et al., 2012). In *C. abyssinica*, three functional *FAD2* genes are present (*CaFAD2-C1* to -*C3*) (Cheng et al., 2013). Even though the functional defect of one of the family members may be complemented by the other members of the gene family (Lynch and Conery, 2000), one or few of these family members usually play a predominant role, so that a primary functional analysis of gene family members is normally required prior to mutation breeding (Wang et al., 2008). Our study was initiated to characterise the effect of the three functional *CaFAD2* genes (*CaFAD2-C1* to -*C3*) to determine the appropriate targets for a targeted mutation approach (TILLING) to develop non-GM crambe lines with novel oil types.

Here we used a construct containing a part of the *CaFAD2-C2* sequence, which is 96 % to 97 % identical to the sequences of other expressed *CaFAD2* genes, to obtain seed-specific silencing (*Napin* promoter) of the endogenous *CaFAD2* gene family members in crambe. The results proved that one RNAi trigger is able to affect the expression of multiple members of the gene family but in different patterns (**Figure 1**). In T_2-plants, the main *FAD2* gene expressed in crambe seeds, *CaFAD2-C3,* is strongly silenced whereas the two lower expressed genes, *CaFAD2-C2* and -*C1* (respectively 4 and 100 times lower than *CaFAD2-C3* at 20 DAP (Cheng et al., 2013), are silenced to different levels ranging from strong- to zero silencing. The reason why no reduction of *CaFAD2* gene expression was detected in two T_1-plants (WG19-5 and WG19-6), which nevertheless showed significant difference for seed-oil composition, is probably due to the heterozygous nature of the seed samples. This can cause that segregating wild-type seeds can be present in the seed bulks used for expression analysis, whereas the oil composition was measured on single seeds. Another possibility is that silencing in these lines took place in the developing seeds at other time points beyond the expression peak at 20 days after pollination (DAP). Indeed, the results of Southern analysis do suggest that the T_1-plant WG19-6 is heterozygous, because a varying number of transgene insertions was detected in its offspring (T_2), and also that WG4-4 (T_1-plant) is most likely homozygous because a single transgene insertion was observed in all randomly tested progeny (T_2) of WG4-4 (**Additional file 4**).

Over two generations (T_1 and T_2), the expression patterns of *CaFAD2* in WG4-4, carrying a single transgene insertion, were similar whereas differences in silencing levels over

generations were observed for WG19-6. The multiple transgene insertions in WG19-6 may cause different types of variation in the T_2 population, which may influence the level of silencing. It is known that trace amount of dsRNA is sufficient to trigger gene silencing (Fire, 1999) and the degree of silencing has no association with insert copy number (Li et al., 2008). Therefore, a plausible explanation to different silencing levels over T_1 and T_2 generations of WG19-6 is that a higher level of homozygosity in the seeds of the T_2 generation leads to a lower number of segregating wild type seeds, which is of importance if the seeds are analysed for expression "in bulk".

The T_2 progenies of two FAD2-RNAi plants (WG4-4 and WG19-6) were tested in detail for the level of gene expression of the different *CaFAD2* gene family members (**Figure 2A** and **2B**) and for oil composition (**Figure 3** and **4**). The results showed that the silencing of gene expression is stably inherited to the subsequent generation. The stability of RNAi-silencing was also reflected at the seed-oil composition, for instance, in the T_2-progeny of WG4-4, the C18:1 level was significantly higher and PUFA level was significantly lower than those of the control plants (**Table 1** and **Figure 3B** to **3D**). The values of important individual oil compounds are close to the average of the parental T_1-plants for both lines, WG4-4 and WG19-6, indicating that the high C18:1 and low PUFA traits were stably transmitted to the subsequent generation.

Correlation and expression regulation analysis

Compared to *CaFAD2-C1* and *CaFAD2-C2*, *CaFAD2-C3* was strong and stably silenced by the seed specific RNAi trigger. On average, silencing of *CaFAD2-C3* is constantly significant in both T_2 families (offspring of WG4-4 and WG19-6). Although some T_2-plants show a clear silencing of *CaFAD2-C1* and *CaFAD2-C2*, the overall family average of these genes are not significantly different from the control for both families. Consequently, the only significant difference in gene expression observed between the two T_2 families is the difference in *CaFAD2-C2* expression (**Figure 2B**).

Taken together with the fact that a change in oil composition was found in all T_2-plants and that out of the three *CaFAD2* genes *CaFAD2-C3* is the highest expressor in developing crambe seeds (Cheng et al., 2013), it is most likely that this gene plays a direct role in seed oil synthesis. This idea is supported by the study on the seed-oil of individual T_2-plants (**Figure 5**). Based on the differential silencing of *CaFAD2-C2* among T_2 plants of both families and the silencing of *CaFAD2-C1* observed in several individual T_2 plants, there is no clear evidence showing that the expression of these two genes is correlated to changes in the seed

oil composition (**Figure 5**). Furthermore, the changes in oil composition due to *CaFAD2-C3*-silencing were not complemented in individual T_2-plants in which *CaFAD2-C1* and *CaFAD2-C2* were unaffected and expressed to wild-type levels.

The trait of high C18:1 is controlled by quantitative trait loci and thus needs coordinated regulation of multiple loci (Guan et al., 2012). Previous studies proved that the *FAD2* gene plays an essential role for this trait (Pham et al., 2011; Yang et al., 2012). In addition, *FAD3*, *FAE1* and other loci with minor effect are likely required to establish high C18:1 pool during oil biosynthesis (Guan et al., 2012; Jagannath et al., 2011). In this study, we also investigated the regulation of *FAD3* and *FAE1* while silencing the *CaFAD2* genes. It is known that *FAD3* and *FAE1* are regulated by abscisic acid (Ruuska et al., 2002). Herein, we found that the expression of *FAE1* is positively related to that of *FAD3* and one of the *FAD2* genes, *CaFAD2-C2* (**Figure 5**). This finding is consistent to previous studies in which these functionally related genes (*FAD3* and *FAE1*) showed coordinated regulation in Arabidopsis and *Brassica napus* (Hu et al., 2009; Ruuska et al., 2002). However, it is unknown why the other two *CaFAD2* genes do not show a similar co-regulation with *FAD3* and *FAE1*.

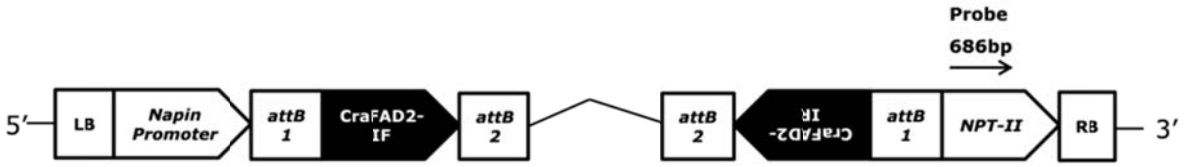
The *CaFAD2-C3* gene is a target gene for crambe oil improvement by mutation breeding

Here, we show that *CaFAD2-C3* is the main *CaFAD2*-gene involved in determining the C18:1 and PUFA contents in crambe oil (**Figure 5**). Embryo tissue is the main compartment of oil storage in crambe seed, so the oil composition in the seeds is mainly determined by that in the embryo. Therefore, *CaFAD2-C3* is most likely responsible for fatty acid synthesis in the embryo. The other two genes, *CaFAD2-C1* and *CaFAD2-C2*, might be responsible to change fatty acid contents in other seed compartments (e.g. seed coat and endosperm). For instance in olive (*Olea europaea*), two FAD2 genes (*OeFAD2-1* and *OeFAD2-2*) are expressed both in two seed compartments, seed coat and embryo. Of these two *OeFAD2* genes, the expression of *OeFAD2-2* was positively correlated to C18:2 content in the seed coat rather than in the embryo under cold condition (D'Angeli et al., 2013).

Conclusions

The finding that the prominent role of *CaFAD2-C3* is further substantiated by the observation that the effect on oil-composition caused by *CaFAD2-C3* silencing is not restored in individual T_2-plants with wild-type expression of *CaFAD2-C2* and *-C1* respectively. This finding suggests that *CaFAD2-C3* may provide an important target gene for TILLING and mutation breeding aiming at pronounced changes of C18:1 and PUFA in crambe oil.

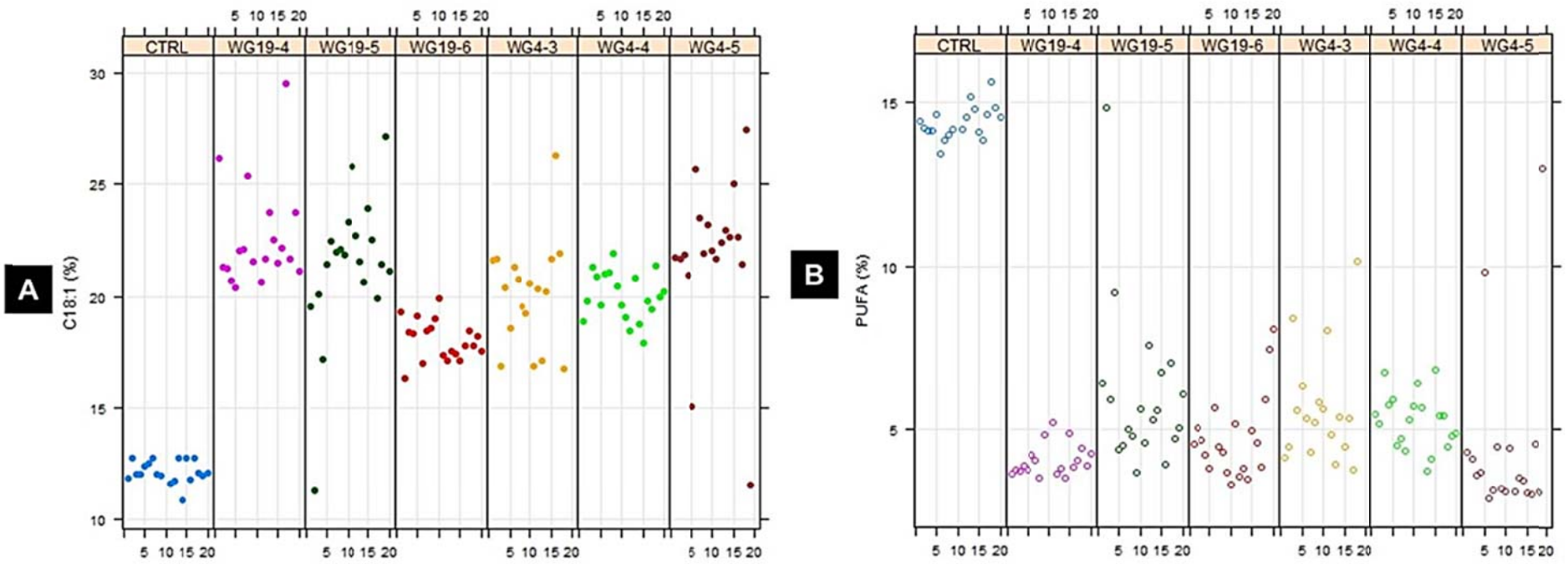
Additional files



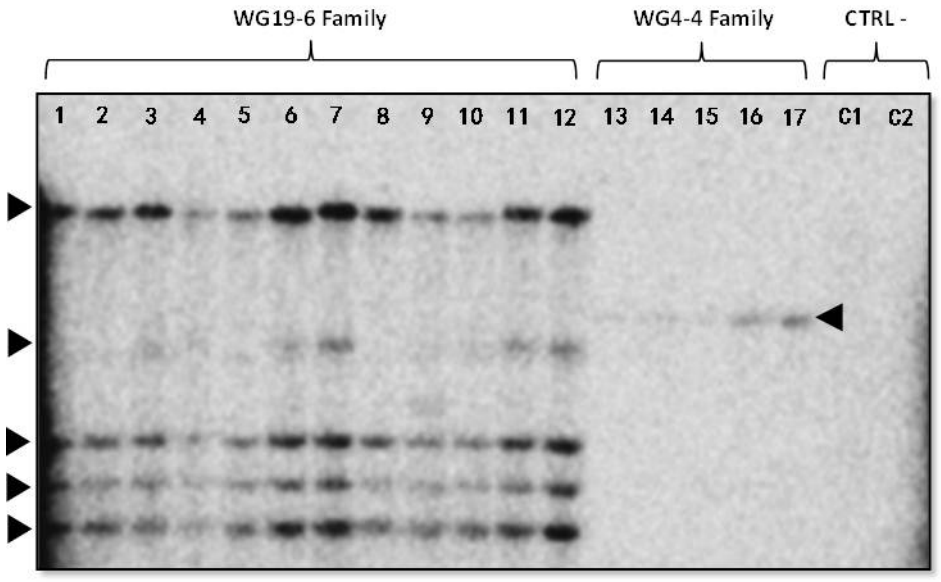
Additional file 1 as PNG. Schematic diagram of T-DNA region of the RNAi construct (not to scale). LB and RB, T-DNA left border and right border, respectively. Napin Promoter, seed specific promoter from *Brassica napus*. The attB1 and attB2, recombination sites used in BP reaction of Gateway®. CaFAD2-IF and CaFAD2-IR, 355 base pairs inverted repeats of crambe FAD2-C2 sequence in forward and reverse orientations, the sequence identities to the genes *CaFAD2-C1*, *-C2* and *-C3* are 96%, 99% and 96% respectively. *nptII*, neomycin hosphotransferase II gene. The broken line represents the sequence that forms stem in hairpin RNA. The arrow indicated npt II probe (686 base pairs) for hybridization in this study.

Category	Gene name	Forward primer	Reverse primer	Product size (bp)
qPCR	*CraFAD2-C1*	CAAGGCTGTGTCCTAAGA	CGTCGTCTCTTTCAAGCGGA	200
	CraFAD2-C2	CGCCATTCCAACACTGGCTCC	GCGAAGCCGTCGTAAGGTCT	191
	CraFAD2-C3	CAAGGCTGTGTCCTAAGC	AACACTTCATCTCTTTCCAAT	200
	FAD3	TTGTGATGTGGTTGGATGCT	TGAGGGATTTGTGGGAAGAG	196
	FAE1	GTGCTGGCGTTATAGCCATT	AGGCTTGTTGGAGAGCAAAA	191
	β-actin 2	ATTCAGATGCCCAGAAGT	CTCATACGGTCAGCGATA	180
Probe	*npt-II*	TGGGCACAACAGACAATCGGCTGC	TGCGAATCGGGAGCGGCGATACCG	686

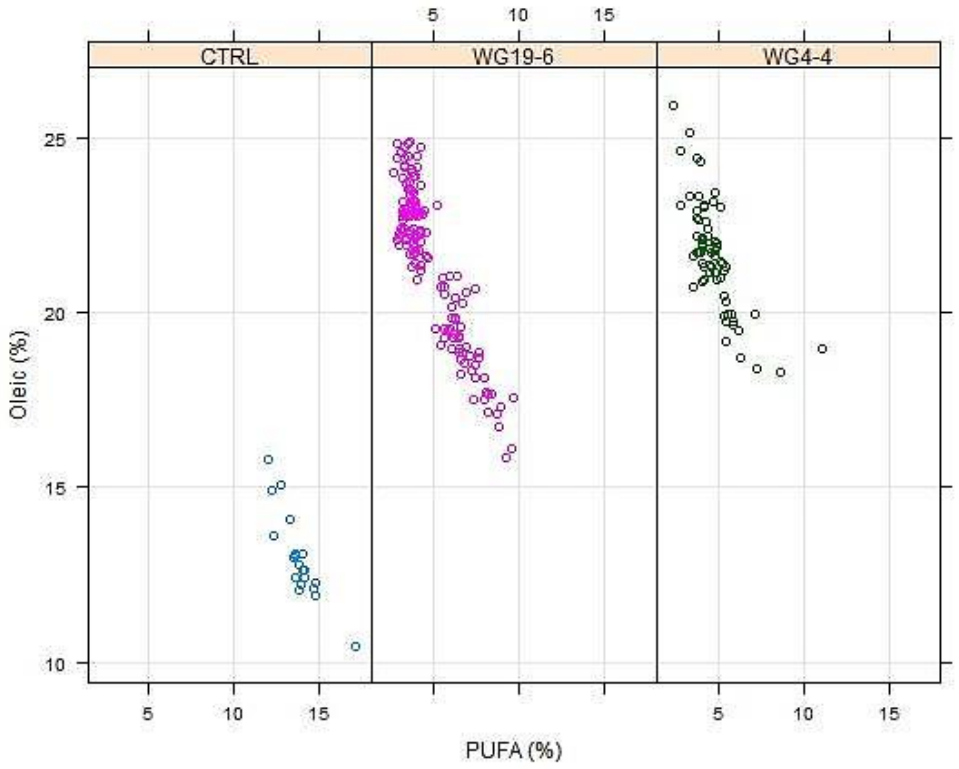
Additional file 2 as DOCX. The primers used in this study.



Additional file 3 as PNG. The levels of C18:1 (A) and PUFA (B) in the single seeds of T_1 plants. For each plant, around 20 single seeds were measured. One point represents a single seed.



Additional file 4 as JPEG. Determination of transgene copy number for T_2 plants of *C. abyssinica* by Southern analysis. The blotting membrane was hybridized with $^{[32P]}$ATP-labelled nptII probe. The family WG19-6 harbours 4 or 5 transgene insertions, and the family WG4-4 harbours a single insertion. Lane 1-12: WG19-6-17; WG19-6-14; WG19-6-13; WG19-6-11; WG19-6-10; WG19-6-9; WG19-6-8; WG19-6-7; WG19-6-5; WG19-6-4; WG19-6-2; WG19-6-1. Lane 13-17: WG4-4-13; WG4-4-10; WG4-4-5; WG4-4-3; WG4-4-1. Lane C1-C2: two wild-type controls.



Additional file 5 as JPEG. The plot of oil composition (C18:1 vs. PUFA) in the seeds of T_2 plants of family WG4-4, WG19-6 and the control (CTRL). Each point represents a single seed of T_2-plant

References

Aldrich, J. and Cullis, C.A. (1993) RAPD analysis in flax: optimization of yield and reproducibility using klen Taq 1 DNA polymerase, chelex 100, and gel purification of genomic DNA. *Plant Molecular Biology Reporter* **11**, 128-141.

Andre, C., Haslam, R.P. and Shanklin, J. (2012) Feedback regulation of plastidic acetyl-CoA carboxylase by 18: 1-acyl carrier protein in Brassica napus. *Proceedings of the National Academy of Sciences* **109**, 10107-10112.

Auld, D.L., Heikkinen, M.K., Erickson, D.A., Sernyk, J.L. and Romero, J.E. (1992) Rapeseed mutants with reduced levels of polyunsaturated fatty acids and increased levels of oleic acid. *Crop science* **32**, 657-662.

Barret, P., Delourme, R., Renard, M., Domergue, F., Lessire, R., Delseny, M. and Roscoe, T. (1998) A rapeseed FAE1 gene is linked to the E1 locus associated with variation in the content of erucic acid. *TAG Theoretical and Applied Genetics* **96**, 177-186.

Beló, A., Zheng, P., Luck, S., Shen, B., Meyer, D.J., Li, B., Tingey, S. and Rafalski, A. (2008) Whole genome scan detects an allelic variant of fad2 associated with increased oleic acid levels in maize. *Molecular Genetics and Genomics* **279**, 1-10.

Blacklock, B.J. and Jaworski, J.G. (2002) Studies into factors contributing to substrate specificity of membrane - bound 3 - ketoacyl - CoA synthases. *European Journal of Biochemistry* **269**, 4789-4798.

Carlsson, A., Clayton, D., Salentijn, E., Toonen, M., Stymne, S., Dyer, W. and Bowles, D. (2007) Oil crop platforms for industrial uses. In: *Outputs from the EPOBIO projects* p. 146. York: CPL Press, Newbury.

Chen, L., Huang, L., Min, D., Phillips, A., Wang, S., Madgwick, P.J., Parry, M.A.J. and Hu, Y.G. (2012) Development and characterization of a new TILLING population of common bread wheat (Triticum aestivum L.). *PLoS ONE* **7**, e41570.

Cheng, J., Salentijn, E., Huang, B., Krens, F., Dechesne, A., Visser, R. and van Loo, E. (2013) Isolation and characterization of the omega-6 fatty acid desaturase (FAD2) gene family in the allohexaploid oil seed crop Crambe abyssinica Hochst. *Molecular Breeding*, 517-531.

Chhikara, S., Dutta, I., Paulose, B., Jaiwal, P.K. and Dhankher, O.P. (2012) Development of an Agrobacterium-mediated stable transformation method for industrial oilseed crop Crambe abyssinica 'BelAnn'. *Industrial crops and products* **37**, 457-465.

Cuperus, F.P., Boswinkel, G. and Derksen, J.T.P. (1996) The processing of new oilseed crops-An economic evaluation. *Journal of the American Oil Chemists' Society* **73**, 1635-1640.

D'Angeli, S., Falasca, G., Matteucci, M. and Altamura, M.M. (2013) Cold perception and gene expression differ in Olea europaea seed coat and embryo during drupe cold acclimation. *New Phytologist* **197**, 123-138.

Falasca, S., Flores, N., Lamas, M., Carballo, S.M. and Anschau, A. (2010) Crambe abyssinica: An almost unknown crop with a promissory future to produce biodiesel in Argentina. *International Journal of Hydrogen Energy* **35**, 5808-5812.

Fire, A. (1999) RNA-triggered gene silencing. *Trends in Genetics* **15**, 358-363.

Gómez-Galera, S., Twyman, R.M., Sparrow, P.A.C., Van Droogenbroeck, B., Custers, R., Capell, T. and Christou, P. (2012) Field trials and tribulations-making sense of the regulations for experimental field trials of transgenic crops in Europe. *Plant biotechnology journal* **10**, 511-523.

Guan, M., Li, X. and Guan, C. (2012) Microarray analysis of differentially expressed genes between Brassica napus strains with high- and low-oleic acid contents. *Plant cell reports* **31**, 929-943.

Hu, Y., Wu, G., Cao, Y., Wu, Y., Xiao, L., Li, X. and Lu, C. (2009) Breeding response of transcript profiling in developing seeds of Brassica napus. *BMC Molecular Biology* **10**, 49.

Jadhav, A., Katavic, V., Marillia, E.F., Michael Giblin, E., Barton, D.L., Kumar, A., Sonntag, C., Babic, V., Keller, W.A. and Taylor, D.C. (2005) Increased levels of erucic acid in Brassica carinata by co-suppression and antisense repression of the endogenous FAD2 gene. *Metabolic engineering* **7**, 215-220.

Jagannath, A., Sodhi, Y.S., Gupta, V., Mukhopadhyay, A., Arumugam, N., Singh, I., Rohatgi, S., Burma, P.K., Pradhan, A.K. and Pental, D. (2011) Eliminating expression of erucic acid-encoding loci allows the identification of "hidden" QTL contributing to oil quality fractions and oil content in Brassica juncea (Indian mustard). *Theoretical and Applied Genetics* **122**, 1091-1103.

Löhden, I. and Frentzen, M. (1988) Role of plastidial acyl-acyl carrier protein: glycerol 3-phosphate acyltransferase and acyl-acyl carrier protein hydrolase in channelling the acyl flux through the prokaryotic and eukaryotic pathway. *Planta* **176**, 506-512.

Lessman, K. and Meier, V. (1972) Agronomic evaluation of crambe as a source of oil. *Crop Science* **12**, 224-227.

Li, J., Brunner, A.M., Shevchenko, O., Meilan, R., Ma, C., Skinner, J.S. and Strauss, S.H. (2008) Efficient and stable transgene suppression via RNAi in field-grown poplars. *Transgenic research* **17**, 679-694.

Li, X., Ahlman, A., Yan, X., Lindgren, H. and Zhu, L.H. (2010) Genetic transformation of the oilseed crop Crambe abyssinica. *Plant cell, tissue and organ culture* **100**, 149-156.

Li, X., Van Loo, E.N., Gruber, J., Fan, J., Guan, R., Frentzen, M., Stymne, S. and Zhu, L.H. (2012) Development of ultra-high erucic acid oil in the industrial oil crop Crambe abyssinica. *Plant biotechnology journal* **10**, 862–870.

Liu, Q., Singh, S.P. and Green, A.G. (2002) High-stearic and high-oleic cottonseed oils produced by hairpin RNA-mediated post-transcriptional gene silencing. *Plant Physiology* **129**, 1732-1743

Lynch, M. and Conery, J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151-1155.

Ma, F. and Hanna, M.A. (1999) Biodiesel production: a review. *Bioresource technology* **70**, 1-15.

Mastebroek, H., Wallenburg, S. and Van Soest, L. (1994) Variation for agronomic characteristics in crambe (Crambe abyssinica Hochst. ex Fries). *Industrial crops and products* **2**, 129-136.

McCallum, C.M., Comai, L., Greene, E.A. and Henikoff, S. (2000) Targeted screening for induced mutations. *Nature Biotechnology* **18**, 455-457.
Mietkiewska, E., Brost, J.M., Giblin, E.M., Barton, D.L. and Taylor, D.C. (2007) Cloning and functional characterization of the fatty acid elongase 1 (FAE1) gene from high erucic Crambe abyssinica cv. Prophet. *Plant biotechnology journal* **5**, 636-645.
Miquel, M. (1992) Arabidopsis mutants deficient in polyunsaturated fatty acid synthesis. Biochemical and genetic characterization of a plant oleoyl-phosphatidylcholine desaturase. *The Journal of Biological Chemistry* **267**, 1502-1509.
Ohlrogge, J. (1995) Lipid biosynthesis. *The Plant Cell* **7**, 957-970.
Ohlrogge, J.B. (1994) Design of new plant products: engineering of fatty acid metabolism. *Plant physiology* **104**, 821-826.
Paulose, B., Kandasamy, S. and Dhankher, O.P. (2010) Expression profiling of Crambe abyssinica under arsenate stress identifies genes and gene networks involved in arsenic metabolism and detoxification. *BMC plant biology* **10**, 108.
Peng, Q., Hu, Y., Wei, R., Zhang, Y., Guan, C., Ruan, Y. and Liu, C. (2010) Simultaneous silencing of FAD2 and FAE1 genes affects both oleic acid and erucic acid contents in Brassica napus seeds. *Plant cell reports* **29**, 317-325.
Pham, A.T., Lee, J.D., Shannon, J.G. and Bilyeu, K. (2010) Mutant alleles of FAD2-1A and FAD2-1B combine to produce soybeans with the high oleic acid seed oil trait. *BMC plant biology* **10**, 195.
Pham, A.T., Lee, J.D., Shannon, J.G. and Bilyeu, K.D. (2011) A novel FAD2-1 A allele in a soybean plant introduction offers an alternate means to produce soybean seed oil with 85% oleic acid content. *TAG Theoretical and Applied Genetics* **123**, 793-802.
Pham, A.T., Shannon, J.G. and Bilyeu, K.D. (2012) Combinations of mutant FAD2 and FAD3 genes to produce high oleic acid and low linolenic acid soybean oil. *TAG Theoretical and Applied Genetics* **125**, 503-515.
Rathore, K.S., Sundaram, S., Sunilkumar, G., Campbell, L.M., Puckhaber, L., Marcel, S., Palle, S.R., Stipanovic, R.D. and Wedegaertner, T.C. (2012) Ultra-low gossypol cottonseed: Generational stability of the seed-specific, RNAi-mediated phenotype and resumption of terpenoid profile following seed germination. *Plant biotechnology journal* **10**, 174-183.
RDevelopment, C.O.R.E.T.E.A.M. (2008) R: A language and environment for statistical computing. Vienna.
Rossak, M., Smith, M. and Kunst, L. (2001) Expression of the FAE1 gene and FAE1 promoter activity in developing seeds of Arabidopsis thaliana. *Plant Molecular Biology* **46**, 717-725.
Ruuska, S.A., Girke, T., Benning, C. and Ohlrogge, J.B. (2002) Contrapuntal networks of gene expression during Arabidopsis seed filling. *Plant Cell* **14**, 1191-1206.
Salas, J.J., Martínez-Force, E. and Garcés, R. (2005) Very long chain fatty acid synthesis in sunflower kernels. *Journal of agricultural and food chemistry* **53**, 2710-2716.
Schaart, J.G., Krens, F.A., Pelgrom, K.T.B., Mendes, O. and Rouwendal, G.J.A. (2004) Effective production of marker-free transgenic strawberry plants using inducible site-specific recombination and a bifunctional selectable marker gene. *Plant biotechnology journal* **2**, 233-240.
Sivaraman, I., Arumugam, N., Sodhi, Y.S., Gupta, V., Mukhopadhyay, A., Pradhan, A.K., Burma, P.K. and Pental, D. (2004) Development of high oleic and low linoleic acid transgenics in a zero erucic acid Brassica juncea L.(Indian mustard) line by antisense suppression of the fad2 gene. *Molecular Breeding* **13**, 365-375.
Slade, A.J., McGuire, C., Loeffler, D., Mullenberg, J., Skinner, W., Fazio, G., Holm, A., Brandt, K.M., Steine, M.N., Goodstal, J.F. and Knauf, V.C. (2012) Development of high amylose wheat through TILLING. *BMC plant biology* **12**, 69.
Tanhuanpää, P., Vilkki, J. and Vihinen, M. (1998) Mapping and cloning of FAD2 gene to develop allele-specific PCR for oleic acid in spring turnip rape (Brassica rapa ssp. oleifera). *Molecular Breeding* **4**, 543-550.
Wang, N., Wang, Y., Tian, F., King, G.J., Zhang, C., Long, Y., Shi, L. and Meng, J. (2008) A functional genomics resource for Brassica napus: Development of an EMS mutagenized population and discovery of FAE1 point mutations by TILLING. *New Phytologist* **180**, 751-765.
Wesley, S.V., Helliwell, C.A., Smith, N.A., Wang, M.B., Rouse, D.T., Liu, Q., Gooding, P.S., Singh, S.P., Abbott, D. and Stoutjesdijk, P.A. (2001) Construct design for efficient, effective and high - throughput gene silencing in plants. *The Plant Journal* **27**, 581-590.
Yang, Q., Fan, C., Guo, Z., Qin, J., Wu, J., Li, Q., Fu, T. and Zhou, Y. (2012) Identification of FAD2 and FAD3 genes in Brassica napus genome and development of allele-specific markers for high oleic and low linolenic acid contents. *TAG Theoretical and Applied Genetics* **125**, 715-729.
Youping, W. and Peng, L. (1998) Intergeneric hybridization between Brassica species and Crambe abyssinica. *Euphytica* **101**, 1-7.
Zanetti, F., Vamerali, T., Bona, S. and Mosca, G. (2006) Can we "cultivate" erucic acid in southern Europe? *Italian Journal of Agronomy* **1**, 3-10.

Chapter 4

High-throughput discovery of induced mutations in the allohexaploid oil crop *Crambe abyssinica* using 454 sequencing

Jihua Cheng[1,2] , Elma M.J. Salentijn[1], Bangquan Huang[2], Christel Denneboom[1], Annemarie C. Dechesne[1], Frans A. Krens[1], Weicong Qi[1,2], Richard G.F. Visser[1], Eibertus N. van Loo[1*]

[1] Wageningen UR Plant Breeding, P.O. Box 16, 6700 AA Wageningen, the Netherlands

[2] College of Life Science, Hubei University, People's Republic of China

Keywords: Brassicaceae, *Crambe abyssinica*, TILLING, *FAD2*, mutation detection, 454 sequencing

To be submitted.

Abstract

Mutagenesis is an important tool for crop improvement, but mutation detection is a challenge in polyploid species because of the presence of multiple homologs. To explore the feasibility of mutation detection in crambe by combining a TILLING (targeting induced local lesion in genome) approach with next generation sequencing technologies (NGS), a mutant population (12,480 individuals, organized in pools of ten genotypes) established by mutagenizing seeds of crambe cultivar 'Galactica' with ethyl methane sulfonate (EMS) was analysed. 454-amplicon sequencing was applied to detect and validate mutations in the omega-6 fatty acid desaturase gene family (*CaFAD2*), which is associated with the oleic acid and polyunsaturated fatty acid (PUFA) contents in plant seed oil.

Of this population, 186 pools (1860 genotypes) were initially screened for mutants. Ten pools that contain plants with mutations in *CaFAD2* were detected, and the mutation density was estimated for this mutant population (one mutation per 382 kbp). Furthermore, homozygous and heterozygous mutants for specific mutations were identified for six mutations, two silent and four missense, that occur in four respective members of *CaFAD2*. No significant difference was found in seed oil composition between the homozygous missense mutants and the segregating wild-type plants. This suggests that the specific amino-acid changes involved have neutral effects on the CaFAD2 function. Although these artificial genetic variations have not resulted in phenotypic changes, we showed that it is possible to identify genotypes with mutations in genes of interest with the mutation detection platform developed here.

Introduction

Crambe (*Crambe abyssinica* Hochst ex. R. E. Fries, 2n = 6x = 90), a non-food oil crop, belongs to the *Brassicaceae* plant family (Rudloff and Wang, 2011). The oil from crambe seeds contains a high content of erucic acid (55% to 65%, C22:1) and thus has a wide range of industrial applications such as slip-agent, nylon, coating and lubricant (Lazzeri et al., 1994). So far, the main source of high C22:1 oil is industrial rapeseed (Cuthbert et al., 2009). In comparison to industrial rapeseed (40% to 50%, C22:1), crambe seed oil contains a higher content of C22:1 (Cuthbert et al., 2009). Furthermore, crambe is not able to cross with food-oil producing *Brassicaceae* species and thus can safely be grown in the neighbourhood of such crops (Rudloff and Wang, 2011). The improvement of crambe oil value relies on a further increase of the C22:1 content and a reduction of polyunsaturated fatty acids (PUFA).

Efforts have been made by conventional breeding and genetic engineering to obtain crambe lines with high C22:1 and low PUFA in their seed oil (Laghetti et al., 1995; Li et al., 2012). However, limited natural variation is available in crambe (Mastebroek et al., 1994) which will eventually become a bottleneck for crambe breeding. Together with the fact that genetic modification is not everywhere readily accepted, this situation calls for various techniques to induce mutations and so artificially enrich genetic variation. Relevant to this, different mutagens have been applied to introduce mutations into genomes. For instance, radiation such as X-rays and neutrons have been used as mutagen, but application of these methods requires special equipment and generates relative large lesions in genomes (Belfield et al., 2012; Gustafsson et al., 1971). Alternatively, chemical mutagens have been used as they are easier to work with and cause mild effects relative to that by ionisation beams. For instance, Ethyl methane sulfonate (EMS), which mainly induces base pair transitions from G/C to A/T, has been widely used to develop mutant populations for various species (reviewed by Sikora et al., 2011).

A mutant population can undergo either phenotypic screening (forward genetics) or selection of genotypes with specific mutations (reverse genetics) (Himelblau et al., 2009; Reddy et al., 2012; Wang et al., 2008). Forward genetic screening is performed by examination of the phenotypes of population individuals, but this is more feasible for easily-detectable phenotypes. As a reverse genetic strategy, TILLING (Targeting Induced Local Lesions IN Genomes), is developed to detect specific mutations in genes of interest, which are held responsible for conferring desirable phenotypes (Slade et al., 2005). Unlike other currently

available reverse genetic strategies that all utilise genetic modification technologies, TILLING is transgene-free whereby the use of chemical mutagens causes a high density of point mutations so that it enables to target all genes in relatively few individuals (Colbert et al., 2001; Sikora et al., 2011). With this strategy, induced single nucleotide polymorphisms (SNPs) in genes of interest are detected in a DNA sample that typically is a pool of DNA from multiple individuals to increase the throughput of detection.

The development of TILLING for plant breeding has relied mainly on improvements in SNPs discovery techniques. For this purpose, several techniques have been developed and improved to reach high throughput mutation detection. Originally, high performance liquid chromatography (HPLC) was employed to detect mutations in diploid species and it is sensitive enough to detect a heterozygous mutation in 8-fold (including the DNA of 8 genotypes) DNA-pools (McCallum et al., 2000). This technique requires the use of several HPLCs to simultaneously run multiple samples, which makes it virtually impossible to detect mutations in high-throughput based on HPLC. The Li-Cor DNA analyser system is used as a high throughput means to detect mutations and has been applied in roughly 20 species (reviewed by Sikora et al., 2011). With this technique, the target gene fragment is amplified with labelled primers after which heteroduplexes are formed between a wild-type and mutant strand. These heteroduplex molecules are cleaved on the mismatch site with a specific enzyme, Cel-I, after which the resulting DNA fragments are separated by size and visualized by fluorescent labelling (e.g. on a Li-Cor system) (Colbert et al., 2001). A single Li-Cor run can be loaded with 96 samples, and the sensitivity is valid up to 16-fold pooling of a sample for diploid species (Sikora et al., 2011).

Mutation detection on a Li-Cor system is also established for polyploid species including soybean, *Brassica napus* and hexaploid wheat (Cooper et al., 2008; Dierking and Bilyeu, 2009; Slade et al., 2005; Wang et al., 2008). To avoid the use of dye-labelled primers for imaging, modified approaches were tested such as ethidium bromide based gels in hexaploid wheat (Dong et al., 2009a; Uauy et al., 2009) and MALDI-TOF (Matrix-assisted laser desorption/ionization-time-of-flight mass spectrometer) in oat (Chawade et al., 2010). However, on either Li-Cor or the modified approaches, the application of TILLING in polyploid species requires the development of allele specific primers to amplify unique fragments derived from a single homolog. This is because in polyploid species multiple highly identical homologs of genes are present (Cooper et al., 2008). In certain cases the design of homolog-specific primers is difficult and forms an obstacle for detection. To

sensitively detect mutations in multiple homoeologous genes, a method using high resolution melting (HRM) was applied in tetraploid and hexaploid wheat (Dong et al., 2009b). But, this technique relies on special equipment and software.

Next generation sequencing (NGS) technology enables the high-throughput discovery of SNPs in multiple homoeologous genes. The high sequencing depth generated by NGS technologies and relative low cost has contributed to its application on SNPs detection both, in target genes and genome-wide (Bundock et al., 2009; Gholami et al., 2012; Han et al., 2011; Reddy et al., 2012; Rigola et al., 2009; Tsai et al., 2011). In the case of non-model species, for which whole genome sequences are not available, SNP detection mostly focused on a subset of target genes rather than on the whole genome. For example, Bundock (2009) sequenced 307 amplicons covering a region of 59 kbp by 454 amplicon sequencing, to discover SNPs in mapped genes of polyploid sugarcane. A further increase of the SNP-detection throughput can be reached by implementing a two- or three- dimensional (2D or 3D) pooling strategy in which SNPs are typically detected from DNA-pools that contain overlapping samples. Tsai (2011) described the use of Illumina DNA sequencing to detect mutations for rice and wheat from both 2D- and 3D pools.

Here, to explore the feasibility to target induced mutations in *C. abyssinica*, we established an EMS mutant population and screened this population for mutations in a well characterised target gene family (*CaFAD2*) (Chapter 2, this thesis) using the NGS technology 454-amplicon sequencing.

Materials and Methods

Establishment of EMS mutant population and M_2-DNA library

The seeds of *C. abyssinica* cultivar 'Galactica' were treated with EMS for mutagenesis. To optimize EMS concentration, a serial concentration of EMS (Sigma, USA) solution (M/V= 0.025%, 0.05%, 0.1%, 0.15%, 0.2%, 0.25% and 0.3%) was used for treatment. In a 50 ml tube, crambe seeds (M_0 seeds, 7 g per concentration) were treated by EMS solution with shaking for 12 h at room temperature. Next, the seeds were washed with Milli-Q water for 4 h in dark followed by drying on filter paper. The mutated seeds (M_1 seeds) were sewed in the greenhouse (22°C, 18h light), and the germination rate was examined. The highest EMS

concentration was considered to be optimal one by which the germination rate of M_1 seeds was higher than 80%.

Using the optimal concentration of 0.3% EMS solution, 70 g of crambe cv. 'Galactica' seeds were treated (700 ml EMS solution per 100 g seeds) following the procedure described above. The resultant M_1 seeds were sown in the field and the M_2 seeds were harvested from 1,000 individual M_1 lines and then sown to developed M_2 lines. From each M_1 line, about ten M_2 lines were produced. The leaf discs with similar size (10×pooled M_2 leaves) were collected from each ten M_2 lines which are progeny of one M_1 lines, and the ripe M_3 seeds were harvested from 12,480 individual M_2 lines and stored.

The M_2-DNA library consists of 1378 M_2-DNA pools, which were extracted from 10×pooled M_2 leaves with using of KingFisher silicate magnetic beads and KingFisher ML robotic magnetic particle processor (Thermo Life Science, USA).

Amplicon library preparation and 454-sequencing

FAD2 gene was chosen as a target for mutation detection. The sequence of crambe *FAD2* gene (*CaFAD2-C3*, accession number: JX964745) was used to predict conserved amino acid domains in FAD2 protein and specific primers were developed with the web tool CODDLE (http://www.proweb.org/coddle/).

Four 454-sequencing libraries were prepared for four sequencing runs with different PCR strategies. For the 1st library, *FAD2* amplicons were amplified from 96 M_2-DNA pools with a specific primer pair (FAD2-Fw, FAD2-Rv, see **Supplementary Table S1**). The sequencing library was formed with purifying the amplicon from individual pool followed by ligating the purified amplicon with adaptors (**Supplementary Table S1**) according to the procedure described by Meyer (2008) . PCR was performed in a 40 µl of reaction containing 100 ng genomic DNA, 1×Taq buffer , 2 unit of Taq DNA polymerase (Goldstar, Eurogentec) plus 0.2 unit of Pfu DNA polymerase (Fermentas), 0.4 mM dNTPs, 0.5 mM of each of the primer. The PCR cycling conditions were 94°C for 3 min followed by 25 cycles at 94°C for 30 s, 57°C for 30 s and 72°C for 1 min, with a final step of 72°C for 5 min.

For the 2nd and 3rd libraries, *FAD2* amplicons were produced using fusion primers, which are specific primers barcoded on 5'-ends with unique ten nucleotides (10-nt) tags and universal tags (25-nt) (**Supplementary Table S1**). For each pool, one single PCR (for the 2nd library) or two independent PCRs (for the 3rd library) were performed with barcode-unique fusion

primers in 20 µl PCR reaction, so the amplicon per PCR was barcoded by two tags (in forward and reverse primers respectively). The PCR reaction contains 4µl of 5×Phusion HF reaction buffer (NEB, USA), 0.2 µl of dNTPs (10mM), 0.5 µl of forward and reverse primers (10uM), 0.1 µl of Phusion DNA polymerase (2 U/µl) and 1 µl of template DNA (100ng). The PCR cycling conditions were 98°C for 30 sec followed by 30 cycles at 98°C for 10 s, 58°C for 10 s and 72°C for 30 s, with a final step of 72°C for 5 min. The PCR products were purified from agarose gel and mixed into libraries.

For the 4th library, a nested PCR strategy was used to amplify fragments of *CaFAD2-C3* gene. First PCR was performed with specific primers (**Supplementary Table S1**) to amplify the full length of *CaFAD2-C3* gene amplicon. The PCR reaction and cycling conditions followed that of the 3rd libraries, however with performing 20 cycles. Using 1 µl aliquot of the product of the first PCR as templates, the nested PCR was then performed to produce short amplicons with fusion primers (**Supplementary Table S1**). The PCR reaction and cycling conditions were the same with that of the 3rd libraries. The products of nested PCR were purified from agarose gel and mixed into a sequencing library.

These four libraries were sequenced on forward and reverse orientations by 454 amplicon sequencing at Greenomics, Wageningen UR.

Sequence analysis

Base-called reads were trimmed for quality and barcode sorting was performed by an in-house developed programme Pattern Research (I Padioleau, EMJ Salentijn, and MJM Smulders, unpublished). In this process, the reads which have identical tags/barcodes were sorted out and these reads originated from a DNA pool per PCR, and then assembly was performed against the reference sequences (accession numbers: JX964741, JX964742, JX964743, JX964744, JX964745, JX964746 and JX964747) with Lasergene Seqman 9 (DNASTAR, USA) and SNPs reports were generated.

Only the EMS-type transitions, G/A to C/T, were considered in this study. The candidate mutation fulfils the following six criteria: i) it is present at a frequency of at least 5% with a mutant reads depth more than five; ii) it is present in the reads from both of two independent PCRs (when applicable); iii) it is present in both forward and reverse reads; iv) it is not located in polymorphic sites between crambe *FAD2* variants; v) it is not present in the ends of reads; vi) it is not located in homo-polymer regions of reads.

Mutant identification and oil composition analysis

Mutants were identified from the M_3 lines of candidate pools. Approximately ten M_3 lines per M_3-seed bag, which is from a M_2 genotype, were grown in the greenhouse. Genomic DNA (M_3-DNA pool) was extracted from pooled leaf discs of the ten M_3 lines and certain *FAD2* fragment was amplified (primers in **Supplementary Table S1**) from M_3-DNA pool. Based on the mutation site, the genotype of M_3-DNA pool was identified to be homozygote, heterozygote or wild-type by direct Sanger-sequencing the *FAD2* fragment (**Figure 6**). For the non-wild-type M_3 pool, M_3 lines or M_4 lines were analysed individually to find out homozygous and heterozygous mutants with the same approach of Sanger sequencing.

To examine the effect of mutations, the ripe seeds from mutants (homozygote and heterozygote) and wild-type were measured for oil composition with gas chromatography (GC). The procedure for oil extraction and GC analysis was described by Cheng (submitted for publication). One-way ANOVA was run to test difference of oil composition among heterozygous mutant, homozygous mutant and wild-type.

Results

Establishment of a crambe EMS-mutant population

Generally, an increase in mutation rate is accompanied by a drop in germination rate and an increased seedling starvation. To determine the most optimal (most mutations and least damaging in general as judged by seed germination rate) EMS concentration for establishment of a crambe mutant population, crambe seeds (*C. abyssinica* cv.'Galactica') were treated with serial dilutions of EMS and the seed germination rate was examined. The results showed that the germination was affected by the EMS treatment whereby the germination rate generally declined with increasing EMS concentration (**Figure 1**). When using 0.3% of EMS for treatment, high seed viability remained (**Figure 1**) and abundant non-wild phenotypes were observed in the M_1 generations. To reach an efficient mutation frequency combined with a good germination and survival rate, the highest EMS concentration (0.3%) that was tested here was chosen to mutagenize crambe seeds. A mutant population, consisting of 12,480 M_2- lines, was established and organized in 1,248 pools of ten M_2 plants each. That the mutation induction was successful was judged by the appearance of presence of phenotypic mutants in the M_2 and M_3 generations (**Figure 2**).

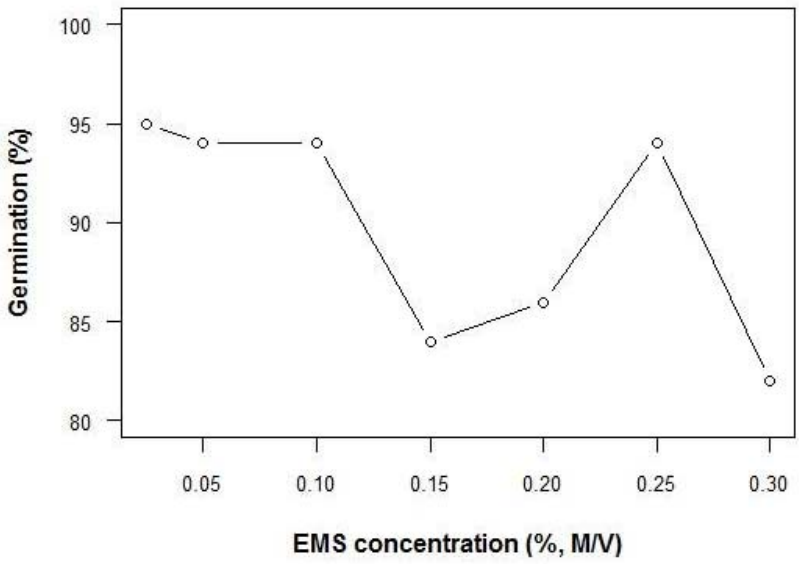


Figure 1. Percentage of seed germination after EMS treatment on crambe cv. 'Galactica'. One thousand seeds (7 g) were treated per concentration.

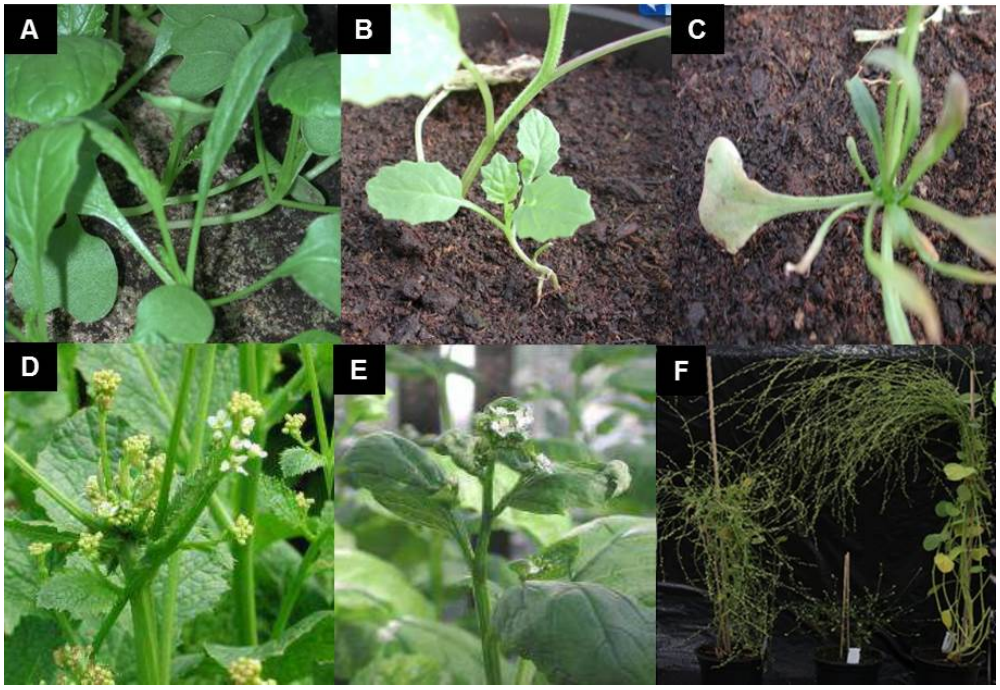


Figure 2. The phenotypes observed in the M2 and M3 crambe lines. Phenotypes in the M2 lines: (A) Seedling with bushy leaves; (B) Dwarfed plant; (C) Abnormal crambe leaves. Phenotypes in the M3 lines: (D) Bushy plants in flowering; (E); (F) Dwarfed plant, wild-type plant on the right.

Detection of EMS-mutations in *CaFAD2* by 454 sequencing

The hexaploid genome of *Crambe abyssinica* (2n = 6x = 90) consists of three homoeologous sub-genomes. This means that for each genomic position, three homoeo-loci are present. To detect genetic variation in this complex genome a high throughput detection platform is needed. Here, we used 454 amplicon sequencing to detect EMS-mutations in the *CaFAD2* genes. The *CaFAD2* gene-family consists of six to seven over 89% homologous genes of which *CaFAD2-C3* is the main candidate for oil-composition. The CaFAD2-C3 protein contains six conserved amino acid domains (**Supplementary Figure S1**). Therefore, a pair of primers (**Supplementary Table S1**) was designed to amplify an amplicon of 727 base pairs covering all the conserved regions. In a pilot study to test the critical sequence depth, this amplicon was amplified from 96 M_2 DNA pools (960 genotypes) and sequenced in bi-orientations in a quarter 454- run (**Table 1**, run 1). The run yielded over 50,000 sequences derived from five different *CaFAD2* genes (-C1, -C2, -C3, -A and -B), labelled by pool and with an average length of 230 base-pairs each, that covered the unique sequence length present in the amplicon by 1.6 times on average (**Table 1**). Analysis of the sequences from this 454 sequencing run yielded many candidate mutations (data not shown).

To validate these, a re-sequencing run (**Table 1,** run 2) was performed using three short amplicons covering the candidate positions. Due to the shorter amplicon sizes and higher sequence yield, the coverage increased to 15 times (**Table 1,** run 2). However, none of the candidate mutations was confirmed (data not shown), suggesting that false-positive mutations were detected in the first run. Instead, due to the higher coverage in the second run, five new and probably more reliable candidate mutations (Mt. 1, 3, 4, 5, and 7) were found (**Table 2**). In pool '918', a missense mutation (Mt. 7) was detected in the *CaFAD2-C3* gene, which has been considered to be an important gene for oil biosynthesis in crambe. The remaining four mutations locate in three pools, of which the genotypes of pool '69b' harbour two mutations in the *CaFAD2-C2* gene, a silent and a missense mutation respectively (Mt.4 and Mt.5, **Table 2**). Meanwhile, the silent mutation in pool '69b' was also detected in pool '69a' (referred to as Mt. 3 in **Table 2**). The probable reason for this is that these two pools, '69a' and '69b', have the same origin, a single M_1 plant '69', so they may harbour the same mutations. In addition, a mutation (Mt. 1) was found in a pseudo -gene, *CaFAD2-A* (**Table 2**).

To validate these five mutations and simultaneously detect new mutations, a next 454-run (run 3) was carried out. Now, two independent PCRs were performed, so that mutations detected in a PCR are able to be validated in another independent PCR to avoid detection of polymorphisms introduced by PCR. This 454-run yielded a similar amount of 454-reads

compared to run 2 and, because one amplicon less was involved, the coverage (21×) was higher in this run (**Table 1**). As expected, the five candidate mutations detected in the run 2 were confirmed (**Table 2**). Furthermore, two new mutations, respectively a missense (Mt. 2) and a silent (Mt. 6) mutation, were observed and validated in two independent PCRs (**Table 2**). A new 454 sequencing library (for run 4) was prepared for mutation detection in the *CaFAD2-C3* gene. As expected, based on the increased number of pools in this run (60 pools), the coverage of this run was reduced by half compared to run 3 (**Table 1**). Eventually, in this run another three mutations (Mt. 8, 9 and 10), a silent and two missense mutations, were found in the *CaFAD2-C3* gene (**Table 2**).

Table 1. Summary of the 454 runs.

Seq-run	Pool (N)	Amplicon (N)	Amplicon_1 (bp)	Amplicon_2 (bp)	Amplicon_3 (bp)	PCR (N)	Bases (Mb) [c]	Reads (N)	Calculated Coverage† (N)	Reads Length Average (bp)	Mutations (N)
1	96	1	722	/	/	1	11,600,000	50,424	1.63	230	0
2	36[a]	3	154	435	446	1	55,348,808	192,869	14.85	287	5
3	34	2	393	374	/	2	55,039,739	174,751	21.10	353	7[b]
4	60	2	393	374	/	2	57,779,162	125,539	12.56	460	3

[a] These 36 pools are a selection out of the pools in run 1. [b] Out of these seven mutations, five mutations detected in run 2 were confirmed and two new mutations were detected and confirmed in this run. [C] Bases=Total sequence length obtained. † Coverage= Bases (bp) / (5 *CaFAD2* genes × 2 alleles × 10 lines × number of pools × amplicon size). "/" = No corresponded amplicon

Table 2. FAD2 mutations detected with 454-amplicon sequencing.

No.	Run	Pool	Gene	PCR	Type	MT reads (Fw)[a]	MT reads (Rv)[a]	WT reads	Freq. (%)	Amino acid change	M3 genotypes[b]
1	2&3†	617	FAD2-A	1	G to A	17	19	151	19.3	/	Homo and Hete
				2		11	12	78	22.8		
2	3	182	FAD2-C1	1	G to A	32	42	501	12.9	Asp to Asn	Homo and Hete
				2		36	40	390	16.4		
3	2&3†	69a	FAD2-C2	1	C to T	36	39	1162	6.1	Ala=	ND
				2		5	4	92	8.9		
4	2&3†	69b	FAD2-C2	1	C to T	35	31	365	15.4	Ala=	Homo and hete
						23	15	238	13.8		
5	2&3†	69b	FAD2-C2	1	C to T	44	58	300	25.4	Leu to Phe	Hete
				2		42	49	184	33.1		
6	3	149	FAD2-C2	1	C to T	4	7	142	7.7	Ser=	ND
				2		8	8	103	13.4		
7	2&3†	918	FAD2-C3	1	C to T	37	56	872	9.6	Leu to Phe	Hete[c]
				2		43	57	1030	8.8		
8	4	960	FAD2-C3	1	C to T	16	11	86	23.9	Asp=	ND
				2		9	13	91	19.5		
9	4	1160	FAD2-C3	1	G to A	21	20	485	7.8	Ala to Thr	ND
				2		87	60	1079	12.0		
10	4	1190	FAD2-C3	1	C to T	69	53	1337	8.4	Pro to Leu	Homo
				2		51	49	1077	8.5		

Note: "/" = non-coding . [a]Number of reads in forward and reverse orientations. [b]Homo and Hete indicate homozygote and heterozygote respectively. [c]Homozygotes were identified in the M_4 lines. †: The data of the run 2&3 are represented by PCR 1 and PCR 2 respectively. ND= not determined.

Table 3 Mutation frequency calculation in the crambe EMS population.

Amplicon[a]	Size (bp, primers excluded)	Lines with enough reads (N)[b]	Alleles (N)	Unique sequence length (bp)[c]	No. of Mt. (Gene ^{Pool})	Mt Freq. (Kb per mutation)
Run2-1-fad2-c1c2c3	112	110	3 × 2 = 6	112 × 110 × 6=739,200	1 ($c3^{918}$)	739
Run2-2-fad2-c1c2c3	395	130	3 × 2 = 6	395 × 130 × 6=308,100	1 (a^{617})	308
Run2-3-fad2-c1c2	396	160	2 × 2 = 4	396 × 160 × 4=253,440	3 ($c2^{69b,3\times}$)	84
Run3-1-fad2-c1c2c3	353	230	3 × 2 = 6	353 × 230 × 6=324,760	1 ($c2^{149}$)	325
Run3-2-fad2-c1c2c3	335	290	3 × 2 = 6	335 × 290 × 6=582,900	1 ($c1^{182}$)	583
Run4-1-fad2-c3	353	530	1 × 2 = 2	353 × 530 × 2=374,180	1 ($c3^{960}$)	374
Run4-2-fad2-c3	335	300	1 × 2 = 2	335 × 300 × 2=201,000	2 ($c3^{1160}$, $c3^{1190}$)	100
					average	382

[a] The primers for the amplicons are specific to different gene/genes, so the sequence reads produced in these amplicons are from different crambe FAD2 genes.

[b] The pools with at least 100 reads per gene were considered.

[c] Unique sequence length = (amplicon size) × (No. of lines) × (No. of *FAD2* alleles)

Mutation frequency of the population

So far, ten mutations were detected in a total of 186 M_2 DNA-pools using 454-sequencing (**Table 2**). Based on this finding, the mutation frequency in this population was estimated to be one mutation per 380 kb (see calculation in **Table 3**). This frequency is comparable to the mutation density of the population of tomato, which was mutagenized by 1.0% of EMS (Minoia et al., 2010) (**Supplementary Table S2**). The mutation frequency in the resultant crambe population is lower compared to the mutation frequency observed in mutant populations of other hexaploid species, for which higher EMS concentrations (0.5% to 1.2%) were used, compared to 0.3% in this study (**Figure 3 and Supplementary Table S2**).

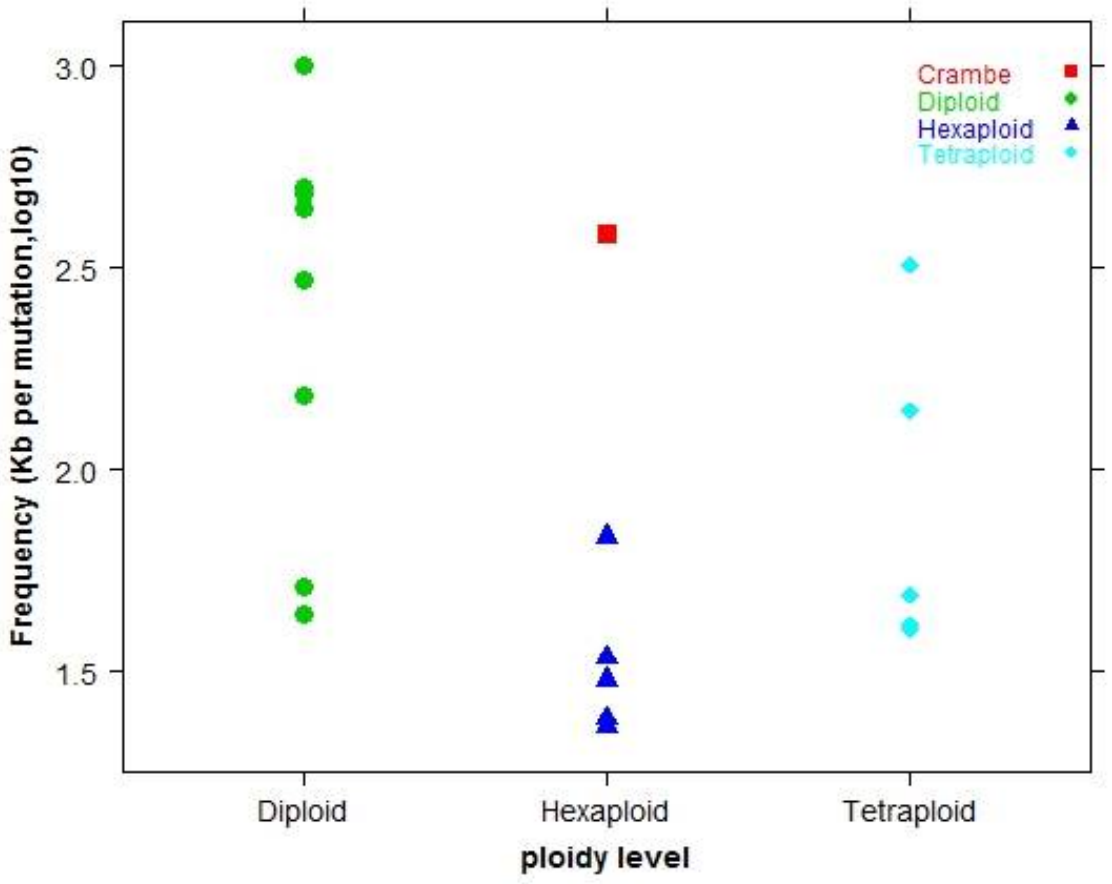


Figure 3. The mutation frequency in crambe EMS population and other published TILLING populations. This figure is produced with the data listed in Supplementary Table S2.

Selection and characterization of CaFAD2 mutant genotypes

Six mutations in the *CaFAD2* genes (*CaFAD2-A*, *-C1*, *-C2* and *-C3*) were chosen for mutant identification. These six mutations include four missense mutations that are introducing

amino acid changes in the protein; of them two in *CaFAD2-C3* and two others in *CaFAD2-C1* and -*C2* respectively. The remaining two are silent mutations that do not result in changes to the amino acid sequence of the protein. To identify the mutants among the ten M_2 genotypes of a M_2-pool, at least ten M_3 lines were grown for each M_2 genotype, so that over a hundred M_3 lines were produced for a candidate M_2-pool (10×10 =100). Because of the possibility that the mutation is heterozygous and segregates in the M_3-progeny, the DNA of ten M_3 lines in a family was pooled and analysed by Sanger sequencing for the predicted mutations. Following this, M_3 lines from a certain M_2 family, which contains the predicted mutation, were analysed individually by Sanger sequencing. In this way, all six mutations were traced back to single M_3-lines and heterozygous and homozygous M_3 lines for the specific mutations were identified (**Figure 4**). For a mutation in *CaFAD2-C3* (Mt. 10), only homozygous M_3-lines were identified. In contrast, only heterozygous M_3-lines were identified for the other two mutations, Mt. 5 and Mt. 7 in *CaFAD2-C2* and *CaFAD2-C3* respectively (**Figure 4**). To further find out the homozygotes of the *CaFAD2-C3* mutant (Mt. 7), 29 M_4 lines were developed and individually identified for the mutation site. The genotypes of these 29 M_4 lines segregated into 6 homozygotes, 15 heterozygotes and 8 wild-types, and this segregation corresponds to a ratio of 1:2:1 (Chi-square test $\chi^2 = 0.333$, $P = 0.846$).

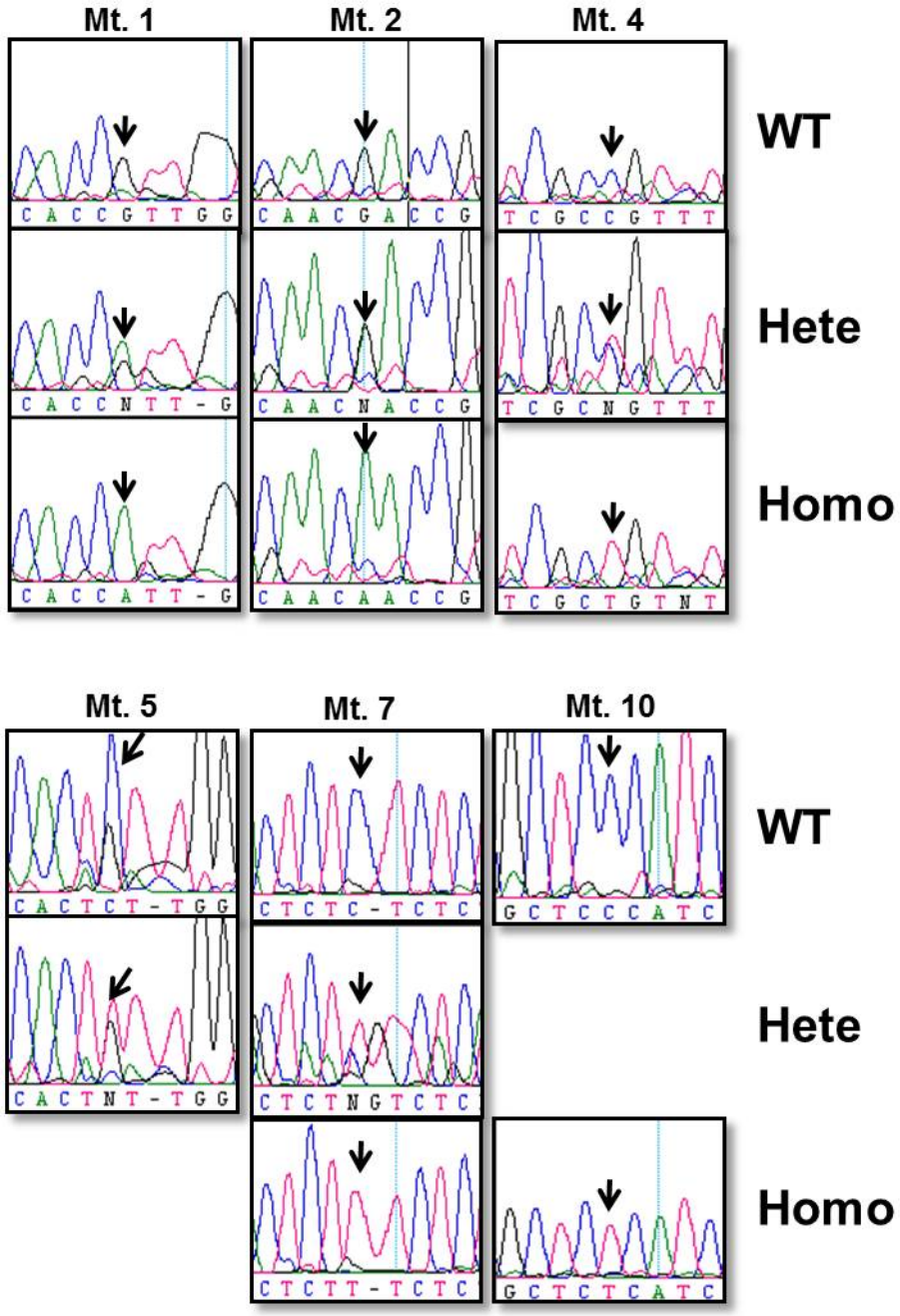


Figure 4. Genotyping the candiate M3 or M4 lines by Sanger sequencing. The DNA sequence traces were from sequencing the M3 lines of candidate mutations with exception of Mt. 7 (the M4 lines were sequenced for Mt.7). The FAD2 fragments were sequenced by Sanger sequencing, and the presence of double peaks (read as "N") indicates heterozygous nature of the line. WT = wild-type, Hete = heterozygote, Homo = homozygote.

The four identified missense mutations result in specific amino acid changes in the CaFAD2 protein (**Supplementary Figure S2**). Overall, none of the amino acid changes are located in the conserved histidine domains. Only a single mutation in *CaFAD2-C3* (Mt. 10) results in a change of a conserved amino acid residue (from Proline to Leucine) (**Supplementary Figure S2**). So, except for this mutation (Mt.10), the detected missense mutations are speculated not to affect the function of CaFAD2 enzyme. Examination of the oil composition in the ripe seeds of M_3-lines showed no significant difference in oil compositions between the mutant- and wild-type seeds for all mutants, including Mt.10 (data not shown).

Targeted mutation detection platform

Summarizing, the criteria for targeted mutation detection in crambe by 454 amplicon sequencing are: 1) coverage in a sample of at least 14 times of the unique sequence length, 2) including DNA of two independent PCRs which are separately labelled and 3) Mutation should be present in both forward and reverse reads.

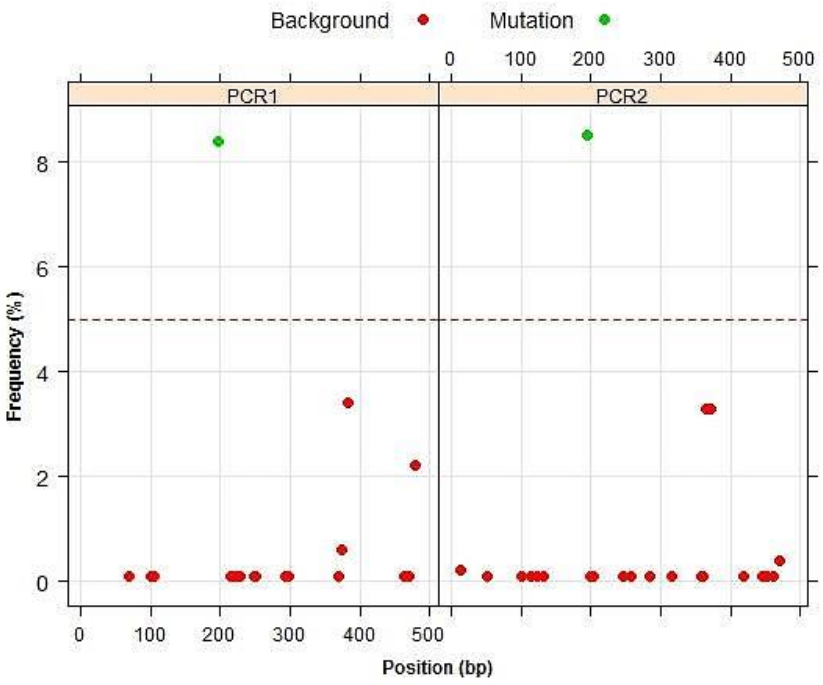


Figure 5. The frequency of all EMS-type transition (G/C to A/T) in pool '1190'. The reads from two independent PCRs were analysed separately. The frequencies of all mutant reads (G/C to A/T) in this pool were ploted against the positions of the contigs. The dotted line (red) is a set threshold (5%) of mutant read frequency in sequence analysis. The red points below this threshold line indicate the background of mutation detection. The green points represent a confirmed mutation detected in two independent PCRs.

For example, as shown in **Figure 5**, the candidate mutation (Mt. 10) was validated in two independent PCRs. Combined with the procedure of mutant identification, a platform for targeted mutation detection by 454 amplicon sequencing has thus now been established for *C. abyssinica* (**Figure 6**).

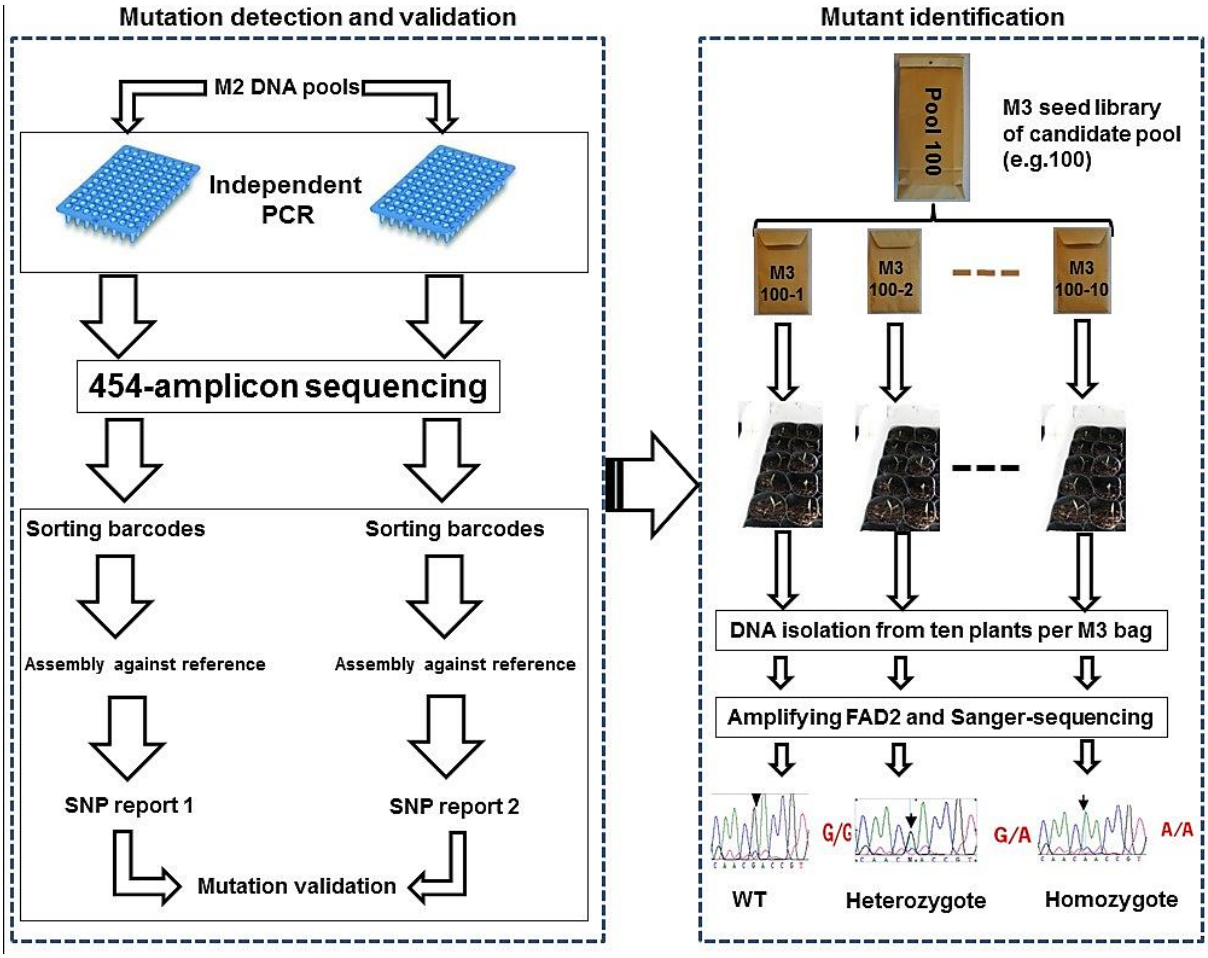


Figure 6. The TILLING platform by 454-amplicon sequencing and mutant identification for C. abyssinica.

Discussion

Establishment of the crambe population

To enrich genetic variation in *C. abyssinica*, we established an EMS-induced mutant population for *C. abyssinica*. The favourite to establish such a population is to introduce as many mutations as possible into a genome without introducing too many background mutations, which would lead to negative effects such as reduced fertility, seed germination, plant vigour (Sikora et al., 2011). So, it is important to optimize the mutagen concentration in association with the mutation density. A higher density of EMS-mutations can be reached in

hexaploid genomes compared to tetraploid and diploid genomes (**Figure 3** and S**upplementary Table S1**). Here, an EMS concentration (0.3%) with a minor effect on seed germination was used. However, higher EMS concentrations were not tested and most likely crambe, being a hexaploid, may resist higher EMS concentrations before seed survival drops. Compared to other hexaploids for which higher EMS concentrations were used, the mutation density found in hexaploid crambe in this work is low and more similar to that obtained in diploid species (**Figure3**).

A saturated mutant population should harbour at least one mutation in every allele of each gene somewhere in the population (Menda et al., 2004). The population size that is required to reach saturation depends on both, genome size and mutation density (Sikora et al., 2011). In case of a population of hexaploid wheat (genome size 17 Gb), it was estimated that a minimum of 5,000 individuals are needed to obtain a saturated mutant population for hexaploid wheat when the mutation frequency is 1:25 kb (Sikora et al., 2011). The mutation frequency in our crambe population (1:382 kb) is 15 times lower than this 1:25 kb in the wheat population (**Table 3**). Although the genome size of *C. abyssinica* is not exactly known, it can be approximately estimated to be 3.5 Gb based on the basis of its 2C-value (= 7.04 pg, (Marie and Brown, 1993)) according to the formula raised by Doležel (2003). So, considering that a hexaploid crambe genome is probably five times smaller than the hexaploid wheat genome (17 Gb), it is speculated that, with a mutation frequency of 1:382 kb, the required population size in crambe is 15,280 individuals (5000 (individuals) × (382 kb/25 kb)/ 5 (genome size ratio) = 15,280) . Therefore, our population (12,480 individuals) is expected to harbour a mutation in approximate 80% of all genes, for all alleles. We detected ten mutations in the *CaFAD2* genes in 186 pools (1,860 genotypes), indicating that the whole population could harbour about 67 mutations in the *CaFAD2* genes (12,480×10/1860 = 67).

Development of a high-throughput TILLING platform for *C. abyssinica*

To detect mutations in this population, we set up a high-throughput approach for mutation detection and validation by 454-amplicon sequencing (**Figure 6**). This approach is able to find mutations in a ten-time genotype pool. Although other high-throughput detection approaches, e.g. Li-Cor and high resolution melting (HRM) are also able to detect mutations in a genotype pool, they are conditionally applied in polyploid species so that mutations were mostly detected from individual lines without pooling until the introduction of next generation

sequencing (Gholami et al., 2012; Han et al., 2011; Slade et al., 2005; Tsai et al., 2011; Wang et al., 2008).

To investigate the feasibility of mutation detection by next generation sequencing in hexaploid crambe, we tested various PCR strategies for amplicon preparation combining with 454-amplicon sequencing. Initially, a 454-run was set up to screen mutations in a large size of FAD2 amplicon (727 bp). In this run, the coverage was low (1.6×) and many false mutations were detected, which can be attributed to the background of sequence errors in 454 sequencing (error rate) and mistakes of PCR reactions (PCR single base substitutions, PCR chimeric) (Quince et al., 2011). The error of 454 sequencing ranges from 1% to 50% and is not randomly distributed depending on some factors such as sequence size and the presence of homopolymer (Gilles et al., 2011). Considering that mismatches are possibly made by PCR, an independent PCR strategy was alternatively employed for amplicon preparation to eliminate PCR-made mistakes. More importantly, it enables to validate the detected mutations with two independent PCRs each other (**Figure 5**).

A coverage of 14 times of the unique sequence length within a sample shows to be enough for a reliable detection of mutations using 454 sequencing. Furthermore, mutation in short amplicon can be found in both forward and reverse reads and the dual detection in both directions guaranteed not to consider sequencing errors.

The core of mutation detection is performing a good sequence analysis, which discriminates real mutations from false mutations and natural variations. For this purpose, a list of criteria was used to evaluate the reliability of the detected mutations (see **Materials and Methods**). The most important criterion is the allele frequency of a mutant allele in a sample which is determined by the composition of a sample (number of lines) and the number of identical alleles that are amplified for a gene. In this work, each sample consists of a DNA-pool of ten crambe M_2-lines. For a pool which contains a single heterozygous mutant, the mutant allele frequency in a sample is expected to be 5% (1 mutated allele per line / (2 alleles per line×10 lines) = 5%).

Characterization of FAD2 mutants

The *CaFAD2* gene was chosen to be a target because FAD2 enzyme plays a role in seed oil biosynthesis by catalysing oleic acid (C18:1) to polyunsaturated fatty acids (PUFA). Therefore, mutagenizing this gene in plants enables to achieve high C18:1 and low PUFA

content in plant seeds oil (Pham et al., 2010; Pham et al., 2011). In this work, we identified the mutants in four crambe *FAD2* genes, *CaFAD2-A*, -*C1*, -*C2* and -*C3* (**Table 2**). The *CaFAD2-A* gene is not expressed in crambe seeds and it encodes a truncated FAD2 protein (Chapter 2, this thesis), so it is reasonable that the mutant in this gene performed the similar oil composition to the wild-type in the seeds oil. The missense mutants are located in the three active *CaFAD2-C* genes (-*C1* to -*C3*) but also showed no significant difference to the wild-type in oil composition in the seeds oil (data not shown). Therefore, it is concluded that the amino acid substitutions in these mutants have neutral effects on the function of FAD2 enzyme. An explanation is that these mutations neither locate in conserved amino acid residues nor in function-essential domains (**Supplementary Figure S2**). Another explanation might be the functional redundancy of genes in polyploid, were often more than one locus is present and homeologous genes may take over the function of the mutated gene (Comai, 2005). However, functional silencing (RNAi) of *CaFAD2* showed that *CaFAD2-C3* is the main gene responsible for oil-composition and that the functional *CaFAD2-C1* and *CaFAD2-C2* genes are not capable to restore the CaFAD2-C3 function completely (Chapter 3, this thesis).

In general, neutral mutations have more chance to be detected because they are easily maintained in a population for a long period of time (Charlesworth et al., 1993). In nature, selection against deleterious alleles by mutation leads to a low rate of this type of mutations (Jiang et al., 2010). Taken together with that narrow natural variation in *C. abyssinica*, mutagenesis is required to artificially increase the amount of deleterious alleles. Therefore, the establishment of the crambe mutant population enriched genetic variation; meanwhile the development of a high-throughput platform is a foundation for material selection for crambe breeding.

Supporting information

Table S1.The primers used in this study.

Category	Run	Amplicon	Forward name	Reverse name	Fw sequence (5'-3')	Rv sequence (5'-3')
454 amplicon preparation	1		FAD2-Fw	FAD2-Rv	CTCCTCCCTCACCCTCTCTC	ACGTGCGTGTCCGTGATGT
	2	1	Re-seq AM1-F	Re-seq AM1-R	TTTCTACTACATCGCCACCAAC	ATTGGTAGTCGCTGAAAGCAT
	2	2	Re-seq AM2-F	Re-seq AM2-R	TCACCCTCTCTCTTACTTAGCTT	AGACGTTCACGGTCGTTG
	2	3	Re-seq AM3-F	Re-seq AM3-R	AGAAGAAATCAGACATCAAGTGGTAC	GAAGACCTTGTTCAACATCCCATA
	3 & 4	1	FAD2-AM1-F	FAD2-AM1-R	GCTCAATTCCTCGCTCTTTC	GTCCAAGAGGGTTGTTGAGG
	3 & 4	2	FAD2-AM2-F	FAD2-AM2-R	CGGCAAATACCTCAACAACC	CCCTTAACCAATCCCACTCA
Full FAD2 fragment		CraFAD2-C1	upFAD2-C1	dnFAD2-C1	GCTATCGTTTATTTATTTTTCTTTC	AGCGATGAGAAGAACAATACAGAGA
		CraFAD2-C2	upFAD2-C2	dnFAD2-C2	CTCAACGCTATCGTTTATTTCTTTC	AGATAAAAGTAGCTTCACAGG
		CraFAD2-C3	upFAD2-C3	dnFAD2-C3	CCACGTACTATCCATTTTTGAAAGT	GGCACAACACAATGGATACTT
		CraFAD2-A	upFAD2-A	dnFAD2-A	GTCTCCACCTTTTGACTCTT	ATAACTCTGTTCTTTCACCATC

Table S2. The EMS induced TILLING populations in various plant species.

Species[a]	Genome size (Mbp)	EMS dosage (%)	Detection methods[b]	Mutation freq[c].	Reference
Arabidopsis	125	0.25	dHPLC, Li-Cor	1:153 Kb	(Colbert et al., 2001; McCallum et al., 2000)
Arabidopsis	125	0.5	Li-Cor	1:51 Kb	(Martín et al., 2009)
Barley	5,100	0.25-0.4	dHPLC	1:1 Mb	(Caldwell et al., 2004)
Barley	5,100	0.25-0.4	Li-Cor	1:500 Kb	(Gottwald et al., 2009)
Brassica napus	1,200	0.6	Li-Cor	1:41 Kb	(Wang et al., 2008)
Brassica oleracea	600	0.4	Li-Cor	1:447 Kb	(Himelblau et al., 2009)
Brassica rapa	529	0.3	CE	1:44 Kb	(Stephenson et al., 2010)
Bread wheat	17,000	0.75-1.2	Li-Cor	1:24 Kb	(Slade et al., 2005)
Bread wheat	17,000	0.5-0.7	AGE, HRM	1:23 Kb	(Dong et al., 2009b)
Bread wheat	17,000	0.9-1.0	PAGE, illumina	1:68 Kb	(Tsai et al., 2011; Uauy et al., 2009)
Bread wheat	17,000	0.8	AGE, PAGE	1:34 Kb	(Chen et al., 2012)
Crambe	***NA***	***0.3***	***454***	***1: 382 Kb***	***This study***
Durum wheat	11,000	0.75-1.0	Li-Cor	1:40 Kb	(Slade et al., 2005)
Durum wheat	11,000	0.7-0.75	PAGE, illumina	1:49 Kb	(Tsai et al., 2011; Uauy et al., 2009)
Maize	2,500	NA	Li-Cor	1:485 Kb	(Till et al., 2004)
Oat	13,000	0.9	MALDI-TOF	1:30 Kb	(Chawade et al., 2010)
Rice	389	1.6	Li-Cor	1:1 Mb	(Wu et al., 2005)
Rice	389	1.5	Li-Cor, illumina	1:294 Kb	(Till et al., 2007; Tsai et al., 2011)
Soybean	1,115	0.6	Li-Cor	1:140 Kb	(Cooper et al., 2008)
Tomato	900	1.0	Li-Cor	1:322 Kb	(Minoia et al., 2010)
Tomato	900	NA	454	NA	(Rigola et al., 2009)

[a] The species are listed in alphabetic order. [b] Detection methods: AGE (agarose gel electrophoresis), CE (capillary electrophoresis), HPLC (denaturing high performance liquid chromatogram), HRM (high-resolution melting), MALDI-TOF (Matrix-assisted laser desorption/ionization-time-of-flight mass spectrometer), PAGE (polyacrylamide gel electrophoresis). [c] mutation frequency. NA= not available

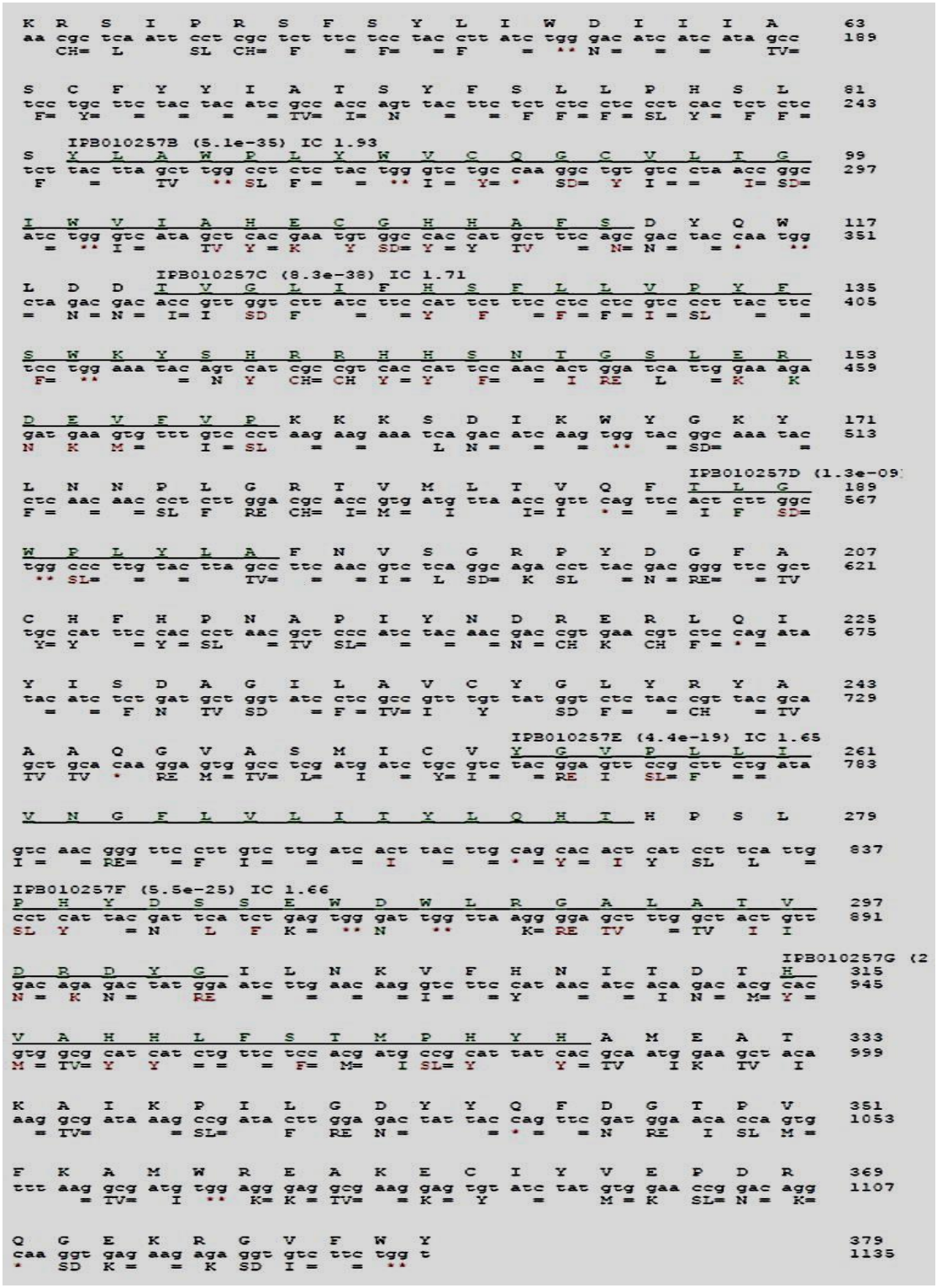


Figure S1. The locations of putative EMS type mutations in CraFAD2-C3 gene. The locations of putative EMS type mutations in CraFAD2-C3 gene. Six conserved amino acids blocks are underlined. The residuals below the nucleotide sequence are putative amino acids from EMS. The '*' and '=' indicate amino acid changes of non-sense and mis-sense mutation respectively.

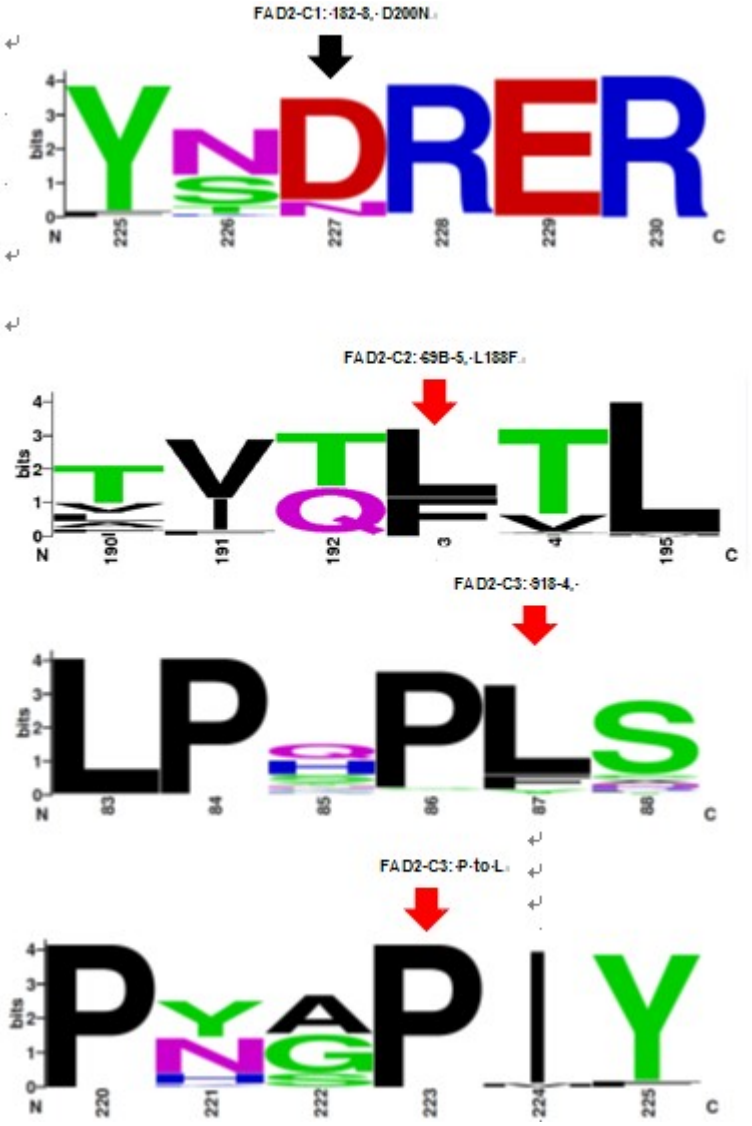


Figure S2. The positions of mutations in the FAD2 WebLogo. The deduced amino acid sequence of CraFAD2-C3 was used to search homologous FAD2 proteins in GenBank and the obtained amino acids were aligned and converted into a FAD2 WebLogo using the web tool WebLogo (http://weblogo.berkeley.edu/). The detected mutations were localized on the WebLogo with arrows.

References

Belfield, E.J., Gan, X., Mithani, A., Brown, C., Jiang, C., Franklin, K., Alvey, E., Wibowo, A., Jung, M., Bailey, K., Kalwani, S., Ragoussis, J., Mott, R. and Harberd, N.P. (2012) Genome-wide analysis of mutations in mutant lineages selected following fast-neutron irradiation mutagenesis of Arabidopsis thaliana. Genome Research 22, 1306-1315.

Bundock, P.C., Eliott, F.G., Ablett, G., Benson, A.D., Casu, R.E., Aitken, K.S. and Henry, R.J. (2009) Targeted single nucleotide polymorphism (SNP) discovery in a highly polyploid plant species using 454 sequencing. Plant Biotechnology Journal 7, 347-354.

Caldwell, D.G., McCallum, N., Shaw, P., Muehlbauer, G.J., Marshall, D.F. and Waugh, R. (2004) A structured mutant population for forward and reverse genetics in Barley (Hordeum vulgare L.). Plant Journal 40, 143-150.

Charlesworth, B., Morgan, M.T. and Charlesworth, D. (1993) The effect of deleterious mutations on neutral molecular variation. Genetics 134, 1289-1303.

Chawade, A., Sikora, P., Bräutigam, M., Larsson, M., Vivekanand, V., Nakash, M.A., Chen, T. and Olsson, O. (2010) Development and characterization of an oat TILLING-population and identification of mutations in lignin and β-glucan biosynthesis genes. BMC Plant Biology 10.

Chen, L., Huang, L., Min, D., Phillips, A., Wang, S., Madgwick, P.J., Parry, M.A.J. and Hu, Y.G. (2012) Development and characterization of a new TILLING population of common bread wheat (Triticum aestivum L. PLoS ONE 7.

Colbert, T., Till, B.J., Tompa, R., Reynolds, S., Steine, M.N., Yeung, A.T., McCallum, C.M., Comai, L. and Henikoff, S. (2001) High-throughput screening for induced point mutations. Plant Physiology 126, 480-484.

Comai, L. (2005) The advantages and disadvantages of being polyploid. Nature Reviews Genetics 6, 836-846.

Cooper, J.L., Till, B.J., Laport, R.G., Darlow, M.C., Kleffner, J.M., Jamai, A., El-Mellouki, T., Liu, S., Ritchie, R. and Nielsen, N. (2008) TILLING to detect induced mutations in soybean. BMC Plant Biology 8, 9.

Cuthbert, R.D., Crow, G. and McVetty, P.B.E. (2009) Assessment of agronomic performance and heterosis for agronomic traits in hybrid high erucic acid rapeseed (HEAR). Canadian Journal of Plant Science 89, 227-237.

Dierking, E.C. and Bilyeu, K.D. (2009) New sources of soybean seed meal and oil composition traits identified through TILLING. BMC Plant Biology 9.

Doležel, J., Bartoš, J., Voglmayr, H., Greilhuber, J. and Thomas, R.A. (2003) Nuclear DNA content and genome size of trout and human (multiple letters). Cytometry Part A 51, 127-129.

Dong, C., Dalton-Morgan, J., Vincent, K. and Sharp, P. (2009a) A modified TILLING method for wheat breeding. The Plant Genome 2, 39-47.

Dong, C., Vincent, K. and Sharp, P. (2009b) Simultaneous mutation detection of three homoeologous genes in wheat by high resolution melting analysis and mutation Surveyor®. BMC Plant Biology 9.

Gholami, M., Bekele, W.A., Schondelmaier, J. and Snowdon, R.J. (2012) A tailed PCR procedure for cost-effective, two-order multiplex sequencing of candidate genes in polyploid plants. Plant Biotechnology Journal 10, 635-645.

Gilles, A., Meglécz, E., Pech, N., Ferreira, S., Malausa, T. and Martin, J.F. (2011) Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. BMC genomics 12.

Gottwald, S., Bauer, P., Komatsuda, T., Lundqvist, U. and Stein, N. (2009) TILLING in the two-rowed barley cultivar 'Barke' reveals preferred sites of functional diversity in the gene HvHox1. BMC Research Notes 2.

Gustafsson, Å., Hagberg, A., Persson, G. and Wiklund, K. (1971) Induced mutations and barley improvement. Theoretical and Applied Genetics 41, 239-248.

Han, Y., Kang, Y., Torres-Jerez, I., Cheung, F., Town, C.D., Zhao, P.X., Udvardi, M.K. and Monteros, M.J. (2011) Genome-wide SNP discovery in tetraploid alfalfa using 454 sequencing and high resolution melting analysis. BMC genomics 12.

Himelblau, E., Gilchrist, E.J., Buono, K., Bizzell, C., Mentzer, L., Vogelzang, R., Osborn, T., Amasino, R.M., Parkin, I.A.P. and Haughn, G.W. (2009) Forward and reverse genetics of rapid-cycling Brassica oleracea. Theoretical and Applied Genetics 118, 953-961.

Jiang, X., Mu, B., Huang, Z., Zhang, M., Wang, X. and Tao, S. (2010) Impacts of mutation effects and population size on mutation rate in asexual populations: A simulation study. BMC Evolutionary Biology 10.

Laghetti, G., Piergiovanni, A.R. and Perrino, P. (1995) Yield and oil quality in selected lines of Crambe abyssinica Hochst. ex R.E. Fries and C. hispanica L. grown in Italy. Industrial Crops and Products 4, 203-212.

Lazzeri, L., Leoni, O., Conte, L.S. and Palmieri, S. (1994) Some technological characteristics and potential uses of Crambe abyssinica products. Industrial Crops and Products 3, 103-112.

Li, X., van Loo, E.N., Gruber, J., Fan, J., Guan, R., Frentzen, M., Stymne, S. and Zhu, L.H. (2012) Development of ultra-high erucic acid oil in the industrial oil crop Crambe abyssinica. Plant Biotechnology Journal 10, 862-870.

Marie, D. and Brown, S.C. (1993) A cytometric exercise in plant DNA histograms, with 2C values for 70 species. Biology of the Cell 78, 41-51.

Martín, B., Ramiro, M., Martínez-Zapater, J.M. and Alonso-Blanco, C. (2009) A high-density collection of EMS-induced mutations for TILLING in Landsberg erecta genetic background of Arabidopsis. BMC Plant Biology 9.

Mastebroek, H.D., Wallenburg, S.C. and van Soest, L.J.M. (1994) Variation for agronomic characteristics in crambe (Crambe abyssinica Hochst. ex Fries). Industrial Crops and Products 2, 129-136.
McCallum, C.M., Comai, L., Greene, E.A. and Henikoff, S. (2000) Targeted screening for induced mutations. Nature Biotechnology 18, 455-457.
Menda, N., Semel, Y., Peled, D., Eshed, Y. and Zamir, D. (2004) In silico screening of a saturated mutation library of tomato. Plant Journal 38, 861-872.
Meyer, M., Stenzel, U. and Hofreiter, M. (2008) Parallel tagged sequencing on the 454 platform. Nature Protocols 3, 267-278.
Minoia, S., Petrozza, A., D'Onofrio, O., Piron, F., Mosca, G., Sozio, G., Cellini, F., Bendahmane, A. and Carriero, F. (2010) A new mutant genetic resource for tomato crop improvement by TILLING technology. BMC Research Notes 3.
Pham, A.T., Lee, J.D., Shannon, J.G. and Bilyeu, K. (2010) Mutant alleles of FAD2-1A and FAD2-1B combine to produce soybeans with the high oleic acid seed oil trait. BMC plant biology 10, 195.
Pham, A.T., Lee, J.D., Shannon, J.G. and Bilyeu, K.D. (2011) A novel FAD2-1 A allele in a soybean plant introduction offers an alternate means to produce soybean seed oil with 85% oleic acid content. TAG Theoretical and Applied Genetics, 1-10.
Quince, C., Lanzen, A., Davenport, R.J. and Turnbaugh, P.J. (2011) Removing Noise From Pyrosequenced Amplicons. BMC Bioinformatics 12.
Reddy, T.V., Dwivedi, S. and Sharma, N.K. (2012) Development of TILLING by sequencing platform towards enhanced leaf yield in tobacco. Industrial Crops and Products 40, 324-335.
Rigola, D., van Oeveren, J., Janssen, A., Bonné, A., Schneiders, H., van der Poel, H.J.A., Van Orsouw, N.J., Hogers, R.C.J., De Both, M.T.J. and Van Eijk, M.J.T. (2009) High-throughput detection of induced mutations and natural variation using KeyPoint™ technology. PLoS ONE 4.
Rudloff, E. and Wang, Y. (2011) Crambe. In: Wild Crop Relatives: Genomic and Breeding Resources pp. 97-116.
Sikora, P., Chawade, A., Larsson, M., Olsson, J. and Olsson, O. (2011) Mutagenesis as a tool in plant genetics, functional genomics, and breeding. International Journal of Plant Genomics 2011.
Slade, A.J., Fuerstenberg, S.I., Loeffler, D., Steine, M.N. and Facciotti, D. (2005) A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. Nature Biotechnology 23, 75-81.
Stephenson, P., Baker, D., Girin, T., Perez, A., Amoah, S., King, G.J. and Østergaard, L. (2010) A rich TILLING resource for studying gene function in Brassica rapa. BMC Plant Biology 10.
Till, B.J., Cooper, J., Tai, T.H., Colowit, P., Greene, E.A., Henikoff, S. and Comai, L. (2007) Discovery of chemically induced mutations in rice by TILLING. BMC Plant Biology 7.
Till, B.J., Reynolds, S.H., Weil, C., Springer, N., Burtner, C., Young, K., Bowers, E., Codomo, C.A., Enns, L.C., Odden, A.R., Greene, E.A., Comai, L. and Henikoff, S. (2004) Discovery of induced point mutations in maize genes by TILLING. BMC Plant Biology 4.
Tsai, H., Howell, T., Nitcher, R., Missirian, V., Watson, B., Ngo, K.J., Lieberman, M., Fass, J., Uauy, C., Tran, R.K., Khan, A.A., Filkov, V., Tai, T.H., Dubcovsky, J. and Comai, L. (2011) Discovery of rare mutations in populations: Tilling by sequencing. Plant Physiology 156, 1257-1268.
Uauy, C., Paraiso, F., Colasuonno, P., Tran, R.K., Tsai, H., Berardi, S., Comai, L. and Dubcovsky, J. (2009) A modified TILLING approach to detect induced mutations in tetraploid and hexaploid wheat. BMC Plant Biology 9.
Wang, N., Wang, Y., Tian, F., King, G.J., Zhang, C., Long, Y., Shi, L. and Meng, J. (2008) A functional genomics resource for Brassica napus: Development of an EMS mutagenized population and discovery of FAE1 point mutations by TILLING. New Phytologist 180, 751-765.
Wu, J.L., Wu, C., Lei, C., Baraoidan, M., Bordeos, A., Madamba, M.R.S., Ramos-Pamplona, M., Mauleon, R., Portugal, A., Ulat, V.J., Bruskiewich, R., Wang, G., Leach, J., Khush, G. and Leung, H. (2005) Chemical- and irradiation-induced mutants of indica rice IR64 for forward and reverse genetics. Plant Molecular Biology 59, 85-97.

Chapter 5

Detection of induced mutations in *CaFAD2* genes by next generation sequencing leading to the production of improved oil composition in *Crambe abyssinica*

Jihua Cheng[1,2] , Elma M.J. Salentijn[1], Bangquan Huang[2], Christel Denneboom[1], Annemarie C. Dechesne[1], Frans A. Krens[1], Weicong Qi[1,2], Richard G.F. Visser[1], Eibertus N. van Loo[1*]

[1] Wageningen UR Plant Breeding, P.O. Box 16, 6700 AA Wageningen, the Netherlands

[2] College of Life Science, Hubei University, People's Republic of China

Keywords: *Crambe abyssinica*, mutation breeding, *FAD2*, Oleic acid, Polyunsaturated fatty acids, Illumina sequencing

To be submitted.

Abstract

Crambe abyssinica is a dedicated oil crop for industrial applications. Crambe seed oil naturally contains up to 60 % of erucic acid (C22:1Δ13) which is a useful raw material for industrial manufacture of plastics and lubricants. A further increase of C22:1 levels and reduction of the polyunsaturated fatty acid (PUFA) content, which naturally makes up around 16 % of the fatty acids in crambe oil, would be a valuable improvement. To reach this goal targeted mutagenesis is performed to identify ethyl methane sulphonate (EMS) induced mutations in *CaFAD2* genes which are expected to be associated with oleic acid (C18:1) and PUFA contents in crambe oil.

Short-read sequencing (Illumina) was applied to screen a mutant population of crambe for EMS induced mutations in *CaFAD2* using two sequence analysis strategies. In the first strategy the mutation frequency was estimated for specific pre-determined positions. In the second strategy every position in a *CaFAD2* fragment was tested for the presence of mutations. Seventeen novel mutations were detected and assigned to the *CaFAD2-C1*, *-C2* and *-C3* genes. Plants carrying specific nonsense alleles (in *CaFAD2-C1,* and *-C3*) and missense mutations (in *CaFAD2-C1* and *-C2*) were identified and segregating populations were tested for oil composition. The homozygous nonsense-mutants for *CaFAD2-C3* showed a moderate alteration in both C18:1 (+25 %) and PUFA (−25 %) contents compared to the wild-type, whereas in the heterozygote plants an intermediate oil composition was observed (+12 % in C18:1 and −11 % in PUFA, relative to wild-type levels). For the nonsense mutation in *CaFAD2-C1* only a single heterozygous plant survived. Between the nonsense-heterozygotes of *CaFAD2-C1* and *-C3*, a significant difference was found only in C18:1 and C22:1 content. There were slight alterations observed in the oil composition of the missense mutants. This study shows mutation detection by a next generation sequencing (NGS) technique and confirms the expectation that knocking out a single *CaFAD2* can lead to changes in crambe oil composition. Combining the

nonsense mutations in different *CaFAD2* genes may result in additional changes in C18:1 and PUFA contents.

Introduction

Mutagenesis as a non-genetically-modified (GM) approach is an important tool in crop improvement because the introduction of genetic variation may induce improved or novel phenotypes (Parry et al., 2009; Slade and Knauf, 2005; Till et al., 2007). Due to stable genotypic changes, these improved phenotypes, can be directly used in breeding programs. Therefore, natural or induced genetic diversity has been used for improvement in a wide range of crops (Comai et al., 2004; Gady et al., 2012; Parry et al., 2009; Till et al., 2007). In particular, the use of mutagenesis to create novel variation is valuable in those crops with restricted natural diversity (Parry et al., 2009). Furthermore, polyploid crops are considered to be perfect candidates for mutation breeding in that due to the gene redundancy of the multiple genes copies in polyploid genomes, mutation accumulation is allowed without severe negative effects (Backes, 2013).

Historically, the use of mutagenesis in breeding has involved screening for mutants with improved traits, so called 'forward genetic' screens. The success of this approach has been proven extensively since large amounts of varieties derived from mutagenesis have been released over the past 70 years (Ahloowalia and Maluszynski, 2001; Ahloowalia et al., 2004). However, there are limitations to forward genetic screens, for instance the time and effort required to identify novel phenotypes within vast number of individuals (Brady and Provart, 2007). Also, the use of forward approaches is hampered by the difficulty of identifying individual mutants with subtle changes in phenotypes, especially for polyploid crops in which mutations have no detectable effect on the plant because of genetic redundancy (Stemple, 2004). Recently, a reverse genetic approach, targeting induced local lesion in genome (TILLING), has been applied for mutant detection and functional analysis of candidate genes which are known to be related to certain traits (Slade et al., 2005). It is considered to be a straightforward approach since the genetic variation in target genes could result in expected traits (Comai and Henikoff, 2006). To efficiently detect mutations in target genes, various techniques have been developed to promote the

application of this approach, and these techniques have conditionally been applied in different crops (reviewed by Sikora et al., 2011).

TILLING has entered into a new era: TILLING by sequencing; since the introduction of next generation sequencing (NGS) technology and availability of whole genome sequences in some crops (Tsai et al., 2011). The NGS technology provides an opportunity to re-sequence whole genomes and detect mutations genome-wide for the crops for which reference genome sequences are available (Tsai et al., 2011; Varshney et al., 2009). Alternatively, for crops without available reference genome, it is still feasible to find mutations locally in candidate genes by sequencing target gene amplicons (Bundock et al., 2009; Marroni et al., 2011; Rigola et al., 2009). An example is the mutation detection in an EMS mutant population for the hexaploid crop *Crambe abyssinica*. In this case, the amplicons of *FAD2* genes were sequenced by 454 sequencing, resulting in the identification of mutations in different loci of the *FAD2* genes (Chapter 4, this thesis).

Following the sequencing of selected amplicons, a routine pipeline is to align the sequencing reads to reference sequences of target genes prior to mutation detection, an approach which needs high quality sequence alignments. However, in case of sequencing candidate genes present as multiple members in a gene family, it is problematic to locate short reads to the corresponding loci by alignment due to high sequence identity (Griffin et al., 2011). Meanwhile, strand switches and PCR drift may take place during PCR amplification so that sequencing reads contain indels or are combinations of sequences of different loci (Abel et al., 2010; Qiu et al., 2001; Schloss et al., 2011; Vandenbroucke et al., 2011). This further makes it problematic to assign some reads to the correct loci by alignment. To date, studies have been done for mutation detection by NGS but these studies all are based on read alignments and assembly (Marroni et al., 2011; Rigola et al., 2009; Tsai et al., 2011; Wells et al., 2013).

C. abyssinica is a hexaploid oil crop and dedicated to industrial oil production because of a high erucic acid (C22:1) content (~60% of the total fatty acids in its seed

oil). The wished improvements to crambe oil include elevating C22:1 and oleic acid (C18:1, naturally ~18% of the fatty acids in the oil) and reducing polyunsaturated fatty acid (PUFA, naturally in total ~15% of the fatty acids in the oil) contents, e.g. linoleic acid (C18:2) and linolenic acid (C18:3). It is known that fatty acid desaturase2 (*FAD2*) plays a critical role for conversion of C18:1 to PUFA (Okuley et al., 1994). It is also known that there are three expressed *FAD2* genes (*CaFAD2-C1*,-*C2* and -*C3*) present in crambe and one of them (*CaFAD2-C3*) is highly related to seed oil biosynthesis (Cheng et al., 2013a; Cheng et al., 2013b). Through regulating *CaFAD2* expression by RNAi, a significant increase and decrease has been found in C18:1 and PUFA respectively in crambe oil (Cheng et al., 2013b). In other crops, selection of natural variation in *FAD2* genes or mutagenizing *FAD2* genes has led to pronounced changes in the seed oil similar to crambe RNAi lines (Beló et al., 2008; Pham et al., 2011; Tanhuanpää et al., 1998). However, the limited natural variation in *CaFAD2* is a bottleneck to improve oil content and composition by selection breeding and calls for mutation breeding (Cheng et al., 2013a).

For this reason, a crambe mutant population has been constructed by EMS-induced mutagenesis of crambe cv. 'Galactica' that can be used for the identification of plants carrying mutations in genes of interest (Chapter 4 of this thesis). To detect mutations in this large mutant population of hexaploid crambe in a high-throughput and efficient way, the possibility of mutation detection by Illumina sequencing in combination with alignment-free sequence analysis approaches was explored, using the *CaFAD2* genes as a target.

Materials and methods

Crambe mutant population

The crambe population was established by mutagenizing seeds of crambe cultivar 'Galactica' with 0.25 % of ethyl methanesulfonate (EMS), 12,480 individual M_2 lines were obtained. A M_2-DNA library was constructed through isolating the genomic DNA from pooled leaves of about 10 lines (1,378 DNA pools). The procedures of

population establishment and M_2 DNA pooling are described in Chapter 4 of this thesis.

FAD2 amplicon preparation and Illumina sequencing

The conserved regions of a crambe FAD2 gene (*CaFAD2-C3*, accession number: JX964745) were predicted by CODDLE (Colbert et al., 2001). Meanwhile, the putative EMS-type transitions, G to A and C to T, were predicted in this gene. Specific primers were developed for an amplicon of 272 bp that covers two conserved histidine-rich domains of the CaFAD2-C3 protein (**Supplementary Figure S1**). The forward and reverse primers were barcoded with 68 unique six-nucleotide (6-nt) indexes on the 5' ends to form 34 forward and 34 reverse fusion primers (**Supplementary Table S1**), by which 1,156 pairs of fusion primers (34×34) can be combined. Of them, 1,100 pairs of fusion primers were used to amplify amplicons from 1,100 DNA pools. The amplicons were purified on agarose gel and paired-end sequenced (100 bp) in a lane of Illumina Hi-seq 2000.

Sequence analysis

The sequence analysis towards detection of mutations and errors was carried out following two approaches.

Approach 1: motif counting

Index sorting

Base-called reads were trimmed for quality and the paired-end reads were sorted in two steps for index. In the first step, all paired-end reads were sorted by the 34 indexes in forward primers. After this step, 34 datasets were obtained and each dataset contains the forward reads with identical index and their corresponding paired-end reads with 34 unique indexes in reverse primers. In the second step, the obtained 34 datasets were individually sorted by the 34 unique indexes in the reverse primers. In this process, all reads were sorted by the combinations of indexes (34×34) to

represent 1,156 pools. In this study 1,100 combinations were used. After this, the forward and reverse reads in each index combination were split.

Selection of candidate mutations

To select pools that contain plants with predicted nonsense or missense EMS mutations in conserved sequence domains of *CaFAD2-C1*, *CaFAD2-C2* and *CaFAD2-C3*, the Illumina reads were screened for the presence of several 9-nt motifs, including wild-type and mutant motifs, which contain 30 specific mutation positions, e.g. CTAACCGGC and CTAATCGGC (wt and mt motifs with a C to T transition). Next, the frequency of a mutation in a pool was calculated by the percentage of mutant reads (N mt-reads) of total reads (N mt-reads + N wt-reads). All motifs are listed in **Supplementary Table S2**; of them a motif (MT_30) was used as a control to count the percentage of a confirmed mutation which was obtained earlier (Chapter 4, this thesis). Candidate mutations were selected to meet the following two criteria: 1) the number of mutant reads is higher than 20 in a pool; 2) the percentage of mutant reads in this pool is higher than 0.2 %.

Assigning mutations to crambe FAD2 genes and mutant identification

Mutant reads in a candidate pool were extracted by an in-house developed program Pattern Research (I Padioleau, EMJ Salentijn, and MJM Smulders, unpublished). The mutant reads were aligned with the reference sequences of crambe FAD2 genes (accession numbers: JX964741, JX964742, JX964743, JX964744, JX964745, JX964746 and JX964747) to locate the selected mutations to the corresponding crambe *FAD2* gene.

Approach 2: reporting SNPs in each position of all reads

SNP-reports for all pools

First, all reads were separated into the different barcode pools by index sorting as done for the approach 1. Second, using a Perl script, reads were individually analyzed to determine the number of A, C, G, T and N calls per position. Counts were made per

nucleotide position, after determining whether the read locally could be allocated to one of the five FAD2 loci. The whole analysis pipeline consists of the following steps:

1. Read a pair of Illumina reads. Change all nucleotide calls with a quality score of less than 20 (Phred > 0.05) to N. The nucleotide calls of N were not taken into account in the later analysis.

2. Allocate the read to one of the five *CaFAD2* loci on the basis of the first 10 nucleotides in the forward read (where several SNPs occur that allow this allocation, **Supplementary Figure S2**).

3. Determine whether an indel occurred in the pair of reads, by counting the number of differences ('edits') between the read and the reference sequence to which the pair was. This was done in the forward and reverse read separately. When the number of 'edits' was higher than 20, the read was discarded as containing an indel. These indels proved to be read errors based on their very low coverage

4. Determine whether a strand switch might have occurred at the position under investigation. When a nucleotide transition was found, it was decided to be a putative mutation or a strand switch by counting the number of 'edits' in the remaining read (after the first 10 bases).

5. An SNP-report was created for all positions on the basis of all reads of a 10-genotype pool, indicating the number of wild-type and nucleotide transition (putative mutation) per position and the frequency of the putative mutation. This created 1000+ SNP-reports (one per 10-genotype pool).

Statistical analysis to filter the SNP-reports for potential mutations

EMS produces other mutations than G to A and C to T only in a 100-fold lower frequency (Gilchrist et al., 2006; Winkler et al., 2005). Those 'non-EMS' loci were considered to be read errors. A statistical test was carried out assuming a Poisson distribution of errors where the expected number of errors is equal to the average number of the non-EMS read errors on a nucleotide position (null hypothesis H_0).

Only when this probability was lower than 1e-3, the potential mutation was reported in the summary report. This probability of falsely calling a read error a mutation (false positive probability, Pfp) was converted to a -log value (LgFP = -Lg(Pfp*1000)). When the LgFP>0, the probability of false detecting a mutation (Pfp) is lower than 1e-3. In case, no read errors were found on a certain position and only potential EMS-mutations, LgFP was capped at 6 (Pfp=1e-3). In case, no EMS-mutation was found, LgFP is -3 (Pfp = 1). In case, only Ns were found at positions, LgFp was set to -4 (Pfp=10).

Mutant identification and phenotyping for oil composition

The mutants were identified in M_3 lines. To identify the candidate mutant lines among the ten lines composing a pool, about ten plants (M_3) were grown per line. Per line, genomic DNA was isolated from the combined leaf samples of all ten M_3 plants of a line. The *CaFAD2* amplicons (*CaFAD2-C1*, *-C2* or *-C3*) were amplified from these isolated DNA samples, and the amplicons were directly sequenced (Sanger sequencing) to find the positive lines containing mutations. Next, genomic DNA was isolated from the individual plants of the positive lines to find mutants by Sanger sequencing *CaFAD2* amplicons. The primers used for the amplicon amplification and sequencing are listed in **Supplementary Table S3**.

For oil measurement, 20 bulk seeds were harvested and measured for homozygous, heterozygous and wild-type plants per mutation. In addition, 20 single seeds were measured for a heterozygote per mutation. The procedures of fatty acid extraction and analysis were performed as described by Cheng et al. (2013a). To compare the fatty acid levels in the mutants and wild-type, individual fatty acid level in a mutant was calculated relative to this fatty acid level in the wild-type.

Results

A coding region of the *CaFAD2* genes (~272 bp in length) were chosen for mutation detection as this region contains two conserved histidine domains required for FAD2 function (Okuley et al., 1994). The *CaFAD2* amplicons for this region were amplified

from 1,100 pools, representing a total of 11,000 plants. The amplicons were uniquely bar-coded per pool and sequenced in a single lane of Illumina Hi-seq 2000. Overall, sequencing produced over 40 million paired-end reads that were assigned to the 1,100 pools based on barcode combinations in paired-end reads. Most of the pools contain 10,000 to 20,000 reads and amplicon coverage was estimated to be around 100 to 1,500 (**Supplementary Figure S3**).

Mutation detection by approach 1

Motif counting per pool

Based on the prediction of non-sense and missense EMS-type transitions (G to A and C to T) in conserved domains of the sequence of *CaFAD2-C3* (**Supplementary Figure S1**), 30 pairs of 9-mer nucleotide sequence motifs (wild-type motifs and the corresponding mutant motifs) have been designed for 30 possible EMS mutation sites in the sequences of the three respective *CaFAD2-C* genes (-C1, -C2 and -C3) (**Supplementary Table S2**). The number of mutant reads (N mt-reads) and the wild-type reads (N wt-reads) containing these specific motifs were counted throughout all the sequences obtained from all 1,100 pools. First, as a control, a motif for a confirmed EMS mutation in pool '960' (MT_30 in **Supplementary Table S2**) was used to count mutant reads throughout all pools to evaluate the possibility of mutation detection by motif counting. Obviously, most of the pools have a low background frequency (< 0.2 %) of reads with this specific transition, with numbers of reads lower than 20 (**Figure 1**). Meanwhile, it was shown that two pools have a high depth of mutant reads (> 80) with a frequency of this mutation that is higher than 0.2 % (**Figure 1**). Furthermore, one of these two pools has been confirmed to have this control mutation. Therefore, to avoid detecting false mutations, all mutation detections were carried out following two criteria, higher than 0.2 % and 20 in frequency and depth of mutant reads respectively.

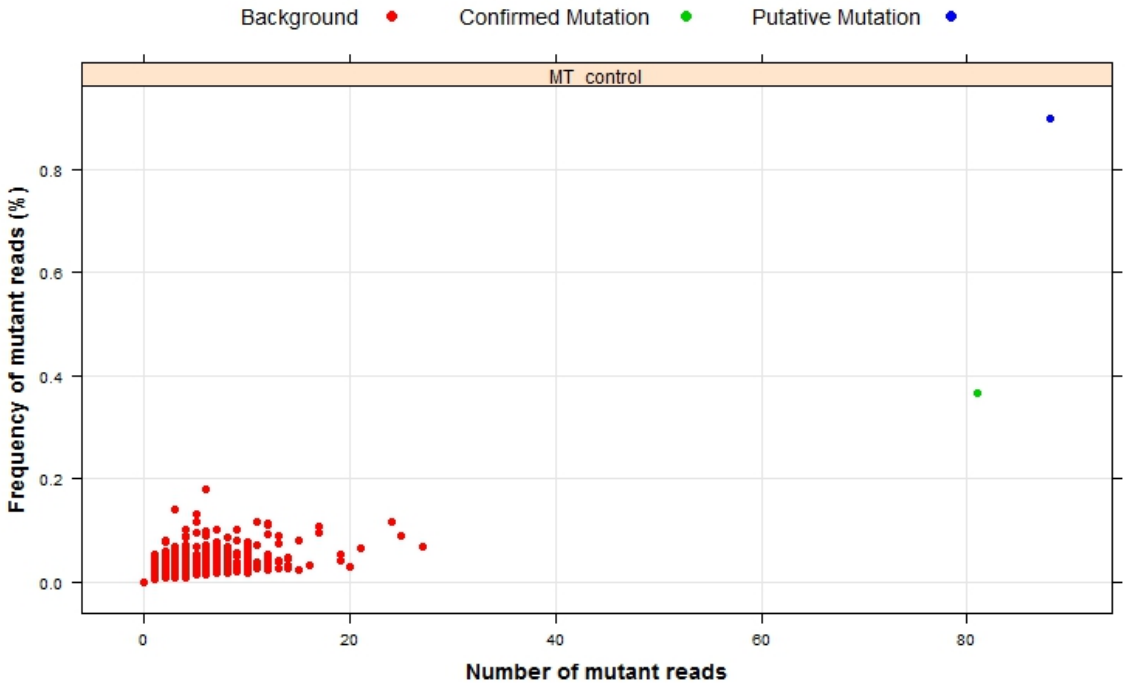


Figure 1. The frequency vs. number of mutant reads for the position (MT_30/MT_control) in all pools. Each point indicates a pool.

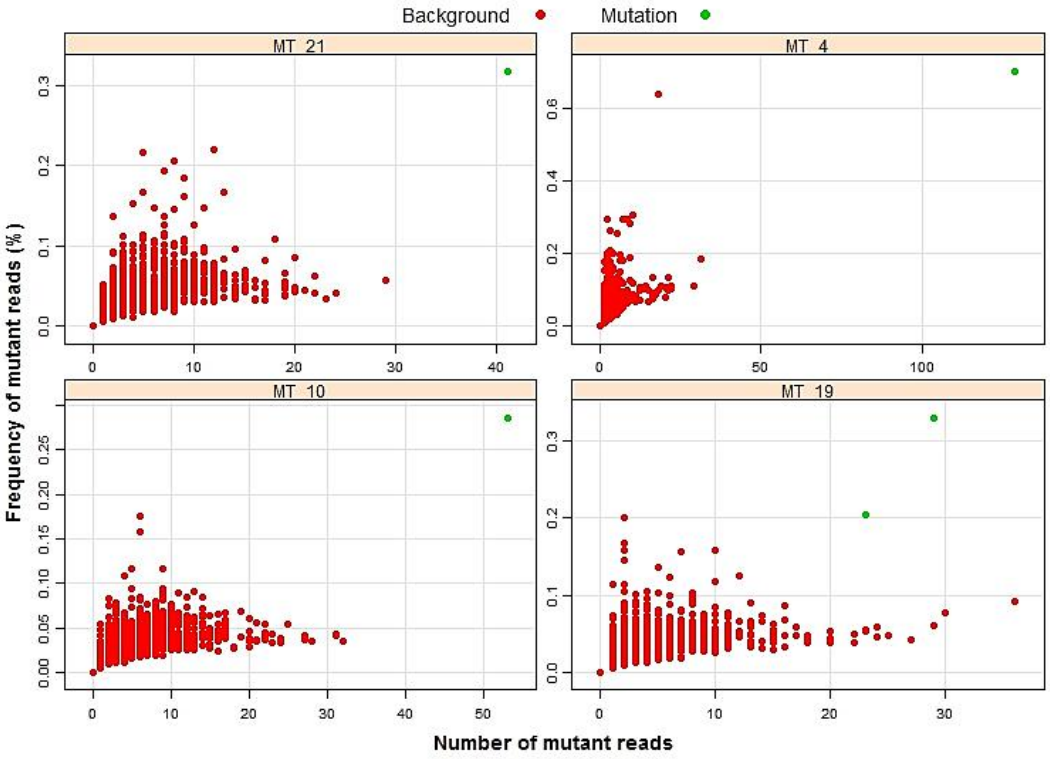


Figure 2. The frequency vs. number of mutant reads for four candidate positions in all pools. Each panel represents a position in all pools, and each point indicates a pool.

Selection of candidate pools

Using these criteria, all putative mutations were screened in the 1,100 pools using the 30 pairs (wild-type and mutant) of motifs. Only four motifs were found to have high frequency (> 0.2 %) and read depth (> 20 reads) in specific pools. For instance, over 100 mutant reads were found in pool '875' for the position MT_4 and they counted over 0.6 % of the total reads (**Figure 2**). All these mutant reads were identified to be mutants of *CaFAD2-C3* and this mutation was thus designated as *875-mt4-c3*, which results in an internal stop codon (**Table 1**).

For two other positions, MT_10 and MT_21, high levels of mutant reads (n=50, frequency 0.3 %) were detected respectively in pool '200' and '323'. The mutant reads were assigned to *CaFAD2-C2* and *-C1* and caused amino acids changes from Gly to Asp (mutation *200-mt10-c2)* and from Ala to Val (mutation *323-mt21-c1*) (**Figure 2, Table 1**).

Mutant reads for the position MT_19, resulting in an internal stop codon, were simultaneously found in two pools, '173' (*173-mt19-c1*) and '303' (*303-mt19-c2*) in 0.2 % and 0.3 % of the total reads for *CaFAD2-C1* and *CaFAD2-C2* respectively (**Figure 2** and **Table 1**). In summary, among five candidate pools 5 unique mutations were detected, one in *CaFAD2-C3*, two in *CaFAD2-C2* and two in *CaFAD2-C1* (**Table 1**).

Table 1. Mutations detected by motif searching

Pool	ID of mutation	Gene	Position	Mutant Motif	N_mt reads	N_wt reads	Frequency (%)	Amino acid transition
173	173-mt19-c1	CaFAD2-C1	MT_19	GTCT*A*GGTC	11285	23	0.2033958	W to *
200	200-mt10-c2	CaFAD2-C2	MT_10	TGTG*A*CCAC	18535	53	0.2859455	G to D
303	303-mt19-c2	CaFAD2-C2	MT_19	GTCT*A*GGTC	8815	29	0.3289847	W to *
323	323-mt21-c1	CaFAD2-C1	MT_21	ATAG*T*CCAC	12910	41	0.3165779	A to V
875	875-mt4-c3	CaFAD2-C3	MT_4	ATCTG*A*GTC	18462	129	0.6987325	W to *
960†	960-mt30-c3	CaFAD2-C3	MT_30	GAT*A*TCTGA	22029	81	0.3663501	D =

† Control mutation.

*=stop codon.

Mutation detection by approach 2

Approach number one investigated the EMS-type transitions in specific candidate positions. To search for mutations in other positions in the selected region of candidate genes, we made an analysis pipeline to check the nucleotide transitions over the full length reads.

Primary allocation of reads to the FAD2 loci

For each pool, uniquely barcoded paired-end FAD2 amplicon reads, were analyzed for their sequence identity and primarily allocated to one of the five FAD2-loci on the basis of the first 10 nucleotides in the forward reads. It showed that the majority of reads (> 95 %) were allocated to the four loci, *CaFAD2-B*, *-C1*, *-C2* and *-C3*, the remaining reads are from the other locus *CaFAD2-A*. Since the loci *CaFAD2-A* and *-B* are not expressed in crambe seeds (Cheng et al., 2013a), no analysis for possible mutations was carried out for the reads of these two loci.

Chimeric reads detection

Strand switches and indels in reads could be produced during PCR amplification (Qiu et al., 2001; Schloss et al., 2011). In particular, strand switches lead to a formation of so-called chimeric reads, e.g. chimeric reads combining *CaFAD2-C1* and *CaFAD2-C2*, which are possibly allocated to incorrect loci.

After the primary allocation of reads to the different *FAD2* loci, the sequence downstream of the first 10 nucleotides of the forward reads was screened for signs of chimeras. This was done by counting the differences ('edits') in the remaining sequence per read (behind the first 10 bases that determines from which locus the sequence is amplified) compared to all *CaFAD2* loci. In our analysis, maximal two 'edits' difference are allowed to confirm the primary allocation (assuming one is a mutation and the other is a read error) in case of the absence of chimeras. For instance read 1 in **Figure 3** has two 'edits' to C1 in the remaining sequence and therefore is still allocated to C1 as primary allocation.

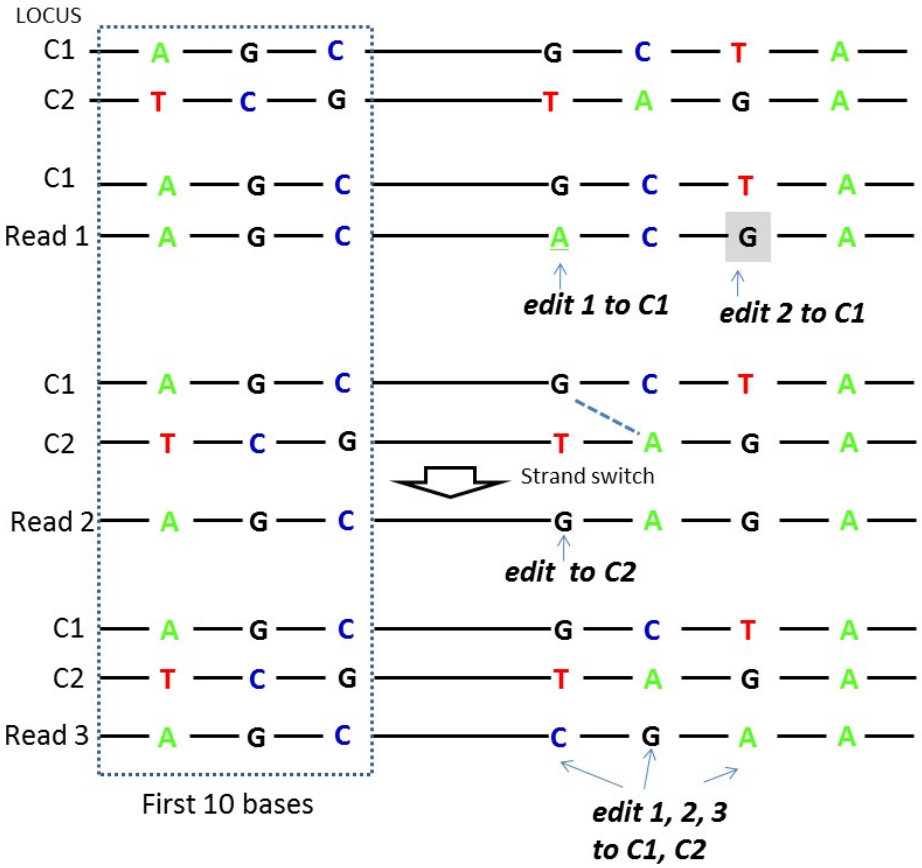


Figure 3. Identification of strand switch during read allocation. Reads are primarily assigned to one of the FAD2 genes by the first 10 base pairs. Differences in the sequence downstream of these 10 base pairs determine if the sequence can truly be allocated to this gene, is representing a strand-switch during PCR or is derived from another locus. For instance, read 1 is derived from a C1 amplicon as determined by the first 10 bp and is different to C1 and C2 by two 'edits' and is therefore assigned to C1; read 2 is primarily allocated to C1 as determined by the first 10 bp, and has two edits difference to C1 but only one 'edit' different to C2 in the remaining sequences, indicating a strand switch between C1 and C2 on the position 'G'; read 3 has the same number of 'edits' to C1 and C2 3, and therefore cannot be allocated to any locus and is discarded from further analysis.

If the number of 'edits' with respect to the initially allocated locus than with respect to one of the other loci, we consider to indicate for a strand switch event, on the condition the number of edits in the other locus is less than two (as depicted for read 2 in **Figure 3)**. This read 2 has two edits compared to C1 and 1 compared to C2, and

therefore it is concluded that a strand switch occurred. In some cases the number of 'edits' from a potential strand switch location was the same (and higher than two) with respect to the initially allocated locus and another locus in which case it was not possible to decide that a strand switch was more likely than a mutation or a series of read errors and then the read was not allocated to one of the loci. Such unallocated parts of reads were discarded for the later analysis. Read 3 shows an example of such a case, where starting with the putative mutation of G to C the number of edits to both C1 and C2 was three. Read 3 could therefore not be unequivocally be allocated to either C1 or C2 and therefore was not allocated and discarded from the analysis (read 3 in **Figure 3**).

Strand switches occurred in a high frequency (10 % overall, even 20 % in some pools). Rarely, indels were found to be present (in less than 0.05 % over all reads)

Reporting mutations and read errors

All nucleotide transitions (mutations and read errors) were counted per pool for each position of the full length reads.. The probability of nucleotide transitions was calculated based on Poisson distribution, and probability of a read error (false positive, Pfp, $P <= 0.001$). As shown in **Figure 4**, the chance on read errors is higher for the final positions of the forward reads (towards position 94) and for the first positions of the reverse reads. Generally, read errors rates are higher for the reverse read compared to the forward read.

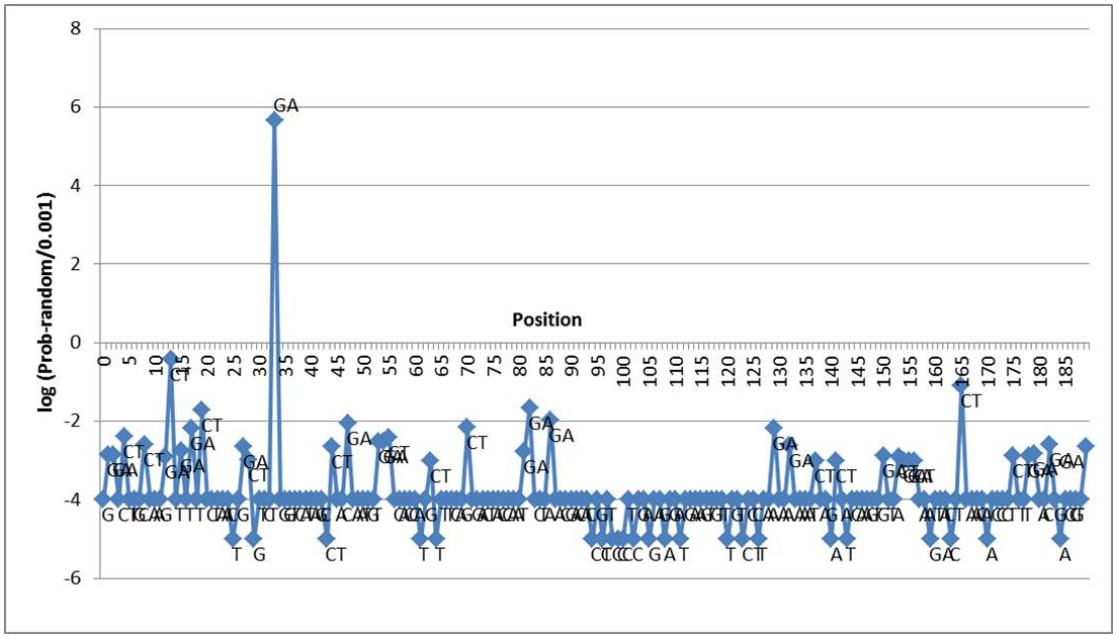


Figure 4. Likelihood of a mutation (LgFP) in 10-genotype pool '875' across the full read of *CaFAD2-C3*. This LgFP was used to report the likelihood of a mutation for each position in the *CaFAD2-C3* sequence. Positions 0-94 are from the forward read and position 95-189 are from the reverse read of the amplicon from the *CaFAD2-C3* gene (**Supplementary Figure S2**). A mutation was detected at position 35 (G to A). The letters indicate the nucleotide calls

Table 2. Mutations detected in the SNP-report of all positions

Pool	Locus	Positions in amplicon (bp)	Total reads (n)	Mutant reads (n)	Frequency (%)	LgFP[b]	Amino acid change
758	c1	132	5719	72	1.3	6	K=
385	c1	141	2208	48	2.2	6	D=
761	c2	58	5657	92	1.6	6	H=
856	c3	31	15003	92	0.6	6	I=
952	c2	147	11392	101	0.9	6	K=
200	c3	35	5114	46	0.9	6	V to I
617	c3	46	2350	11	0.5	4.6	H=
18.1	c3	34	2788	18	0.6	5.8	W to *
183	c2	88	4793	52	1.1	6	D=
688	c2	86	5996	35	0.6	6	D to N
995	c3	46	4810	91	1.9	6	H=
1021	c1	89	3040	22	0.7	5.3	T=
960†[a]	c3	141	7719	50	0.6	6	D=
323[a]	c1	42	3115	35	1.1	6	A to V
200[a]	c2	54	5290	35	0.7	6	G to D
303[a]	c2	33	3767	19	0.5	6	W to *
875[a]	c3	34	14881	120	0.8	6	W to *

*=stop codon †The control mutation

[a] Positive pools detected by approach 1

b LgFP = likelihood of falsely deciding that a mutation is present. LgFP = - Lg (Pfp*1000). Pfp= probability of falsely deciding that a mutation occurred with m mutant reads and given the number of non-EMS read errors found on this position in this pool*0.001), using a Poisson distribution of the number of read errors; when LgFP > 0 this probability is < 0.001

In total, 17 mutations were found by approach 2 including the control mutation ('960') throughout 1,100 pools (**Table 2)**. The mutation frequency varied from 0.5 % up to 2.2 % in pools (**Table 2**) and is generally higher than that observed in approach 1

(0.2 % to 0.6 % in **Table 1**). Of the 16 novel mutations (the control mutation excluded), three are nonsense mutations, four are missense mutations and the remaining nine mutations are silent (**Table 2**). Meanwhile, four mutations (in pool '200', '303', '323' and '875') were also detected by approach 1; however, the mutation in pool '173' (*173-mt19-c1*) detected by the approach 1 was not confirmed by approach 2 **(Table 1** and **Table 2)**.

Estimation of mutation density in the population

Six and seventeen mutations (control mutation included) were detected by two approaches respectively. Of them, five mutations including the control mutation were found by both approaches. Therefore, 18 unique mutant pools were found from a total of 1,100 10-genotype pools (from about 11,000 genotypes) by these two approaches. Only typical EMS-mutations are reported (G to A and C to T). In the amplicons analyzed, about 77 nucleotide positions are either G or C (the potential mutation sites, 39 in forward and 28 in reverse reads of *CaFAD2-C3* locus). Therefore, around 2.5 M bases of potential mutation sites were analyzed (77 bases per amplicon×3 loci per line×11,000 lines=2.5 M bases). Given that 18 mutations were found, the mutation frequency amounts to slightly lower than 1 mutation in 140 kbp (18/2.5 Mbp).

Identification of mutant lines

Several interesting mutations were observed. Four mutations were nonsense mutations that introduced stop codons (pool '18.1', '173', '303' and '875') and two were missense mutations that introduced changes in a conserved Histidine-domain (pool '200' Gly108Asp and pool '323' Ala104Val). The nonsense mutations in pool '18.1' and '875' caused a stop codon in the same position (amino acid position: 101) of CaFAD2-C3 protein (**Supplementary Figure S4**). Therefore, five unique mutations (in pool '173', '303', '875', '200' and '323') were chosen to identify the corresponding mutant plants in the M_3 generation.

The M_3 progeny of the ten lines composing a pool were grown and by sequencing *CaFAD2-C* amplicons, the M_3 progenies containing the mutants were identified.. In

this screen five positive lines were found for four unique mutations. The lines 875-2, 200-5 and 173-5 were found to contain the mutations, *875-mt4-c3*, *200-mt10-c2* and *173-mt19-c1* respectively (**Table 3**) and two lines, 323-2 and 323-7, contain the mutation *323-mt21-c1*. Lines harboring mutation *303-mt19-c2* were not found (**Table 3**). Subsequently, by sequencing the target amplicons of individual M_3 plants the homozygous and heterozygous mutants (M_3) corresponding to *875-mt4-c3*, *200-mt10-c2* and *323-mt21-c1* were identified in the respective lines 875-2, 200-5 and 173-5 (**Supplementary Figure S5**). In these M_3 progenies a Mendelian segregation of the mutations was observed (1 wt/wt : 2 wt/mt : 1 mt/mt). For 173-5, the M_3 progeny consisted of a single heterozygous plant for the mutation *173-mt19-c1* (**Table 3**).

Table 3. Mutant identification from the positive lines

ID of mutations	Positive lines	No. of M_3 individuals	No. of Homozygotes (mt/mt)	No. of heterozygotes (mt/wt)	No. of wildtype (wt/wt)
173-mt19-c1	173-5	1	0	1	0
200-mt10-c2	200-5	12	2	6	4
303-mt19-c2	nd	nd	nd	nd	nd
323-mt21-c1	323-2	6	1	4	1
323-mt21-c1	323-7	7	1	6	0
875-mt4-c3	875-2	9	2	6	1

nd=not determined

Oil composition in the mutants

The oil composition of the segregating and genotyped M_3 progenies was measured on twenty bulk seeds per plant of respectively homozygous mutants (“mt”), homozygous

wild type ("wt") and heterozygous mutants ("mtwt"). In addition, per mutation, twenty single seeds were analyzed for the heterozygotes. The contents of C18:1, C18:2, C18:3 and C22:1 were quantified for the seeds. For the mutants, the level of individual fatty acid compound is calculated relative to the level of this compound in the wild-type.

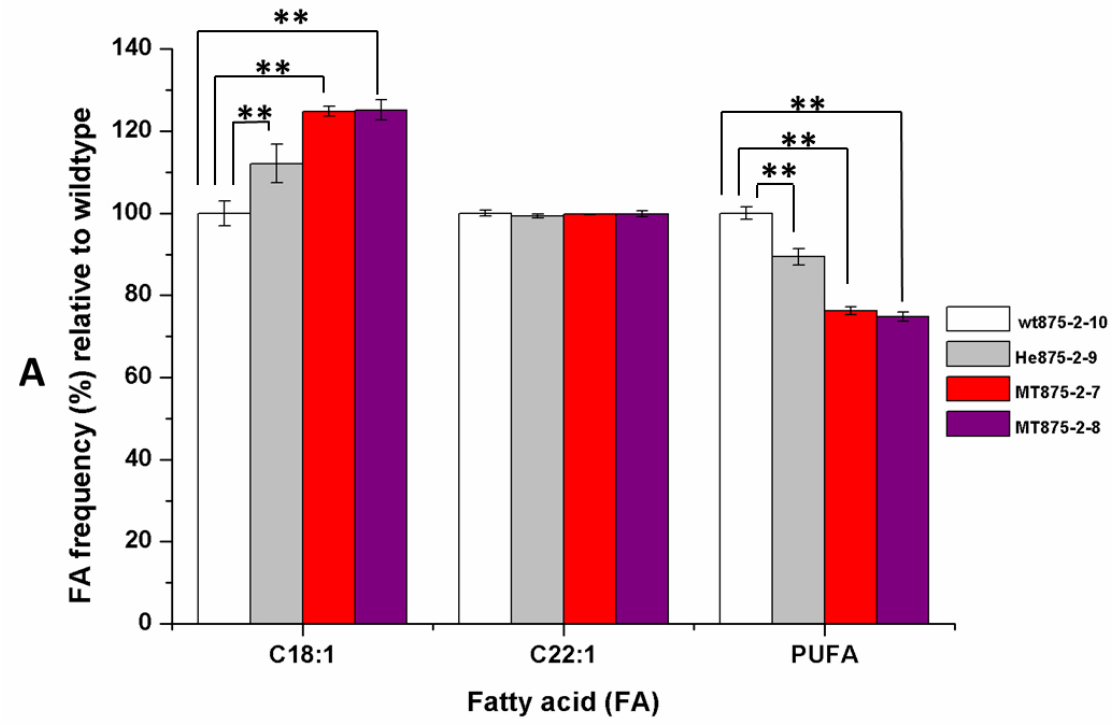
Two mutants (MT875-2-7 and MT875-2-8) that carry the homozygous nonsense mutation 875-mt4-c3 in *CaFAD2-C3* both showed an increase by 25 % (P <0.01) in C18:1 content compared to the wild-type and a concomitant 25 % decrease (P <0.01) in PUFA content (**Figure 5A**). The heterozygote (He875-2-9) for the same mutation showed an intermediate oil phenotype with changes in C18:1 and PUFA of respectively +12 % and −11 % (P <0.01) as compared to the wild-type (**Figure 5A**). Regarding the C22:1content, no difference was found between the mutants and wild-type (**Figure 5A**).

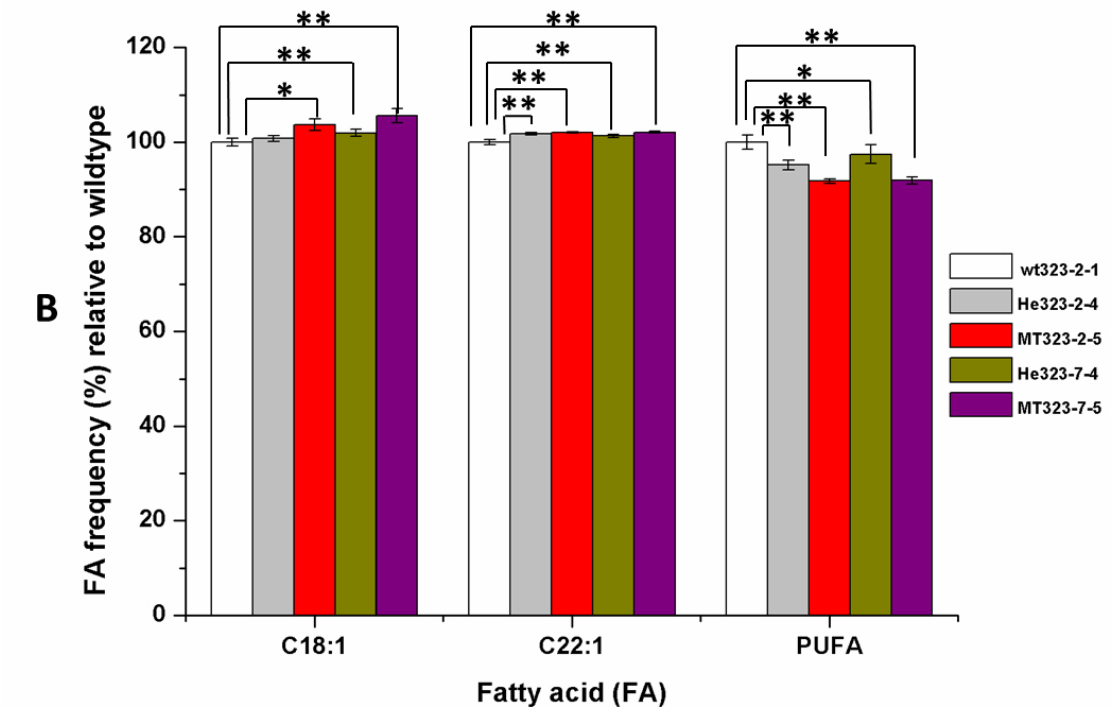
For the nonsense mutation in *CaFAD2-C1* (173-mt19-c1), only a single heterozygous mutant survived (He173-5-1). This line showed a wider range of oil composition than heterozygotes in the single seeds from other lines (**Supplementary Figure S6**). Since only mutation 875-mt4-c3 is nonsense and showed an effect on oil composition, the corresponding heterozygous mutant He875-2-6 was chosen as a control to compare with He173-5-1 for oil composition. It was shown that C18:1 and C22:1 contents in line He173-5-1 were significantly lower (−2 % and −7 % respectively; $P < 0.05$; $P <$ 0.01) compared to He875-2-6 but both mutants showed similar PUFA content.

The oil composition in lines carrying the homozygous missense mutation 323-mt21-c1 (Ala toVal, in *CaFAD2-C1*) showed minor but significant alterations in oil composition. The homozygous mutants for 323-mt21-c1 (MT323-2-5 and MT323-7-5) showed a detectable increase (3.7 % and 5.5 % respectively, $P < 0.01$) in C18:1 content and both showed a reduction (−8 %, $P < 0.01$) in PUFA content compared to the wide-type. For the two corresponding heterozygous mutants (He323-2-4 and He323-7-4), it was also shown that C18:1 content changed only in He323-7-4 (+2 %,

$P < 0.05$) together with the PUFA content changed in both mutants (−4.8 %, He323-2-4, $P < 0.01$ and −2.5 %, He323-7-4, $P < 0.05$) (**Figure 5B**).

Meanwhile, it is worth noting that these mutants showed a slight but consistent increase in C22:1 content (homozygotes +2 %, heterozygotes +1.5 %, $P < 0.01$) of the wild-type level (wt 64.2%, homozygous mt 65.5%), a change that is not observed in the *CaFAD2-C3* knockouts, carrying the nonsense mutation 875-mt4-c3 (**Figure 5B**). The remaining missense homozygous mutants for 200-mt10-c2, MT200-5-3 and MT200-5-6, both showed an oil composition typical for the wild-type, although the heterozygous mutant (He200-5-12) of this mutation did show a significant alteration in C18:1 (4.4 %, $P < 0.01$) and PUFA (−6 %, $P < 0.01$) contents compared to the wild-type. These mutants harbour amino acid changes in *CaFAD2-C2* (from Gly to Asp) (**Figure 5C**).


A
FA frequency (%) relative to wildtype
140
120
100
80
60
40
20
0
**
**
**
**
**
**
C18:1
C22:1
PUFA
Fatty acid (FA)
wt875-2-10
He875-2-9
MT875-2-7
MT875-2-8


B
FA frequency (%) relative to wildtype
120
100
80
60
40
20
0
**
**
*
**
**
**
**
*
**
**
C18:1
C22:1
PUFA
Fatty acid (FA)
wt323-2-1
He323-2-4
MT323-2-5
He323-7-4
MT323-7-5

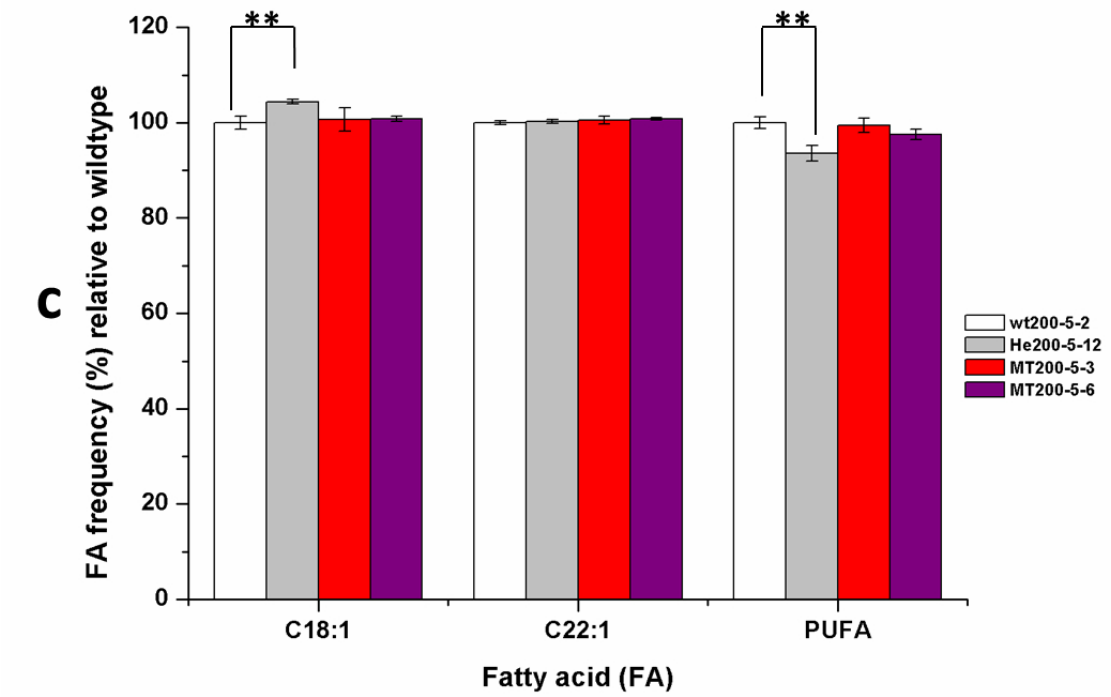


Figure 5. The fatty acid levels relative to the wild-type in the identified mutants for CaFAD2-C3 (a), -C1 (b) and -C2 (c). MT=homozygous mutants, He=heterozygous mutants, wt=homozygous wild-type. Twenty seeds were measured in bulk. *, **= Significance levels $P < 0.05$, $P < 0.01$

Discussion

Mutation breeding is an alternative to selection breeding and has been proved to be a tool to overcome limitations of natural genetic variation (Ahloowalia and Maluszynski, 2001; Ahloowalia et al., 2004). By either forward screens for phenotypes or reverse screens for mutations in target genes, agronomic traits have been improved in many crops (Alonso and Ecker, 2006; Caldwell et al., 2004; Himelblau et al., 2009; Le Signor et al., 2009; Men et al., 2002; Wu et al., 2005). In previous work, to obtain higher C18:1 and lower PUFA contents in crambe seeds, mutations were screened in the *CaFAD2* gene family by 454 sequencing in a fraction of the mutants (1,860 out of 12, 480 lines) of the same crambe population and ten mutations were found, however, these mutants showed similar oil compositions to the wild-type (Cheng et al., unpublished). This is telling that more promising mutants are required to obtain novel phenotypes in the seed oil. Illumina sequencing enabled to

screen almost the whole population (11,000 out of 12, 480 lines) for mutation detection in a short and conserved region of the *CaFAD2* genes.

Two strategies to mutation detection

The different crambe *FAD2* loci share over 89 % sequence identity, in particular the homology of the three expressed *FAD2* genes (*CaFAD2-C1*, *-C2* and *-C3*) is over 95 % in sequences (Cheng et al., 2013a). Distinguishing homeologous loci is already difficult with short sequencing reads (Griffin et al., 2011). Taken together with the presence of read errors and mutations in reads, it may be more difficult to match short reads to the corresponding loci by alignment. When using stringent parameters in alignments, mutant reads could be lost; otherwise reads cannot be distinguished for locus matching. To explore the feasibility of mutation detection without read alignments, a simple counting approach (approach 1) was used to estimate the frequencies of given mutations per pool (**Figure 1** and **2**). This approach needs no consideration of assigning reads to the single locus but relies on the frequency of a local mutant nucleotide motif (MT-motif, containing a putative mutation position in 9-mer nucleotides) in the total reads containing local wide-type (WT-) and MT-motifs per pool (freq=Nwt/(Nmt+Nwt)).

Estimation of mutation frequency by approach 1

In principle, if a heterzygote is present in a pool, the frequency of this mutation estimated by local motifs ranged from 1 % to 5 % depending on how many loci this motif matches to (1 % for five loci, 5 % for one locus). However, in practice the frequency of a mutation in a pool is hard to be estimated due to at least two factors: 1) the bias to different loci from potentially unequal amplification (Chen et al., 2012); 2) not all 10 genotypes in a pool are fully and equally represented by chance, it may be that the specific mutant is underrepresented so that the frequency of this mutation is lower than 1 %.

Fortunately, a previously confirmed mutation (960-mt30-c3 in **Table 1**) could be used as a control to evaluate the frequency of a real mutation. The frequency of the control

mutation is slightly over 0.2 % (**Figure 1**) and the local WT-motif of this mutation matches to three *CaFAD2-C* loci (**Supplementary Table 2**). The primers used for amplicon preparation are designed based on the sequence of *CaFAD2-C3* and thus more specific to *CaFAD2-C* loci (see **Results**: approach 2). Therefore, a frequency of 0.2 % can be used as a cut-off for mutation detection in this motif counting approach. The subsequent confirmation of the detected mutations proved the reliability of this approach (**Table 3**).

Although approach 1 with local matching motifs successfully detected mutations for the given putative positions, it is not efficient when all nucleotide positions need to be checked for potential mutations. Therefore, a new approach (approach 2) was made to check nucleotide transitions over the full read length of the CaFAD2 amplicon (**Figure 4**). This approach also requires no alignment but to allocate each paired-end read to the corresponding *CaFAD2* locus by reading each single read.

Importance of correct allocations for reads

To fully report nucleotide transitions in all positions, it is critical to allocate reads as possible as correctly to five crambe *FAD2* loci. Therefore, the allocation has taken into account not only information (the first 10 bases) of the forward read but also the 'edits' in the remaining sequence (behind the first 10 bases) for potential strand switches (**Figure 3**). Still, it is impossible for some reads to decide to which locus the remaining sequence (pair-end reads) belongs when low quality scores were present (nucleotide calls as N). In the case where after the position under investigation, no more differential SNPs can be found, the remaining sequence was allocated to the locus to which the read was primarily allocated. This may induce some erroneous allocations as it might be chimeric reads indeed. Consequently, this could lead to detecting false mutations which are actually nucleotides crossed from other crambe *FAD2* loci. More importantly, mutations detected in the switched regions may be assigned to the wrong locus. These reads with ambiguous allocations were discarded in the later analysis.

The false negative rate is harder to quantify because of the low number of potential mutations followed up. The mutants were identified for four mutations with exception of the nonsense mutation in *CaFAD2-C2* (303-mt19-c2, in **Table 3**). The reason why it failed to identify mutants for 303-mt19-c2 is probably that the M_2 plant to this mutation died when developing and subsequently no M_3 seeds were harvested. This reason can also explain why only a heterozygous mutant was found for 173-mt19-c1 (**Table 3**).

Insight into *FAD2* gene function for crambe oil improvement

Three crambe *FAD2* loci are expressed in the seeds, however the function of each of these three genes is still unclear (Cheng et al., 2013a). Suppressing expression of *CaFAD2* genes by RNAi suggested that *CaFAD2-C3* might play a relatively more important role in seed oil than the others (Cheng et al., 2013b). The nonsense mutants in *CaFAD2-C1* and *-C3* give a new insight into the functions of these two loci. The effects of the mutations in *CaFAD2-C1* and *-C3* indicate that both genes play roles in seed oil biosynthesis. It is known that *CaFAD2-C1* and *-C3* showed coordinated regulation on expression in the FAD2-RNAi crambe (Cheng et al., 2013b). It is interesting to investigate whether and how *CaFAD2-C1* expression is regulated in the *CaFAD2-C3* nonsense mutant, for instance knocking out of *CaFAD2-C3* could up-regulate expression of *CaFAD2-C1* or vice versa.

Meanwhile, the nonsense mutant of *CaFAD2-C3* showed no differences on other phenotypes in the greenhouse, e.g. seed germination (data not shown). Since only a heterozygous mutant of *CaFAD2-C1* was studied, it is unknown whether nonsense of *CaFAD2-C1* will result in negative effects on the plant. Crossings can be made between nonsense mutants of *CaFAD2-C1* and *-C3* to further enhance oil quality for crambe if neither of them shows negative agronomic traits. This was evident in the double *FAD2* mutant of soybean which showed a higher C18:1 content than the single gene mutants (Pham et al., 2010).

Supporting information

Table S1. The fusion primers used in this study for *CaFAD2* amplicon

Forward Primers	Sequence (5'-3')	Reverse Primers	Sequence (5'-3')
amp2_F1	ATCACGGGGTCTGCCAAGGCTGTGTCC	**amp2_R1**	GTAGCCCACGGTGCGTCCAAGAGGGT
amp2_F2	CGATGTGGGTCTGCCAAGGCTGTGTCC	**amp2_R2**	TACAAGCACGGTGCGTCCAAGAGGGT
amp2_F3	TTAGGCGGGTCTGCCAAGGCTGTGTCC	**amp2_R3**	TTGACTCACGGTGCGTCCAAGAGGGT
amp2_F4	TGACCAGGGTCTGCCAAGGCTGTGTCC	**amp2_R4**	GGAACTCACGGTGCGTCCAAGAGGGT
amp2_F5	ACAGTGGGGTCTGCCAAGGCTGTGTCC	**amp2_R5**	TGACATCACGGTGCGTCCAAGAGGGT
amp2_F6	GCCAATGGGTCTGCCAAGGCTGTGTCC	**amp2_R6**	GGACGGCACGGTGCGTCCAAGAGGGT
amp2_F7	CAGATCGGGTCTGCCAAGGCTGTGTCC	**amp2_R7**	CTCTACCACGGTGCGTCCAAGAGGGT
amp2_F8	ACTTGAGGGTCTGCCAAGGCTGTGTCC	**amp2_R8**	GCGGACCACGGTGCGTCCAAGAGGGT
amp2_F9	GATCAGGGGTCTGCCAAGGCTGTGTCC	**amp2_R9**	TTTCACCACGGTGCGTCCAAGAGGGT
amp2_F10	TAGCTTGGGTCTGCCAAGGCTGTGTCC	**amp2_R10**	GGCCACCACGGTGCGTCCAAGAGGGT
amp2_F11	GGCTACGGGTCTGCCAAGGCTGTGTCC	**amp2_R11**	CGAAACCACGGTGCGTCCAAGAGGGT
amp2_F12	CTTGTAGGGTCTGCCAAGGCTGTGTCC	**amp2_R12**	GAATGACACGGTGCGTCCAAGAGGGT
amp2_F13	AGTCAAGGGTCTGCCAAGGCTGTGTCC	**amp2_R13**	CCACTCCACGGTGCGTCCAAGAGGGT
amp2_F14	AGTTCCGGGTCTGCCAAGGCTGTGTCC	**amp2_R14**	GCTACCCACGGTGCGTCCAAGAGGGT
amp2_F15	ATGTCAGGGTCTGCCAAGGCTGTGTCC	**amp2_R15**	ATCAGTCACGGTGCGTCCAAGAGGGT
amp2_F16	CCGTCCGGGTCTGCCAAGGCTGTGTCC	**amp2_R16**	GCTCATCACGGTGCGTCCAAGAGGGT
amp2_F17	GTCCGCGGGTCTGCCAAGGCTGTGTCC	**amp2_R17**	AGGAATCACGGTGCGTCCAAGAGGGT
amp2_F18	GTGAAAGGGTCTGCCAAGGCTGTGTCC	**amp2_R18**	CTTTTGCACGGTGCGTCCAAGAGGGT
amp2_F19	GTGGCCGGGTCTGCCAAGGCTGTGTCC	**amp2_R19**	TAGTTGCACGGTGCGTCCAAGAGGGT
amp2_F20	GTTTCGGGGTCTGCCAAGGCTGTGTCC	**amp2_R20**	CCGGTGCACGGTGCGTCCAAGAGGGT
amp2_F21	CGTACGGGGTCTGCCAAGGCTGTGTCC	**amp2_R21**	ATCGTGCACGGTGCGTCCAAGAGGGT
amp2_F22	GAGTGGGGGTCTGCCAAGGCTGTGTCC	**amp2_R22**	TGAGTGCACGGTGCGTCCAAGAGGGT
amp2_F23	ACTGATGGGTCTGCCAAGGCTGTGTCC	**amp2_R23**	CGCCTGCACGGTGCGTCCAAGAGGGT
amp2_F24	ATTCCTGGGTCTGCCAAGGCTGTGTCC	**amp2_R24**	GCCATGCACGGTGCGTCCAAGAGGGT
amp2_F25	CGTGATGGGTCTGCCAAGGCTGTGTCC	**amp2_R25**	AAAATGCACGGTGCGTCCAAGAGGGT
amp2_F26	ACATCGGGGTCTGCCAAGGCTGTGTCC	**amp2_R26**	TGTTGGCACGGTGCGTCCAAGAGGGT
amp2_F27	GCCTAAGGGTCTGCCAAGGCTGTGTCC	**amp2_R27**	ATTCCGCACGGTGCGTCCAAGAGGGT
amp2_F28	TGGTCAGGGTCTGCCAAGGCTGTGTCC	**amp2_R28**	AGCTAGCACGGTGCGTCCAAGAGGGT
amp2_F29	CACTGTGGGTCTGCCAAGGCTGTGTCC	**amp2_R29**	GTATAGCACGGTGCGTCCAAGAGGGT
amp2_F30	ATTGGCGGGTCTGCCAAGGCTGTGTCC	**amp2_R30**	TCTGAGCACGGTGCGTCCAAGAGGGT
amp2_F31	GATCTGGGGTCTGCCAAGGCTGTGTCC	**amp2_R31**	GTCGTCCACGGTGCGTCCAAGAGGGT
amp2_F32	TCAAGTGGGTCTGCCAAGGCTGTGTCC	**amp2_R32**	CGATTACACGGTGCGTCCAAGAGGGT
amp2_F33	CTGATCGGGTCTGCCAAGGCTGTGTCC	**amp2_R33**	GCTGTACACGGTGCGTCCAAGAGGGT
amp2_F34	AAGCTAGGGTCTGCCAAGGCTGTGTCC	**amp2_R34**	ATTATACACGGTGCGTCCAAGAGGGT

Table S2. The list of motifs used in motif counting

Putative positions	mutation	Wild-type motifs	mutant motifs	Remarks*
MT_1		CTAACCGGC	CTAA*T*CGGC	CaFAD2-C3
MT_2		ACCGGCATC	ACCG*A*CATC	CaFAD2-C3
MT_3		ATCTGGGTC	ATCT*A*GGTC	CaFAD2-C1, -C2 and -C3
MT_4		ATCTGGGTC	ATCTG*A*GTC	CaFAD2-C1, -C2 and -C3
MT_5		ATAGCTCAC	ATAG*T*TCAC	CaFAD2-C1, -C2 and -C3
MT_6		GCTCACGAA	GCT*T*ACGAA	CaFAD2-C1, -C2 and -C3
MT_7		CACGAATGT	CAC*A*AATGT	CaFAD2-C1, -C2 and -C3
MT_8		GAATGTGGC	GAAT*A*TGGC	CaFAD2-C1, -C2 and -C3
MT_9		TGTGGCCAC	TGT*A*GCCAC	CaFAD2-C1, -C2 and -C3
MT_10		TGTGGCCAC	TGTG*A*CCAC	CaFAD2-C1, -C2 and -C3
MT_11		GGCCACCAT	GGCT*A*CCAT	CaFAD2-C1, -C2 and -C3
MT_12		CATGCTTTC	CATG*T*TTTC	CaFAD2-C1, -C2 and -C3
MT_13		GCTTTCAGC	GCTTTCA*A*C	CaFAD2-C1, -C2 and -C3
MT_14		TACCAATGG	TAC*T*AATGG	CaFAD2-C1, -C2 and -C3
MT_15		CAATGGCTA	CAAT*A*GCTA	CaFAD2-C1, -C2 and -C3
MT_16		CAATGGCTA	CAATG*A*CTA	CaFAD2-C1, -C2 and -C3
MT_17		CTAACTGGC	CTAA*T*TGGC	CaFAD2-C2
MT_18		ACTGGCGTC	ACTG*A*CGTC	CaFAD2-C2
MT_19		GTCTGGGTC	GTCT*A*GGTC	CaFAD2-C1 and -C2
MT_20		GTCTGGGTC	GTCTG*A*GTC	CaFAD2-C1 and -C2
MT_21		ATAGCCCAC	ATAG*T*CCAC	CaFAD2-C1 and -C2
MT_22		GCCCACGAA	GCC*T*ACGAA	CaFAD2-C1 and -C2
MT_23		CACGAATGT	CAC*A*AATGT	CaFAD2-C1 and -C2
MT_24		CTAACAGGC	CTAA*T*AGGC	CaFAD2-C1
MT_25		ACAGGCGTC	ACAG*A*CGTC	CaFAD2-C1
MT_26		CTTCTTAGG	CTTCTTAG*A*	CaFAD2-C1, -C2 and -C3
MT_27		CTTCTTAGG	CTTCTTA*A*G	CaFAD2-C1, -C2 and -C3
MT_28		GTACCACTT	GTAC*T*ACTT	CaFAD2-C1, -C2 and -C3
MT_29		GTACCACTT	GTA*T*CACTT	CaFAD2-C1, -C2 and -C3
MT_30		GATGTCTGA	GAT*A*TCTGA	CaFAD2-C1, -C2 and -C3

* motifs specific to which of CaFAD2 gene(s)

Table S3 The primers used for mutant identification by Sanger sequencing

	amplification		sequencing	
Amplicon	**Forward (5'-3')**	**Reverse (5'-3')**	**Forward (5'-3')**	**Reverse (5'-3')**
CaFAD2-C1	GCTATCGTTTATTTATTTTTCTTTC	AGCGATGAGAAGAACAATACAGAGA	GCTATCGTTTATTTATTTTTCTTTC	CACGGTGCGTCCAAGAGGGT
CaFAD2-C2	CTCAACGCTATCGTTTATTTCTTTC	AGATAAAAGTAGCTTCACAGG	CTCAACGCTATCGTTTATTTCTTTC	CACGGTGCGTCCAAGAGGGT
CaFAD2-C3	CCACGTACTATCCATTTTTGAAAGT	AACACAACACAATGGATAATT	CCACGTACTATCCATTTTTGAAAGT	AACACAACACAATGGATAATT

Figure S1 Prediction of EMS-type transition (G to A and C to T) in the partial sequence of CaFAD2-C3. The positions of primers were highlighted. The wild-type sequence encodes the amino acids above the line, and the amino acid transitions from EMS mutations were shown below the line. The red amino acids and * indicate the changes to severe missense and truncation.

```
Y   F   S   L   L   P   H   S   L   S   Y   L   A   W   P   L   Y   W      90
tac ttc tct ctc ctc cct cac tct ctc tct tac tta gct tgg cct ctc tac tgg    270
  =   =  F   F = F = SL  Y =  F   F =  F    =     TV   ** SL  F =   =  **

V   C   Q   G   C   V   L   T   G   I   W   V   I   A   H   E   C   G      108
gtc tgc caa ggc tgt gtc cta acc ggc atc tgg gtc ata gct cac gaa tgt ggc    324
I =  Y= *   SD=  Y  I = =    I= SD=   =  ** I =     TV  Y = K    Y  SD=

                                                IPB010257C (8.3e-38) IC 1.71
H   H   A   F   S   D   Y   Q   W   L   D   D   T   V   G   L   I   F      126
cac cat gct ttc agc gac tac caa tgg cta gac gac acc gtt ggt ctt atc ttc    378
Y = Y   TV    =  N= N =   = *    ** =   N = N =  I= I   SD  F     =   =

H   S   F   L   L   V   P   Y   F   S   W   K   Y   S   H   R   R   H      144
cat tct ttc ctc ctc gtc cct tac ttc tcc tgg aaa tac agt cat cgc cgt cac    432
Y    F    = F = F = I = SL    =   =  F=  **       =  N  Y   CH= CH  Y =

H   S   N   T   G   S   L   E   R   D   E   V   F   V   P   K   K   K      162
cat tcc aac act gga tca ttg gaa aga gat gaa gtg ttt gtc cct aag aag aaa    486
Y    F=   =  I   RE   L    = K    K   N   K   M =     I = SL    =   =

S   D   I   K   W   Y   G   K   Y   L   N   N   P   L   G   R   T   V      180
tca gac atc aag tgg tac ggc aaa tac ctc aac aac cct ctt gga cgc acc gtg    540
 L  N =   =   =  **   = SD=       = F =   =   = SL  F   RE  CH=  I= M =
```

Figure S2. Natural variants in the reads among five FAD2 genes. Yellow highlights=positions of primer

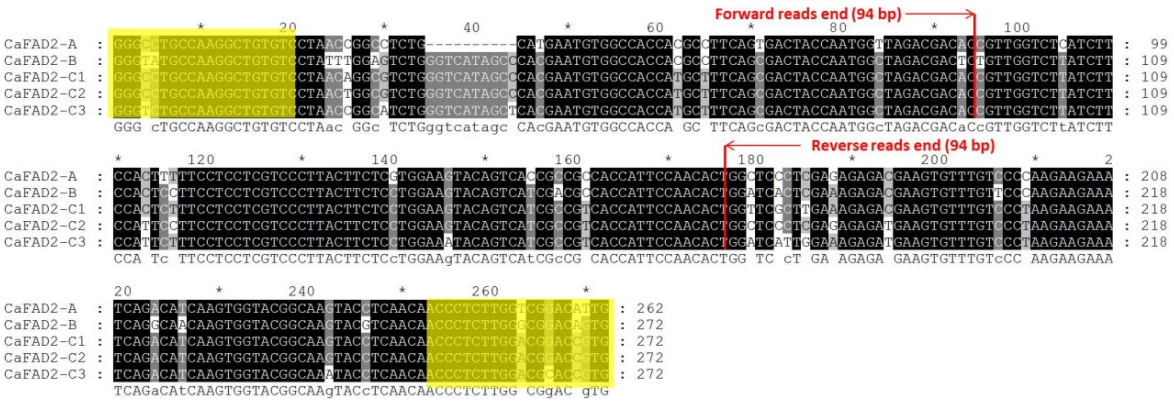


Figure S3. The histograms of reads numbers (A) and coverage (B) in the Illumina sequencing

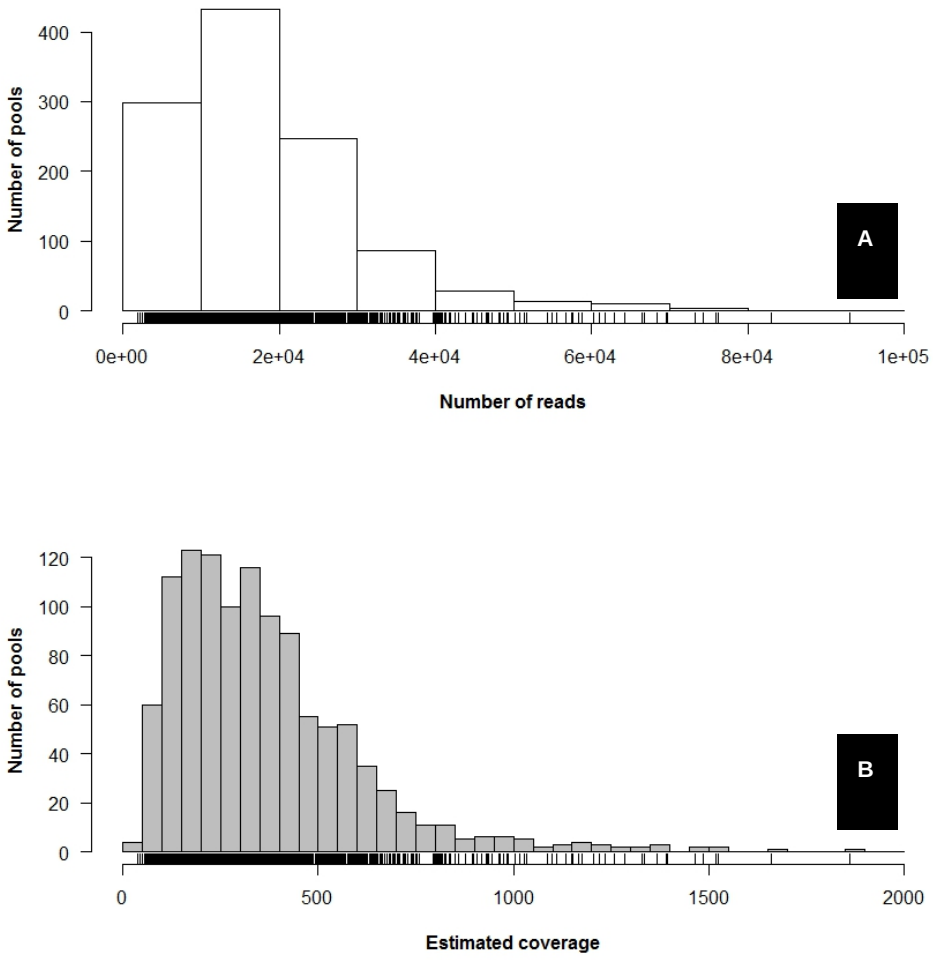


Figure S4. The locations of mutations found by the approach 2. Mutations are indicated by pool number; gene-position; amino acid change.

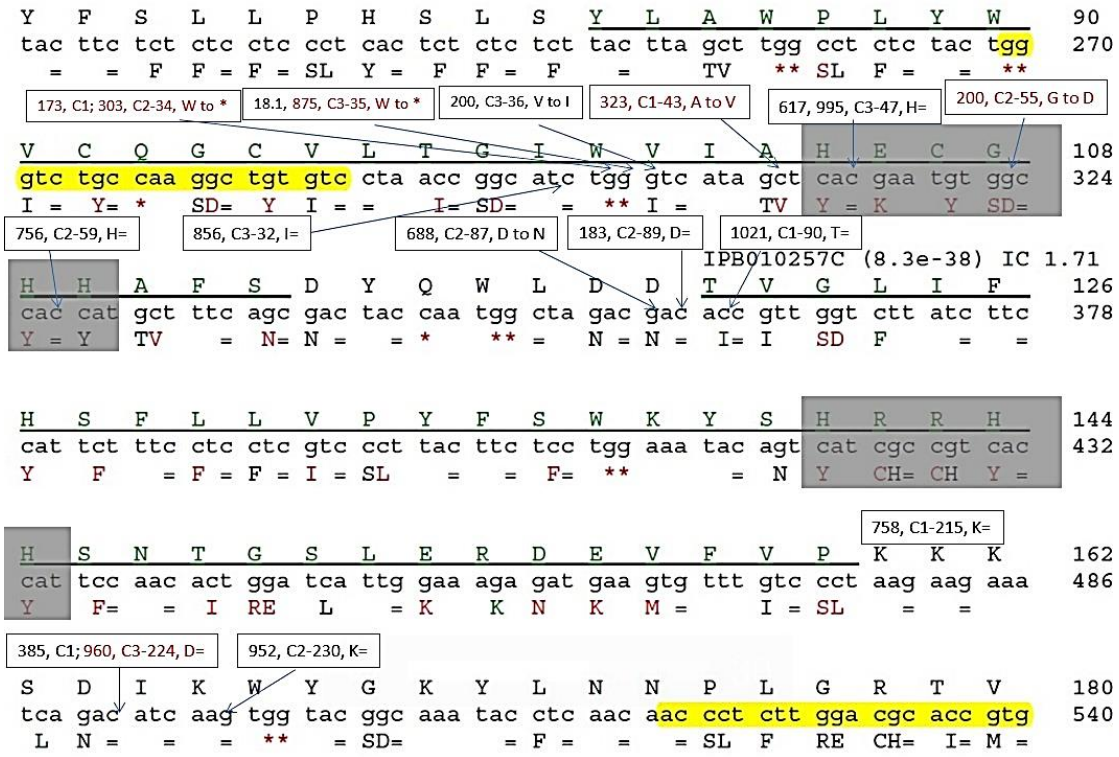


Figure S5. Mutant identification by Sanger sequencing. Red=homozygous mutants; grey=homozygous wild-type.

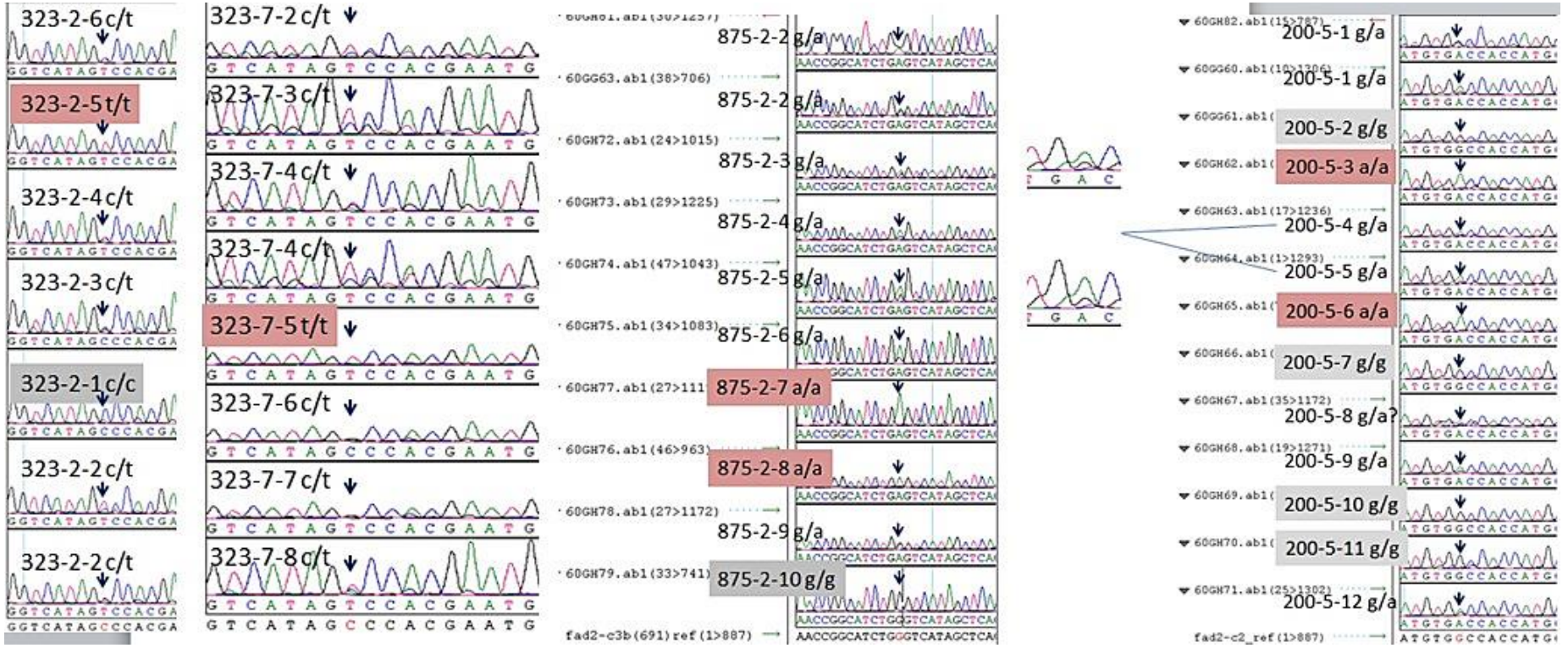
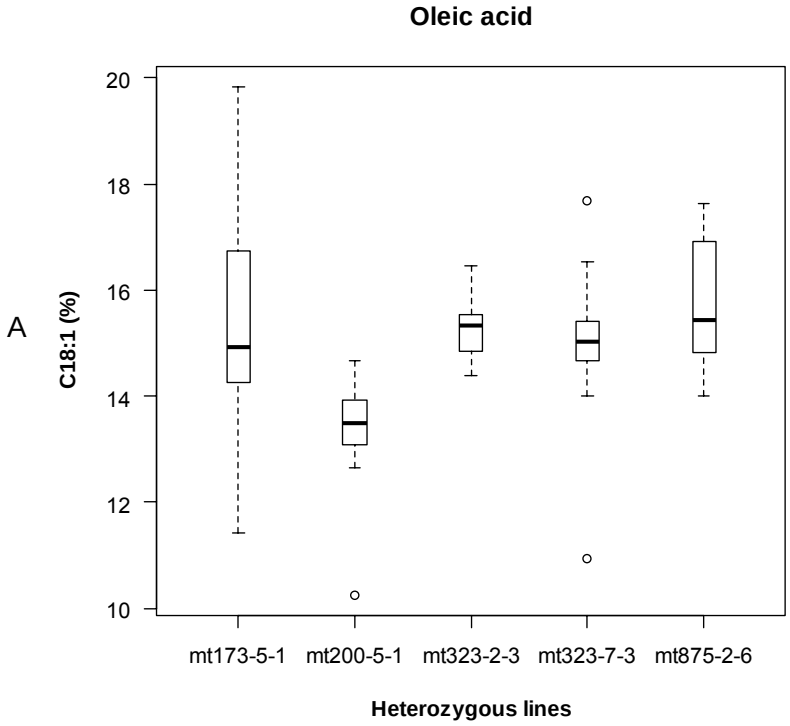
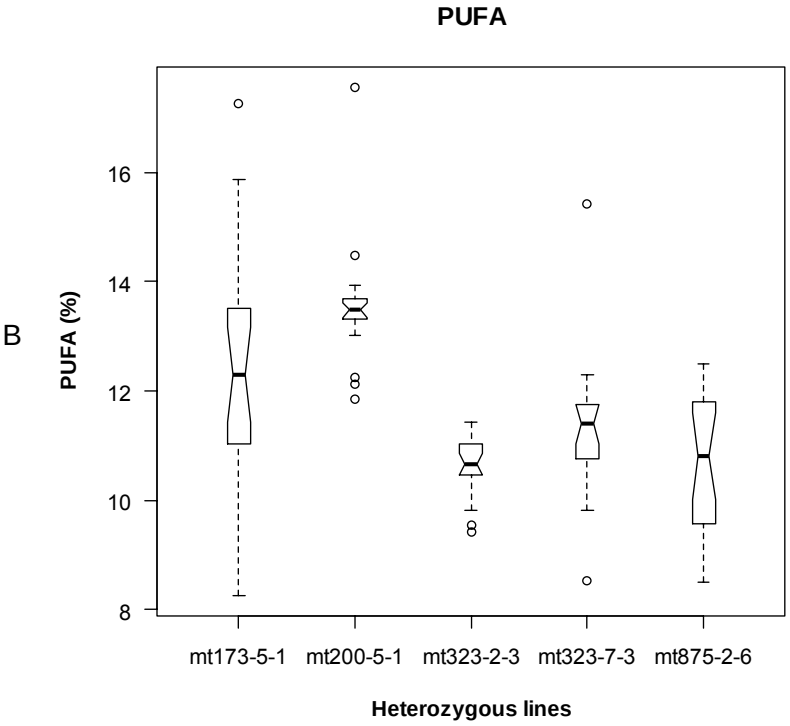
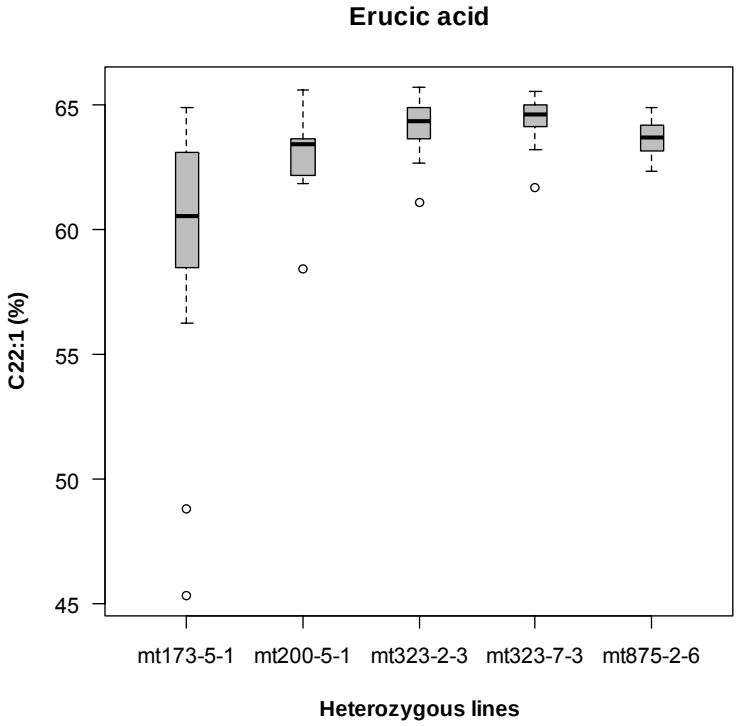


Figure S6 The contents of C18:1 (a), PUFA (b) and C22:1 (c) in single seeds of the heterozygote mutants





C


Erucic acid
C22:1 (%)
65
60
55
50
45
mt173-5-1
mt200-5-1
mt323-2-3
mt323-7-3
mt875-2-6
Heterozygous lines

References

Abel, H.J., Duncavage, E.J., Becker, N., Armstrong, J.R., Magrini, V.J. and Pfeifer, J.D. (2010) SLOPE: a quick and accurate method for locating non-SNP structural variation from targeted next-generation sequence data. Bioinformatics 26, 2684-2688.

Ahloowalia, B. and Maluszynski, M. (2001) Induced mutations-A new paradigm in plant breeding. Euphytica 118, 167-173.

Ahloowalia, B., Maluszynski, M. and Nichterlein, K. (2004) Global impact of mutation-derived varieties. Euphytica 135, 187-204.

Alonso, J.M. and Ecker, J.R. (2006) Moving forward in reverse: genetic technologies to enable genome-wide phenomic screens in Arabidopsis. Nature Reviews Genetics 7, 524-536.

Backes, G. (2013) TILLING and EcoTILLING. In: Diagnostics in Plant Breeding (Varshney, R.K. ed) pp. 145-165. Springer.

Beló, A., Zheng, P., Luck, S., Shen, B., Meyer, D.J., Li, B., Tingey, S. and Rafalski, A. (2008) Whole genome scan detects an allelic variant of fad2 associated with increased oleic acid levels in maize. Molecular Genetics and Genomics 279, 1-10.

Brady, S.M. and Provart, N.J. (2007) Extreme breeding: Leveraging genomics for crop improvement. Journal of the Science of Food and Agriculture 87, 925-929.

Bundock, P.C., Eliott, F.G., Ablett, G., Benson, A.D., Casu, R.E., Aitken, K.S. and Henry, R.J. (2009) Targeted single nucleotide polymorphism (SNP) discovery in a highly polyploid plant species using 454 sequencing. Plant Biotechnology Journal 7, 347-354.

Caldwell, D.G., McCallum, N., Shaw, P., Muehlbauer, G.J., Marshall, D.F. and Waugh, R. (2004) A structured mutant population for forward and reverse genetics in Barley (Hordeum vulgare L.). The Plant Journal 40, 143-150.

Chen, X., Listman, J.B., Slack, F.J., Gelernter, J. and Zhao, H. (2012) Biases and errors on allele frequency estimation and disease association tests of next-generation sequencing of pooled samples. Genetic Epidemiology 36, 549-560.

Cheng, J., Salentijn, E.M.J., Huang, B., Krens, F.A., Dechesne, A.C., Visser, R.G.F. and van Loo, E.N. (2013a) Isolation and characterization of the omega-6 fatty acid desaturase (FAD2) gene family in the allohexaploid oil seed crop Crambe abyssinica Hochst. Molecular Breeding 32, 517-531.

Cheng, J., Zhu, L., Salentijn, E.M., Huang, B., Gruber, J., Dechesne, A.C., Krens, F.A., Qi, W., Visser, R.G. and van Loo, E.N. (2013b) Functional analysis of the omega-6 fatty acid desaturase (CaFAD2) gene family of the oil seed crop Crambe abyssinica. BMC plant biology 13, 146.

Colbert, T., Till, B.J., Tompa, R., Reynolds, S., Steine, M.N., Yeung, A.T., McCallum, C.M., Comai, L. and Henikoff, S. (2001) High-throughput screening for induced point mutations. Plant Physiology 126, 480-484.

Comai, L. and Henikoff, S. (2006) TILLING: practical single-nucleotide mutation discovery. The Plant Journal 45, 684-694.

Comai, L., Young, K., Till, B.J., Reynolds, S.H., Greene, E.A., Codomo, C.A., Enns, L.C., Johnson, J.E., Burtner, C. and Odden, A.R. (2004) Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. The Plant Journal 37, 778-786.

Gady, A.L.F., Vriezen, W.H., Van de Wal, M.H.B.J., Huang, P., Bovy, A.G., Visser, R.G.F. and Bachem, C.W.B. (2012) Induced point mutations in the phytoene synthase 1 gene cause differences in carotenoid content during tomato fruit ripening. Molecular Breeding 29, 801-812.

Gilchrist, E.J., O'Neil, N.J., Rose, A.M., Zetka, M.C. and Haughn, G.W. (2006) TILLING is an effective reverse genetics technique for Caenorhabditis elegans. BMC genomics 7, 262.

Griffin, P.C., Robin, C. and Hoffmann, A.A. (2011) A next-generation sequencing method for overcoming the multiple gene copy problem in polyploid phylogenetics, applied to Poa grasses. BMC Biology 9, 19.

Himelblau, E., Gilchrist, E.J., Buono, K., Bizzell, C., Mentzer, L., Vogelzang, R., Osborn, T., Amasino, R.M., Parkin, I.A. and Haughn, G.W. (2009) Forward and reverse genetics of rapid-cycling Brassica oleracea. Theoretical and Applied Genetics 118, 953-961.

Le Signor, C., Savois, V., Aubert, G., Verdier, J., Nicolas, M., Pagny, G., Moussy, F., Sanchez, M., Baker, D. and Clarke, J. (2009) Optimizing TILLING populations for reverse genetics in Medicago truncatula. Plant Biotechnology Journal 7, 430-441.

Marroni, F., Pinosio, S., Di Centa, E., Jurman, I., Boerjan, W., Felice, N., Cattonaro, F. and Morgante, M. (2011) Large-scale detection of rare variants via pooled multiplexed next-generation sequencing: towards next-generation Ecotilling. The Plant Journal 67, 736-745.

Men, A.E., Laniya, T.S., Searle, I.R., Iturbe-Ormaetxe, I., Gresshoff, I., Jiang, Q., Carroll, B.J. and Gresshoff, P.M. (2002) Fast neutron mutagenesis of soybean (Glycine soja L.) produces a supernodulating mutant containing a large deletion in linkage group H. Genome Letters 1, 147-155.

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

Parry, M.A., Madgwick, P.J., Bayon, C., Tearall, K., Hernandez-Lopez, A., Baudo, M., Rakszegi, M., Hamada, W., Al-Yassin, A. and Ouabbou, H. (2009) Mutation discovery for crop improvement. Journal of Experimental Botany 60, 2817-2825.

Pham, A.T., Lee, J.D., Shannon, J.G. and Bilyeu, K.D. (2010) Mutant alleles of FAD2-1A and FAD2-1B combine to produce soybeans with the high oleic acid seed oil trait. BMC Plant Biology 10, 195.

Pham, A.T., Lee, J.D., Shannon, J.G. and Bilyeu, K.D. (2011) A novel FAD2-1 A allele in a soybean plant introduction offers an alternate means to produce soybean seed oil with 85% oleic acid content. TAG Theoretical and Applied Genetics 123, 793-802.
Qiu, X., Wu, L., Huang, H., McDonel, P.E., Palumbo, A.V., Tiedje, J.M. and Zhou, J. (2001) Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based cloning. Applied and Environmental Microbiology 67, 880-887.
Rigola, D., van Oeveren, J., Janssen, A., Bonné, A., Schneiders, H., van der Poel, H.J.A., Van Orsouw, N.J., Hogers, R.C.J., De Both, M.T.J. and Van Eijk, M.J.T. (2009) High-throughput detection of induced mutations and natural variation using KeyPoint™ technology. PLoS One 4, e4761.
Schloss, P.D., Gevers, D. and Westcott, S.L. (2011) Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PloS One 6, e27310.
Sikora, P., Chawade, A., Larsson, M., Olsson, J. and Olsson, O. (2011) Mutagenesis as a tool in plant genetics, functional genomics, and breeding. International Journal of Plant Genomics 2011.
Slade, A.J., Fuerstenberg, S.I., Loeffler, D., Steine, M.N. and Facciotti, D. (2005) A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. Nature Biotechnology 23, 75-81.
Slade, A.J. and Knauf, V.C. (2005) TILLING moves beyond functional genomics into crop improvement. Transgenic Research 14, 109-115.
Stemple, D.L. (2004) TILLING-a high-throughput harvest for functional genomics. Nature Reviews Genetics 5, 145-150.
Tanhuanpää, P., Vilkki, J. and Vihinen, M. (1998) Mapping and cloning of FAD2 gene to develop allele-specific PCR for oleic acid in spring turnip rape (Brassica rapa ssp. oleifera). Molecular Breeding 4, 543-550.
Till, B.J., Comai, L. and Henikoff, S. (2007) TILLING and EcoTILLING for crop improvement. In: Genomics-assisted crop improvement pp. 333-349. Springer.
Tsai, H., Howell, T., Nitcher, R., Missirian, V., Watson, B., Ngo, K.J., Lieberman, M., Fass, J., Uauy, C., Tran, R.K., Khan, A.A., Filkov, V., Tai, T.H., Dubcovsky, J. and Comai, L. (2011) Discovery of rare mutations in populations: Tilling by sequencing. Plant Physiology 156, 1257-1268.
Vandenbroucke, I., Van Marck, H., Verhasselt, P., Thys, K., Mostmans, W., Dumont, S., Van Eygen, V., Coen, K., Tuefferd, M. and Aerssens, J. (2011) Minor variant detection in amplicons using 454 massive parallel pyrosequencing: experiences and considerations for successful applications. BioTechniques 51, 167-177.
Varshney, R.K., Nayak, S.N., May, G.D. and Jackson, S.A. (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends in Biotechnology 27, 522-530.
Wells, R., Trick, M., Fraser, F., Soumpourou, E., Clissold, L., Morgan, C., Pauquet, J., Bancroft, I. and Mondonville, F. (2013) Sequencing-based variant detection in the polyploid crop oilseed rape. BMC Plant Biology 13, 1-11.
Winkler, S., Schwabedissen, A., Backasch, D., Bökel, C., Seidel, C., Bönisch, S., Fürthauer, M., Kuhrs, A., Cobreros, L. and Brand, M. (2005) Target-selected mutant screen by TILLING in Drosophila. Genome Research 15, 718-723.
Wu, J.-L., Wu, C., Lei, C., Baraoidan, M., Bordeos, A., Madamba, M.R.S., Ramos-Pamplona, M., Mauleon, R., Portugal, A. and Ulat, V.J. (2005) Chemical-and irradiation-induced mutants of indica rice IR64 for forward and reverse genetics. Plant Molecular Biology 59, 85-97.

Chapter 6

General discussion

Our life nowadays heavily relies on petroleum because it is not only the source of energy for e.g. transportation, but also provides industrial feedstocks which are mainly derivatives of petroleum. A forthcoming problem is an expected shortage of available petroleum as a consequence of the constant growth of the human population and finite nature of petroleum reserves on Earth. As a solution to this problem, scientists have looked for other natural resources to replace petroleum. Some plant species, so-called industrial oil crops, produce oils that have similar characteristics as petroleum and are considered to be an interesting sustainable alternative to petroleum. These industrial oil crops include not only some currently utilized major food oil crops such as rapeseed, soybean, maize and sunflower but also some minor oil crops such as *Crambe abyssnica* (Cheng et al., 2013; Dyer et al., 2008; Vanhercke et al., 2013). Of them, high erucic acid rapeseed (HEAR) cultivars and crambe both can be used to produce industrial oils with a high content of erucic acid (C22:1). Erucic acid is a source of industrial feedstock but at the same time is an anti-nutritional compound which is not accepted in food oils (Vogtmann et al., 1973).

Industrial oil crops have to be grown in restricted regions to prevent outcrossing with food oil cultivars, through which industrial oil can enter the food chain. As an advantage, crambe plants naturally cannot cross with related food oil crops, as is the case in rapeseed compared to HEAR (Rudloff and Wang, 2011).

Crambe oil contains substantial amounts of C22:1 (over 60 %) and a fraction of other fatty acids: oleic acid (C18:1, 18 %), linoleic acid (C18:2, 9 %), linolenic acid (C18:3, 7 %) and others (**Table 1, Chapter 2**). For application, purification has to be carried out to extract C22:1 from the raw oil. Obviously, higher C22:1 content in oil is beneficial to reduce the cost of purification. It is reported that the 10 % increase in C22:1 content in rapeseed would reduce the purification cost by a half (Jadhav et al., 2005). Meanwhile, the polyunsaturated fatty acids (PUFA) in crambe oil, C18:2 and C18:3, are prone to oxidation under storage thus negatively affect the life of shelf for crambe oil. Therefore, it is encouraged to improve the value of crambe oil by elevating C22:1 content or/and reducing the content of PUFA.

Breeding strategies for *C. abyssinica*: genetic modification or mutation breeding

Breeding for C. abyssinica with high C18:1 content

All breeding strategies rely on natural or artificial variation. To breed crops with desirable phenotypes, various breeding strategies can be employed. Traditional selection is applied to

select improved or novel phenotypes based on natural variation in accessions. Although phenotypic selection is labor intensive and particularly relies on natural variation, it is widely used in plant breeding (Rommens et al., 2007). Initially, to explore the natural variation in *C. abyssinica*, we investigated the oil composition among several crambe breeding lines under field and greenhouse conditions. Under both conditions, these lines showed minor but significant variations in C18:1 and C18:3 contents (Table 1, Chapter 2). Next, we assumed that these variations are linked to either genetic- or expression variation of the fatty acid desaturase 2 (FAD2) gene, which is associated to the contents of C18:1 and PUFA in plants (Okuley et al., 1994). However, in the FAD2 gene there is no genetic variation found among these crambe lines and only limited variation in gene expression; furthermore, the variation in expression of FAD2 cannot explain the minor but significant variation in oil composition (Figure 4, Chapter 2). This might reflect that natural variation in C. Abyssinica is limited, which fits with earlier claims for this crop (Lessman, 1975; Mastebroek et al., 1994).

Selection in breeding programmes is based on the fact that genetic variation is induced at the DNA level but only screened phenotypically. Variation can be introduced into candidate species from its wild relatives by intergenic hybridization (Rommens et al., 2007; Wang and Luo, 1998; Wang et al., 2006; Wang et al., 2003). An example is the somatic hybridzation between *C. abyssinica* and *Brassica napus*, two members of *Brassicaceae* family, that introduced variation of the FAE1 gene into B. napus (Wang et al., 2004). However, it is known that intergenic crossing or somatic hybridization is normally difficult and the phenotypes of the offspring are unpredictable; furthermore, tissue culture is usually required to establish the offspring (Griga et al., 1995; Mohan Jain, 2001). Therefore, it is hard to gain desirable phenotypes by applying these breeding approaches for crambe improvement.

Candidate gene transformation

When a certain phenotype is targeted and the genes related to this phenotype are known, a straight-forward way to establish the phenotype is by over-expression or silencing the target genes. In many studies, it is evident that the phenotype of "higher C18:1, lower PUFA" in plant oils is related to the omega-6 fatty acid desaturase 2 (*FAD2*) gene (Hu et al., 2006; Liu et al., 2002a; Pirtle et al., 2001; Stoutjesdijk et al., 2002). In *C. abyssinica*, the content of C18:1 was elevated by over 8 % (from 12 % to 20 %) together with a reduction of PUFA by over 6 % (from 12 % to 5 %) in seed oil as a consequence of silencing crambe *FAD2* expression (**Figure 4, Chapter 3**). To improve the content of C22:1 in crambe, overexpression of fatty acid elongase (*FAE1*) in combination with silencing *CaFAD2* enabled

to increase the contents of both C18:1 and C22:1 (Li et al., 2012). Undoubtedly, these genetic engineering approaches are promising for crambe if genetically modified organisms (GMO) are widely accepted all over the world which at the moment is not yet the case.

Crambe oil improvement by mutation breeding

To introduce variation to *C. abyssinica* by a non-GMO approach, chemical mutagenesis was applied to crambe cultivar 'Galactica' (**Chapters 4 and 5**). Mutagenesis is an important tool to generate genetic variation for crop improvement and is particularly valuable in those crops with limited natural diversity (Parry et al., 2009). During the past few decades, more than 70 cultivars derived from mutation breeding by the use of chemical mutagen or irradiation have been released for various crops (Ahloowalia and Maluszynski, 2001; Ahloowalia et al., 2004). Furthermore, polyploid genomes are tolerant to mutation accumulation because of genetic redundancy (Backes, 2013). Therefore, as a hexaploid, *C. abyssinica* is a perfect candidate to apply mutagenesis. In this study, a mutant population (over 12,000 mutants) was generated in the high yielding cultivar 'Galactica', using the mutagen ethyl methane sulphonate (EMS) and indeed mutants with improved oil quality were found within this population (**Chapters 4 and 5**). This population harbors at maximum, one mutation per 140 kb and potentially can be used as a source for crambe improvement. Alternative to screening mutations from M_2 generation, mutations can also be found within M_1 generation as applied in tomato and potato breeding (Okabe et al., 2013; Uitdewilligen, 2012). Obviously, this strategy needs less mutant individuals to recover mutations generated compared to the use of M_2 generations; however, the individuals arising from the mutagenized seeds (the M_1 generation) may be chimeric for the introduced mutations. So, a M_1 generation is usually selfed to obtain homozygous M_2 mutants, which is a problem especially for self-incompatible crops. We applied the M_2 generation of 1000 M_1 lines for mutation detection and several homozygous mutants were identified which can be directly used to develop crambe cultivars.

The isolation and functional analysis of crambe FAD2 genes

The gene *FAD2* is associated with the contents of C18:1 and PUFA in plant seeds; therefore, knocking out *FAD2* or silencing this gene has resulted in a higher content of C18:1 and/or a lower content of PUFA in many crops (Kang et al., 2011; Liu et al., 2002b; Mietkiewska et al., 2008; Pham et al., 2010). The mutation breeding program in our study is targeting *FAD2* in *C. abyssinica* and thus it is essential to know how many *FAD2* genes are present in the crambe genome and what the function of each of the genes is.

Typically, the *FAD2* gene is present as a single locus in diploid species whereas multiple loci are present in polyploid genomes (Okuley et al., 1994; Scheffler et al., 1997). So, the hexaploid *C. abyssinica* is expected to harbor a *FAD2* gene family with multiple members. Various strategies have been used for novel gene isolation, for instance map-based cloning (Tanksley et al., 1995), T-DNA insertion as well as transposon tagging (Walbot, 1992). In case of isolating genes which were previously reported in other species, PCR cloning based on sequence information is often applied. However, using this strategy it is sometimes hard to detect all variants due to primer specificity. The conventional way of PCR cloning was applied to isolate *FAD2* genes in crambe (*CaFAD2*) based on the Arabidopsis *FAD2* gene sequence and five genes (*CaFAD2-A, -B, -C1, -C2* and *-C3*) were found in this way. However, this result is not consistent to the Southern analysis which indicates that two more *CaFAD2* genes may be present (**Figure 2, Chapter 2**). Therefore, multiple sets of primers were developed to generate *FAD2* amplicons for deep amplicon sequencing. High-throughput 454 sequencing showed the presence of in total seven *FAD2* variants (**Supplementary Figure S7, Chapter 2**).

Gene function analysis in C. abyssinica

For genes in polyploid species, usually multiple loci are present as orthologs or paralogs because of speciation and genome duplication (Lawton-Rauh, 2003). Also, these duplicated loci are sometimes expressed in different tissues. For the seven detected *CaFAD2* genes in *C. abyssinica*, sequence analysis suggests that three out of them (*CaFAD2-C1, -C2 and -C3*) encode the full length protein and thus are functional. In contrast, the remaining four gene copies encode truncated proteins because of the presence of nucleotide deletions in the coding sequences and are termed pseudo genes (**Figure S5, Chapter 2**). Moreover, this indication was substantiated by the temporal and spatial expression studies in crambe plants, where only three *CaFAD2* genes with full coding sequences were expressed in the tissues investigated (**Supplementary Figure S9, Chapter 2**). The case of *FAD2* genes in *Brassica napus* also matches to this situation. *B. napus* is a tetraploid species and derives from two ancestors, *Brassica rapa* and *Brassica oleracea* (U, 1935) and there are four *BnFAD2* genes present in the *B. napus* genome (Scheffler et al., 1997). Similar to the *CaFAD2* genes, three of them were expressed constitutively and the proteins were localized in the endoplasmic reticulum (ER) whereas the remaining one encodes a truncated FAD2 protein (Lee et al., 2013). Surprisingly, the gene encoding truncated FAD2 was also expressed in the roots and developing seeds, which is different to the case of *C. abyssinica*, and the further study on the

promoter indicated that the truncated protein resides in the nucleus and chloroplasts of *B. napus* (Lee et al., 2013).

To study the functions of the three actively expressed *CaFAD2* genes, the endogenous *CaFAD2* genes were down-regulated by RNAi and two generations of transgenic lines were analyzed for *CaFAD2* expression and oil composition in the seeds. All three *CaFAD2* genes (-*C1,* -*C2* and -*C3*) were down-regulated in the transgenic progenies, though to varying levels among the different transgenic families (**Figure 2A, Chapter 3**). Naturally, *CaFAD2-C3* shows the highest expression level in the seed tissue followed by *CaFAD2-C2* and *CaFAD2-C1* (**Figure 3, Chapter 2**). The RNAi study showed that *CaFAD2-C3* was the major gene related to the seed oil whereas *CaFAD2-C1* and -*C2* down-regulation showed a minor effect on oil composition (Figure 5, Chapter 3). This finding is strongly supported by the analysis of a nonsense mutant of *CaFAD2-C3*, which showed a significant increase in C18:1 and reduction in PUFA content (**Figure 5A, Chapter 5**). However, the similar phenotypic effect on oil composition in the nonsense mutants for *CaFAD2-C1* and -*C3* indicates that both genes play a similar role in the seed oil biosynthesis (**Figure 5A, Chapter 5**). All these findings suggest that at least one *CaFAD2* gene (*CaFAD2-C1* or *CaFAD2-C3*) needs to be knocked out to achieve higher C18:1 and lower PUFA contents and that knocking out both genes could result in more pronounced changes in crambe oil.

Choice for a mutation detection approach

In this study, we applied two high-throughput sequencing platforms, 454 and Illumina, to detect mutations in the *FAD2* genes. In the program of classic mutation breeding, mutants were traditionally found by forward screening for novel phenotypes, however, it is not feasible to directly detect the mutants when the investigated phenotypes are not easily detectable, e.g. oil composition. Furthermore, in polyploid species the possible resultant phenotype from mutation could be masked by the wild-type allele because of the nature of genetic redundancy (Stemple, 2004). TILLING (targeting induced local lesion in genomes), a reverse genetic approach, is more often applied for gene function analysis to detect mutations in candidate genes towards mutants with desirable phenotypes. Many methods are capable to detect single nucleotide transitions on the DNA level including capillary electrophoresis (McCallum et al., 2000), Li-cor (Colbert et al., 2001) as well as MALDI-TOF (Chawade et al., 2010) and some of these methods have been conditionally applied to polyploid species (Reviewed by Till et al., 2007). Among these methods, Li-cor is the most commonly used

thanks to its high-throughput properties. In this study, we initially tested the possibility of mutation detection with the *Cel I* enzyme that cuts mismatch positions in heteroduplex molecules by Li-cor but it failed to apply this system to the hexaploid crambe (data not shown). High resolution melting (HRM) analysis also has been applied to analyze genetic variations in PCR amplicons without a requirement of post-PCR separation, thus it is a very efficient method for reverse genetics screens (Uitdewilligen, 2012). However, all these detection methods require the development of specific primers, which is particularly difficult in polyploids e.g. hexaploid *C. abyssinica* where genes are present as multiple loci and share high identity in sequences. This indicates that other high-throughput means are needed for mutation detection in the crambe.

DNA sequencing is considered by 'rule of thumb' as the most straight forward method for mutation detection. Recently, the efficiency of mutation detection has been rapidly facilitated by the high-throughput next generation sequencing (NGS). Due to these developments in technology the detection of mutations in complex situations is now possible, such as mutations in genes belonging to a multigene family in a polyploid crop (Chen et al., 2013; Gholami et al., 2012; Kaur et al., 2012). We successfully detected mutations in the candidate genes (*CaFAD2*) within the established crambe mutant population using amplicon sequencing with two different next generation sequencing platforms (454 and Illumina). In three 454 sequencing runs, ten mutants were found in *CaFAD2* amplicons (size ranges from 110 bp to 396 bp) within 186 pools (1,860 lines). In these runs one mutation was found per 382 kbp (mutation density) (Chapter 4). A number of 17 mutations was detected by Illumina sequencing (one lane, 25 Gb) in a *CaFAD* fragment of 272 bp amplified from 1,100 pools (11,000 lines), indicating a mutation density of around one per 140 kbp for the whole population (**Chapter 5**).

Although 454 sequencing enables to generate reads with a larger size than Illumina sequencing (to date 400 bp vs. 100 bp in one direction) (Bräutigam and Gowik, 2010), the much higher throughput (6 Gbp vs. 0.4 Gbp reads for yield per run) (Bräutigam and Gowik, 2010), and relatively lower costs make Illumina sequencing more practical and powerful in the application of mutation discovery from a vast number of lines.

Improving oil quality for *C. abyssinica* by knocking out the *CaFAD2* genes

To improve crambe oil quality by mutagenizing *CaFAD2* genes, mutation detection was performed targeted to the expressed *CaFAD2* genes (-*C1*, -*C2* and -*C3*) by 454 and Illumina

sequencing resulting in respectively 9 and 17 mutations (n=26). Half of them (n=13) are either missense (n=9) resulting in an amino acid change, or nonsense mutations that are introducing a stop codon (n=4, two in *CaFAD2-C3*, the other two in *CaFAD2-C1* and -*C2*) whereas the others (n=13) are silent mutations that don't have any impact on the amino acid composition (**Table 2 in Chapter 4 and Table 1 and 2 in Chapter 5**).

The nonsense mutations result in truncated FAD2 proteins and thus it is promising to find homozygous nonsense mutants with altered oil composition compared to the wild-type. Nonsense mutants corresponding to *CaFAD2-C1* and *CaFAD2-C3* were identified and analyzed for oil composition. As expected, both mutants showed a higher C18:1 content as was determined in the seed oil composition of segregating M_3 progenies. For the nonsense mutant of *CaFAD2-C1* only a single M_3 plant was able to germinate which appeared to be heterozygous for the mutation. The C18:1 level in the seed oil of this plant increased with approximately 17 % as compared to the general C18:1 level in wildtype crambe plants. In the M_3 progeny of the nonsense mutant of *CaFAD2-C3* the C18:1 fraction in the oil increased from 0.138 in the wildtype, to 0.155 (+12 % change) in the heterozygous progeny and to 0.173 (+25 % change) in the homozygous progeny. As expected based on the proposed actions for *CaFAD2-C1* and -*C3*, the PUFA concentrations reduced proportionally (-11 % change in the *CaFAD2-C3* heterozygote and −25 % change in the *CaFAD2-C3* homozygote) (**Figure 5A**). This indicates that it is feasible to improve the quality of crambe oil by knocking out either *CaFAD2-C1* or *CaFAD2-C3*. Although a nonsense mutation for *CaFAD2-C2* was also detected, the corresponding mutant line was not identified. Therefore, it is unknown whether knocking out *CaFAD2-C2* would result in an oil phenotype similar to the nonsense mutants for *CaFAD2-C1* and -*C3*. It was also reported that FAD2 in other crops plays a role under abiotic stress, e.g. salt tolerance and cold resistance (Shi et al., 2012; Zhang et al., 2012). Further studies are necessary to unravel additional functions for the crambe FAD2 genes.

Besides the nonsense mutations that generate functional knockouts, also several missense mutations (n=9) were found that are resulting in an amino acid change in the FAD2 protein. Two of the missense mutants were identified in the M_3 generation. A mutation in *CaFAD2-C1* that introduces an alanine to valine in position 104 and a mutation in *CaFAD2-C2* that introduces a glycine to asparagine change in position.108 of the wild-type CaFAD2 protein.

In both cases the homozygous missense mutants showed relatively minor alterations in oil composition (up to 4 % increase in C18:1 and 8 % decrease in PUFA) (**Figure 5B, Chapter 5**). Nevertheless, it is possible to gain more significant changes in crambe oil by stacking this missense mutation with other missense/nonsense mutations. This turned out to be true in the case of soybean where, although a homozygous *FAD2-1B* missense mutant allele alone was not capable of producing a high oleic acid phenotype, a high oleic acid phenotype was recovered from the homozygous double missense (*FAD2-1A* and *FAD2-1B*) mutants (Pham et al., 2010). The two single mutants (*FAD2-1A* and *FAD2-1B*) both enabled to achieve a moderate increased in C18:1 content (by 17 % and 8 % respectively). The mutation in FAD2-1A caused an amino acid change from Ser to Asn at the position 117 of the protein and this position is not located but closed to the conserved histidine domain. What is more, combining these mutations eventually led to more pronounced changes in the oil composition (by 60 % C18:1 increase) (Pham et al., 2010). Since in crambe *CaFAD2-C1* and *CaFAD2-C3* are both related to seed oil composition, it is expected to gain an additional improvement in crambe oil quality by crossing the nonsense/missense mutants.

Meanwhile, crambe mutants in the high yielding cultivar 'Galactica' can be used to develop crambe lines with novel fatty acids types in their oil, or with oils containing wax esters. A wax ester is an ester of a fatty acid and a fatty alcohol by wax ester synthase (WS) and this compound is found as a storage of energy in some plants, notably jojoba and is of importance for cosmetic and pharmaceutical applications (Li et al., 2008; Miwa, 1971). Taken together with the fact that crambe seeds originally contain a high content of C22:1, the crambe lines with higher C18:1 obtained here can be used as parents for WS gene transformation to produce medium-long-chain (C18) and very-long-chain (C22) wax esters (W. Qi 2014; Wageningen UR Plant Breeding; unpublished results).

Strategies towards the increase of C22:1 content in crambe

The improvement of crambe oil aims directly to elevate erucic acid (C22:1) content. It is theoretically possible to elevate C22:1 content by increasing the C18:1 level in the oil since oleoyl (C18)-CoA derived from C18:1 is the precursor of C22:1 synthesis. Because C18:1 in seeds can also be converted into C18:2 by FAD2, blocking the conversion from C18:1 to C18:2 is expected to increase not only the concentration of C18:1 but also the pool of oleoyl (C18)-CoA. Therefore, it can be hypothesized that disrupting the function of FAD2 would promote the synthesis of C22:1. There are at least two ways to disrupt a gene function:

1) gene silencing/suppression of the gene expression by genetic modification (GM); 2) mutation breeding to generate a truncated protein (Non-GM).

In this thesis, we identified a homozygous nonsense mutants for *CaFAD2-C3*, in this single mutant the content of C18:1 increased moderately (by at maximum 25%) but the increasing C18:1 pool did not lead to a concomitant elevation of C22:1 (**Figure 5A, Chapter 5**). Interestingly, the missense mutant of *CaFAD2-C1* that showed only a minor effect on C18:1 and PUFA (up to ~8 % changes), unexpectedly showed a minor but significant increase in the C22:1 content (by 2 %, **Figure 5B in Chapter 5**). Therefore, it seems that the little additional accumulation of C18:1 in the missense mutant was eventually partially converted into C22:1 whereas all the increased C18:1 accumulation in the nonsense mutants of *CaFAD2-C3*, which is the major *CaFAD2* gene expressed in seeds was only stored as C18:1 in the seeds and did not result in any change in C22:1 level. Analyzing the data on the metabolic flux to C22:1 in these mutants would be helpful to answer the question why only the missense mutants showed an elevated content in C22:1. Maybe *CaFAD2-C1* and *CaFAD2-C2*, which are the genes that contain the missense mutations, act differently in the oil biosynthesis pathway compared to *CaFAD2-C3*.

The function of CaFAD2 was also explored through down-regulation of the expression of this gene by using of RNA interference (RNAi). Many studies have been done to investigate the effect on C18:1 and C22:1 contents by down-regulating the expression of endogenous *FAD2* genes. The hypothesis that silencing FAD2 leads to an increase of C18:1 and C22:1 was proven to be true in case of *Brassica carinata*, in which the contents of both, C18:1 and C22:1, were increased by down-regulation of *FAD2* expression (Jadhav et al., 2005); however, it failed so far to gain similar results in *Arabidopsis* (Stoutjesdijk et al., 2002), soybean (Wagner et al., 2011) and other *Brassica* species (Stoutjesdijk et al., 2000). For *C. abyssinica*, Li et al. (2012) investigated three generations of FAD2-RNAi transformants and found that significant increase in C18:1 did not lead to increased C22:1 content. We studied different progenies of the same transformants as Li et al. (2012) and we found that in the T_2 seeds both, C18:1 and C22:1 were significantly increased compared to the control. Unfortunately, this increase in C22:1 was not present in the T_3 seeds. These results suggest that it is difficult to elevate C22:1 in crambe through the silencing of *FAD2* alone and reveals that there are most probably other limitations than the available pool of oleoyl-CoA in crambe.

Fatty acid elongase 1(FAE1) was shown to be associated with the variation of C22:1 content and control biosynthesis of very-long-chain fatty acid (VLCFA) (Barret et al., 1998; Fourmann et al., 1998; Millar and Kunst, 1997). Although some studies showed that down regulation of *FAE1* expression was able to significantly reduce the C22:1 content in plant oil, over-expression of endogenous *FAE1* alone led to only a minor increase in C22:1 content for *B. napus* and *B. juncea* (Kanrar et al., 2006; Nath et al., 2009). The enzyme FAE1 in crambe has substrate specificity for C20-fatty acids (Mietkiewska et al., 2007). Crambe cv. Galactica contains over 60 % of C22:1 and this content nearly reached a theoretical peak (66%) in wildtype crambe. This might suggest that over expression of the *FAE1* gene in crambe is not a promising strategy for a further elevation of the C22:1 content.

To increase C22:1 in crambe oil with genetic engineering, an attempt is to combine silencing *FAD2*, over-expressing *FAE1* and introducing LPAAT (lysophospholipid acyltransferase), which transfers acyl-fatty acid to the sn-2 position of triacylglycerol, specific to C22-fatty acids (Lassner et al., 1995). By combining *FAD2* silencing and overexpression of *FAE1*of *C. abyssinica*, 10 % increase in C22:1 was found in transgenic *B. carinata* seeds (Mietkiewska et al., 2008). In transgenic rapeseed, C22:1 content was increased by 60 % with co-expression of *BnFAE1* and an erucoyl-CoA-specific LPAAT (Han et al., 2001). Recently, genetic engineering has been applied in *C. abyssinica* to increase C22:1. Through over-expression of *BnFAE1* and erucoyl-CoA-specific LPAAT, C22:1 content in crambe oil increased to 70 % when endogenous *FAD2* was silenced in combination with over-expression of *FAE1* and C22-specific LPAAT, and this content further increased up to 76.9 % in the successive generations (Li et al., 2012).

In our study, the crambe lines with *FAD2* silencing (GM) and *CaFAD2* knockouts (non-GM) produced more pure crambe oil (containing less PUFA). The knockout mutants of cultivar 'Galactica' can be used directly to generate new (non-GM) cultivars by stacking mutations in different *FAD2* genes. In addition, the mutant population developed in this study can also be used to isolate mutations in the candidate genes for other important traits, e.g. in order to reduce the glucosinolate content in seed meal. Crambe defatted meal contains 4-6 % (w/w) of glucosinolates, with epiprogoitrin accounting for > 90 % of the total and this feature limits the use of the meal as feed (Finiguerra et al., 2001). Stacking mutations in the different candidate genes could produce crambe lines with dual quality for applications in both industry and animal husbandry.

References

Ahloowalia, B. and Maluszynski, M. (2001) Induced mutations-A new paradigm in plant breeding. Euphytica 118, 167-173.

Ahloowalia, B., Maluszynski, M. and Nichterlein, K. (2004) Global impact of mutation-derived varieties. Euphytica 135, 187-204.

Backes, G. (2013) TILLING and EcoTILLING. In: Diagnostics in Plant Breeding (Varshney, R.K. ed) pp. 145-165. Springer.

Barret, P., Delourme, R., Renard, M., Domergue, F., Lessire, R., Delseny, M. and Roscoe, T. (1998) A rapeseed FAE1 gene is linked to the E1 locus associated with variation in the content of erucic acid. TAG Theoretical and Applied Genetics 96, 177-186.

Bräutigam, A. and Gowik, U. (2010) What can next generation sequencing do for you? Next generation sequencing as a valuable tool in plant research. Plant Biology 12, 831-841.

Chawade, A., Sikora, P., Bräutigam, M., Larsson, M., Vivekanand, V., Nakash, M.A., Chen, T. and Olsson, O. (2010) Development and characterization of an oat TILLING-population and identification of mutations in lignin and β-glucan biosynthesis genes. BMC Plant Biology 10.

Chen, X., Li, X., Zhang, B., Xu, J., Wu, Z., Wang, B., Li, H., Younas, M., Huang, L., Luo, Y., Wu, J., Hu, S. and Liu, K. (2013) Detection and genotyping of restriction fragment associated polymorphisms in polyploid crops with a pseudo-reference sequence: A case study in allotetraploid Brassica napus. BMC Genomics 14.

Cheng, J., Salentijn, E.M.J., Huang, B., Krens, F.A., Dechesne, A.C., Visser, R.G.F. and van Loo, E.N. (2013) Isolation and characterization of the omega-6 fatty acid desaturase (FAD2) gene family in the allohexaploid oil seed crop Crambe abyssinica Hochst. Molecular Breeding 32, 517-531.

Colbert, T., Till, B.J., Tompa, R., Reynolds, S., Steine, M.N., Yeung, A.T., McCallum, C.M., Comai, L. and Henikoff, S. (2001) High-throughput screening for induced point mutations. Plant Physiology 126, 480-484.

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

Finiguerra, M.G., Iori, R. and Palmieri, S. (2001) Soluble and total myrosinase activity in defatted Crambe abyssinica meal. Journal of Agricultural and Food Chemistry 49, 840-845.

Fourmann, M., Barret, P., Renard, M., Pelletier, G., Delourme, R. and Brunel, D. (1998) The two genes homologous to Arabidopsis FAE1 co-segregate with the two loci governing erucic acid content in Brassica napus. Theoretical and Applied Genetics 96, 852-858.

Gholami, M., Bekele, W.A., Schondelmaier, J. and Snowdon, R.J. (2012) A tailed PCR procedure for cost-effective, two-order multiplex sequencing of candidate genes in polyploid plants. Plant Biotechnology Journal 10, 635-645.

Griga, M., Stejskal, J. and Beber, K. (1995) Analysis of tissue culture-derived variation in pea (Pisum sativum L.)-preliminary results. Euphytica 85, 335-339.

Han, J., Lühs, W., Sonntag, K., Zähringer, U., Borchardt, D.S., Wolter, F.P., Heinz, E. and Frentzen, M. (2001) Functional characterization of β-ketoacyl-CoA synthase genes from Brassica napus L. Plant Molecular Biology 46, 229-239.

Hu, X., Sullivan-Gilbert, M., Gupta, M. and Thompson, S.A. (2006) Mapping of the loci controlling oleic and linolenic acid contents and development of fad2 and fad3 allele-specific markers in canola (Brassica napus L.). TAG Theoretical and Applied Genetics 113, 497-507.

Jadhav, A., Katavic, V., Marillia, E.F., Michael Giblin, E., Barton, D.L., Kumar, A., Sonntag, C., Babic, V., Keller, W.A. and Taylor, D.C. (2005) Increased levels of erucic acid in Brassica carinata by co-suppression and antisense repression of the endogenous FAD2 gene. Metabolic Engineering 7, 215-220.

Kang, J., Snapp, A.R. and Lu, C. (2011) Identification of three genes encoding microsomal oleate desaturases (FAD2) from the oilseed crop Camelina sativa. Plant Physiology and Biochemistry 49, 223-229.

Kanrar, S., Venkateswari, J., Dureja, P., Kirti, P.B. and Chopra, V.L. (2006) Modification of erucic acid content in Indian mustard (Brassica juncea) by up-regulation and down-regulation of the Brassica juncea Fatty Acid Elongation1 (BjFAE1) gene. Plant cell reports 25, 148-155.

Kaur, S., Francki, M.G. and Forster, J.W. (2012) Identification, characterization and interpretation of single-nucleotide sequence variation in allopolyploid crop species. Plant Biotechnology Journal 10, 125-138.

Lassner, M.W., Levering, C.K., Davies, H.M. and Knutzon, D.S. (1995) Lysophosphatidic acid acyltransferase from meadowfoam mediates insertion of erucic acid at the sn-2 position of triacylglycerol in transgenic rapeseed oil. Plant Physiology 109, 1389-1394.

Lawton-Rauh, A. (2003) Evolutionary dynamics of duplicated genes in plants. Molecular Phylogenetics and Evolution 29, 396-409.

Lee, K.R., In Sohn, S., Jung, J.H., Kim, S.H., Roh, K.H., Kim, J.B., Suh, M.C. and Kim, H.U. (2013) Functional analysis and tissue-differential expression of four FAD2 genes in amphidiploid Brassica napus derived from Brassica rapa and Brassica oleracea. Gene 531, 253-262.

Lessman, K.J. (1975) Variation in crambe, Crambe abyssinica hochst. Journal of the American Oil Chemists Society 52, 386-389.

Li, F., Wu, X., Lam, P., Bird, D., Zheng, H., Samuels, L., Jetter, R. and Kunst, L. (2008) Identification of the wax ester synthase/acyl-coenzyme a:diacylglycerol acyltransferase WSD1 required for stem wax ester biosynthesis in Arabidopsis. Plant Physiology 148, 97-107.

Li, X., van Loo, E.N., Gruber, J., Fan, J., Guan, R., Frentzen, M., Stymne, S. and Zhu, L.H. (2012) Development of ultra-high erucic acid oil in the industrial oil crop Crambe abyssinica. Plant Biotechnology Journal 10, 862-870.

Liu, Q., Singh, S.P. and Green, A.G. (2002a) High-stearic and high-oleic cottonseed oils produced by hairpin RNA-mediated post-transcriptional gene silencing. Plant Physiology 129, 1732.
Liu, Q., Singh, S.P. and Green, A.G. (2002b) High-stearic and high-oleic cottonseed oils produced by hairpin RNA-mediated post-transcriptional gene silencing. Plant Physiology 129, 1732-1743
Mastebroek, H.D., Wallenburg, S.C. and van Soest, L.J.M. (1994) Variation for agronomic characteristics in crambe (Crambe abyssinica Hochst. ex Fries). Industrial Crops and Products 2, 129-136.
McCallum, C.M., Comai, L., Greene, E.A. and Henikoff, S. (2000) Targeted screening for induced mutations. Nature Biotechnology 18, 455-457.
Mietkiewska, E., Brost, J.M., Giblin, E.M., Barton, D.L. and Taylor, D.C. (2007) Cloning and functional characterization of the fatty acid elongase 1 (FAE1) gene from high erucic Crambe abyssinica cv. Prophet. Plant biotechnology journal 5, 636-645.
Mietkiewska, E., Hoffman, T.L., Brost, J.M., Giblin, E.M., Barton, D.L., Francis, T., Zhang, Y. and Taylor, D.C. (2008) Hairpin-RNA mediated silencing of endogenous FAD2 gene combined with heterologous expression of Crambe abyssinica FAE gene causes an increase in the level of erucic acid in transgenic Brassica carinata seeds. Molecular Breeding 22, 619-627.
Millar, A.A. and Kunst, L. (1997) Very-long-chain fatty acid biosynthesis is controlled through the expression and specificity of the condensing enzyme. Plant Journal 12, 121-131.
Miwa, T.K. (1971) Jojoba oil wax esters and derived fatty acids and alcohols: gas chromatographic analyses. Journal of the American Oil Chemists' Society 48, 259-264.
Mohan Jain, S. (2001) Tissue culture-derived variation in crop improvement. Euphytica 118, 153-166.
Nath, U.K., Wilmer, J.A., Wallington, E.J., Becker, H.C. and Möllers, C. (2009) Increasing erucic acid content through combination of endogenous low polyunsaturated fatty acids alleles with Ld-LPAAT + Bn-fae1 transgenes in rapeseed (Brassica napus L.). Theoretical and Applied Genetics 118, 765-773.
Okabe, Y., Ariizumi, T. and Ezura, H. (2013) Updating the Micro-Tom TILLING platform. Breeding science 63, 42.
Okuley, J., Lightner, J., Feldmann, K., Yadav, N. and Lark, E. (1994) Arabidopsis FAD2 gene encodes the enzyme that is essential for polyunsaturated lipid synthesis. Plant Cell 6, 147-158.
Parry, M.A., Madgwick, P.J., Bayon, C., Tearall, K., Hernandez-Lopez, A., Baudo, M., Rakszegi, M., Hamada, W., Al-Yassin, A. and Ouabbou, H. (2009) Mutation discovery for crop improvement. Journal of Experimental Botany 60, 2817-2825.
Pham, A.T., Lee, J.D., Shannon, J.G. and Bilyeu, K.D. (2010) Mutant alleles of FAD2-1A and FAD2-1B combine to produce soybeans with the high oleic acid seed oil trait. BMC Plant Biology 10, 195.
Pirtle, I.L., Kongcharoensuntorn, W., Nampaisansuk, M., Knesek, J.E., Chapman, K.D. and Pirtle, R.M. (2001) Molecular cloning and functional expression of the gene for a cotton Δ-12 fatty acid desaturase (FAD2). Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression 1522, 122-129.
Rommens, C.M., Haring, M.A., Swords, K., Davies, H.V. and Belknap, W.R. (2007) The intragenic approach as a new extension to traditional plant breeding. Trends in Plant Science 12, 397-403.
Rudloff, E. and Wang, Y. (2011) Crambe. Wild Crop Relatives: Genomic and Breeding Resources, 97-116.
Scheffler, J.A., Sharpe, A.G., Schmidt, H., Sperling, P., Parkin, I.A.P., Lühs, W., Lydiate, D.J. and Heinz, E. (1997) Desaturase multigene families of Brassica napus arose through genome duplication. TAG Theoretical and Applied Genetics 94, 583-591.
Shi, J., Cao, Y., Fan, X., Li, M., Wang, Y. and Ming, F. (2012) A rice microsomal delta-12 fatty acid desaturase can enhance resistance to cold stress in yeast and Oryza sativa. Molecular Breeding 29, 743-757.
Stemple, D.L. (2004) TILLING-a high-throughput harvest for functional genomics. Nature Reviews Genetics 5, 145-150.
Stoutjesdijk, P.A., Hurlestone, C., Singh, S.P. and Green, A.G. (2000) High-oleic acid Australian Brassica napus and B. juncea varieties produced by co-suppression of endogenous Δ12-desaturases. Biochemical Society Transactions 28, 938-940.
Stoutjesdijk, P.A., Singh, S.P., Liu, Q., Hurlstone, C.J., Waterhouse, P.A. and Green, A.G. (2002) hpRNA-mediated targeting of the Arabidopsis FAD2 gene gives highly efficient and stable silencing. Plant Physiology 129, 1723.
Tanksley, S.D., Ganal, M.W. and Martin, G.B. (1995) Chromosome landing: A paradigm for map-based gene cloning in plants with large genomes. Trends in Genetics 11, 63-68.
Till, B.J., Comai, L. and Henikoff, S. (2007) TILLING and EcoTILLING for crop improvement. In: Genomics-assisted crop improvement pp. 333-349. Springer.
U, N. (1935) Genome analysis in Brassica with special reference to the experimental formation of B. napus and peculiar mode of fertilization. Japanese Journa of Botany 7, 389-452.
Uitdewilligen, J. (2012) Discovery and genotyping of existing and induced DNA sequence variation in potato. In: Plant Breeding Department. Wageningen: Wageningen University.
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 biotechnology journal 11, 197-210.
Vogtmann, H., Clandinin, D.R. and Hardin, R.T. (1973) Utilization of rapeseed oils of high and low erucic acid contents. I. Digestibility and energy utilization. Nutrition and Metabolism 15, 252-266.
Wagner, N., Mroczka, A., Roberts, P.D., Schreckengost, W. and Voelker, T. (2011) RNAi trigger fragment truncation attenuates soybean FAD2-1 transcript suppression and yields intermediate oil phenotypes. Plant biotechnology journal 9, 723-728.
Walbot, V. (1992) Strategies for mutagenesis and gene cloning using transposon tagging and T-DNA insertional mutagenesis. Annual Review of Plant Physiology and Plant Molecular Biology 43, 49-82.
Wang, Y. and Luo, P. (1998) Intergeneric hybridization between Brassica species and Crambe abyssinica. Euphytica 101, 1-7.

Wang, Y., Sonntag, K., Rudloff, E., Wehling, P. and Snowdon, R.J. (2006) GISH analysis of disomic Brassica napus-Crambe abyssinica chromosome addition lines produced by microspore culture from monosomic addition lines. Plant Cell Reports 25, 35-40.
Wang, Y.P., Snowdon, R.J., Rudloff, E., Wehling, P., Friedt, W. and Sonntag, K. (2004) Cytogenetic characterization and fae1 gene variation in progenies from asymmetric somatic hybrids between Brassica napus and Crambe abyssinica. Genome 47, 724-731.
Wang, Y.P., Sonntag, K. and Rudloff, E. (2003) Development of rapeseed with high erucic acid content by asymmetric somatic hybridization between Brassica napus and Crambe abyssinica. Theoretical and Applied Genetics 106, 1147-1155.
Zhang, J., Liu, H., Sun, J., Li, B., Zhu, Q., Chen, S. and Zhang, H. (2012) Arabidopsis Fatty Acid Desaturase FAD2 Is Required for Salt Tolerance during Seed Germination and Early Seedling Growth. PloS One 7, e30355.

Summary

Petroleum (fossil oil) and its derivatives are providing feedstocks for many industrial products. The expected shortage of petroleum and the undesired greenhouse gas emissions from fossil oil call for seeking viable alternatives for this resource. Therefore, several oil crops are considered as alternative to fossil oil for the sustainable production of oil types that resemble petroleum derivatives. For instance, high erucic acid rapeseed (HEAR) is an important source of high erucic acid (C22:1) oil, which is an industrial oil-type with many applications, e.g. in manufacturing plastics.

Crambe abyssinica L (hexaploid; 2n = 6x = 90 chromosomes), an under-utilized non-food oil crop, contains an even higher content of C22:1 in its seed-oil than HEAR. Taken together with the fact that *C. abyssinica* cannot inter-cross with other food oil crops, which diminishes the risk of mixing its industrial oil with oil of the food oil chain, it is thus very promising to use *C. abyssinica* as a renewable source for industrial oil types.

The commercial implementation of *C. abyssinica* needs a further increase of the C22:1 level and/or a reduction of the polyunsaturated fatty acid (PUFA) level in the seed oil. Crambe oil is usually separated into a C22:1 fatty acid fraction and 'top fatty acid' fraction consisting of C18 fatty acids. A low PUFA and very high C18:1 content of the top fatty acids is desired by industry. Further, a high PUFA content makes the oil prone to oxidation and therefore shortens the shelf-life of this particular oil. This thesis aims to develop a molecular breeding program to establish crambe lines with high contents of C22:1 and lower contents of PUFA in its seed oil. The project is carried out in the framework of the EU project ICON (Industrial Crops producing added value Oils for Novel chemicals). In **Chapter 1,** the crop crambe is introduced and the key genes involved in the oil biosynthesis pathway are described. The enzyme fatty acid desaturase 2 (FAD2) is known to be involved in the biosynthesis of PUFA, and in several crops this gene has demonstrated its value as target gene for the development of improved oil qualities

Chapter 2 describes the isolation and characterization of the FAD2 gene family (*CaFAD2*) in crambe cultivar 'Galactica'. The genetic variation of this gene family in the hexaploid *C. abyssinica* was investigated and the expression levels of the gene family members were determined in several breeding lines of crambe with different seed oil compositions.The results show that the *CaFAD2* gene family is composed of seven copies of which only three (*CaFAD2-C1*, *CaFAD2-C2* and *CaFAD2-C3*) are expressed in the tissues investigated.

CaFAD2-C3 is predominantly expressed in the seeds, indicating that this gene could be important in crambe seed oil synthesis. The expression profiles of the *CaFAD2* genes were investigated in a set of breeding lines, which harbour variation in oil composition. It was shown that only minor expression variation was present among the lines for the *CaFAD2* genes and this variation could not explain the variation in oil composition in the selected lines investigated. The limited natural variation among breeding lines of crambe supports the choice of methods for inducing and targeting genetic variation as described in this thesis.

To investigate the function of each of the three expressed *CaFAD2* genes, transgenic crambe lines were developed to silence endogenous *CaFAD2* with RNAi (**Chapter 3**). Two primary (T0) transformants were obtained with variable insertion numbers and they were propagated to two successive generations (T1 and T2). The expression levels of the *CaFAD2* genes and oil composition in the seeds of these two successive generations were studied. It was found that the expression level of *CaFAD2-C3* was highly positively correlated with the level of PUFA in the seed oil. This result suggests that *CaFAD2-C3* plays a critical role in seed oil composition in crambe. Silencing of gene expression and accompanying alterations in oil composition were observed in both generations, indicating that the silencing of *CaFAD2* expression was stably inherited. It was demonstrated that it is feasible to modify oil composition in crambe seeds by targeting *CaFAD2*. All three expressed *CaFAD2* variants were silenced to a large extent, and this occurred with a concomitant reduction in PUFA. This reduction in PUFA was most strongly correlated with the extent of silencing of *CaFAD2-C3*, followed by that of *CaFAD2-C1* and *CaFAD2-C2*.

On the basis of this RNAi-proof of concept of gene function, a molecular mutation breeding program was set up by producing a crambe mutant population of cultivar 'Galactica', induced by EMS (ethyl methane sulfonate) and by developing a mutation detection method for the *CaFAD2* genes using 454-amplicon sequencing (**Chapter 4**). The aim was to detect mutations from the pooled DNAs of mutant genotypes (pools of ten M2 plants) and to discover mutants with changed oil composition in the seed oil. In total, over 12,000 individual mutants (M2) were obtained. The DNA of M2 mutants was organized into 1,200 mutant pools to reduce the screening efforts. Using 454 amplicon sequencing mutations in the *CaFAD2* genes were discovered. On analysing the 454- sequences of 168 pools from 1680 genotypes, ten mutations were detected among the three expressed *CaFAD2* genes, including four 'missense' mutations; missense mutations are point mutations resulting in an amino acid change.

In the 454 sequencing only pools of ten genotypes were found containing a genotype with a mutation. For each pool of ten genotypes, a seed pool was available (M3 or M4 generation) from which the missense mutants were selected. The oil composition of these missense mutants was investigated to evaluate the effect of these mutations. Because of the location of missense mutations in non-conserved domains for FAD2 function, these mutants showed, as expected, an oil composition in the seeds typical of wild type genotypes.

To find mutants with desirable phenotypes in oil, an Illumina sequencing platform was used to screen the whole population for mutations in *CaFAD2*. This was done in combination with two different sequencing analysis strategies: 1) motif search and 2) screening all differences (**Chapter 5**). Here, 17 novel mutations in a conserved region of CaFAD2 were identified (four nonsense, four missense and nine silent) within 1,100 mutant pools. For each of the three different *CaFAD2* genes, a nonsense mutation was found; one for both*CaFAD2-C1* and *CaFAD2-C2* and two for *CaFAD2-C3*. Next, by sequencing the *CaFAD2* amplicons from the ten lines composing a pool, the mutant lines (M3-lines) that carry the specific mutations were selected. The heterozygous nonsense mutant (M3 generation) and homozygous nonsense mutant (M3 generation) for CaFAD2-C1 and CaFAD2-C3 respectively were identified but the nonsense mutation in *CaFAD2-C2* was not found back after screening the individual plants in the positive pool. The mutant phenotypes were determined by analysing the fatty acid composition of the seed oil of offspring of the mutant genotypes usingGC. The homozygote nonsense mutation of *CaFAD2-C3* showed a significant increase in C18:1 (+25 %, relative to the level of the wild-type) and an accompanying reduction in PUFA contents (-25 % compared to the wild type). The heterozygote of the same mutation in *CaFAD2-C3* showed an intermediate change with an increase in C18:1 of 12 %) and decrease in PUFA of 11 %, both compared to the wild-type. These alterations were also found in the heterozygote nonsense mutation of *CaFAD2-C1*.

Also, the missense mutant genotypes (M3 generation) for *CaFAD2-C1* and *CaFAD2-C2* were identified and the seed oil composition of heterozygous, homozygous and wild-type individuals of these lines were compared. The missense mutants of *CaFAD2-C1* (alanine to valine) and *CaFAD2-C2* (glycine to asparagine) both showed minor but significant alterations in the content of C22:1 (in addition to minor alterations in PUFA and C18:1, only in the missense mutant of *CaFAD2-C1*). It is demonstrated that it is feasible to improve the oil of crambe, elevating C18:1 and reducing PUFA contents, by knocking out a single *CaFAD2* gene. It is also demonstrated that the use of Illumina sequencing technology enables to

identify mutations in candidate genes from a vast number of mutants for the hexaploid *C. abyssinica*.

In the General discussion (**Chapter 6**), the results of the mutant characterisations are discussed with regard to the function of *CaFAD2* genes and the metabolic fluxes in the mutants. We propose that it is possible to further improve the crambe oil quality by combining mutations in the different *CaFAD2* genes. Furthermore, related to the results of FAD2 mutant studies in other species, the feasibility and strategies for increasing C22:1 content by mutation breeding and genetic engineering are discussed. In addition, the mutation detection platform for *C. abyssinica* using next generation sequencing was compared with other mutation detection platforms for different crops.

.

Samenvatting

Fossiele olie (aardolie) wordt gebruikt als grondstof voor veel industriële producten. De winbare voorraden fossiele olie slinken snel en het gebruik van fossiele olie leidt uiteindelijk ook in de levenscyclus van industriële producten tot ongewenste broeikasgasemissie. Daarom is de zoektocht naar bruikbare alternatieven waaronder plantaardige oliën in volle gang. *Crambe abyssinica*, (crambe, Afrikaanse bolletjeskool, een hexaploid (2n=6x =90 chromosomen) is een veelbelovend duurzaam olie gewas dat zaadolie levert met een zeer hoog gehalte (60%) erucazuur (C22:1). Dit type olie heeft veel industriële toepassingen en wordt onder andere gebruikt bij de productie van plastics en kan fossiele olie vervangen als grondstof voor industriële producten. Dit proefschrift is gericht op het ontwikkelen van een moleculair veredelingsprogramma voor crambe, met name voor de selectie van crambe lijnen met verbeterde oliesamenstelling. Het project is (deels) uitgevoerd in het kader van het EU FP7 project ICON (Industrial Crops producing added value Oils for Novel chemicals)

In **Hoofdstuk 1** wordt het gewas crambe geïntroduceerd en worden genen die een rol spelen bij de biosynthese van zaadolie besproken. De kwaliteit van crambe olie kan verbeterd worden door een verdere verhoging van het C22:1 gehalte en verlaging van het gehalte meervoudig onverzadigde vetzuren (Polyunsaturated fatty acids, PUFA's). Crambe olie wordt meestal gescheiden in een C22:1-vetzuurfractie en een C18-vetzuurfractie. De industrie vraagt om een C18-vetzuurfractie met veel C18:1 en weinig PUFA. Verder zijn deze PUFA's gevoelig voor oxidatie, hetgeen een negatieve invloed heeft op de houdbaarheid van de olie. Het enzym fatty acid desaturase 2 (FAD2) is betrokken bij de biosynthese van PUFAs en is doelgen voor moleculaire mutatieverdeling.

Hoofdstuk 2 beschrijft de identificatie van de crambe genen die coderen voor FAD2. De genetische variatie en de expressie profielen van deze *CaFAD2* genen zijn bestudeerd. De resultaten tonen aan dat de *CaFAD2* gen familie uit zeven genen bestaat waarvan er slechts drie actief zijn (*CaFAD2-C1*, *CaFAD2-C2* en *CaFAD2-C3*). De expressie van *CaFAD2-C3* domineert in zaden wat aangeeft dat dit gen hoogst waarschijnlijk een belangrijke functie heeft bij PUFA synthese in de zaadolie. De expressie van de drie actieve *CaFAD2* genen is onderzocht in een selectie van crambe lijnen die verschillen in olie samenstelling. Er zijn in dit onderzoek geen verschillen in *CaFAD2* expressie gevonden die correleren met de kleine verschillen in olie samenstelling in de geselecteerde lijnen. Uit dit onderzoek blijkt verder dat

de natuurlijke variatie in het uitgangsmateriaal van crambe beperkt is. Dit onderbouwt de keuze voor het gebruik van methoden voor het induceren van genetische variatie in crambe en het gericht opsporen van specifieke mutaties, zoals beschreven in dit proefschrift.

Met behulp van 'RNAi gene silencing' is de functie van de drie actieve *CaFAD2* genen nader onderzocht (**Hoofdstuk 3**). De expressie (mRNA) niveaus van individuele *CaFAD2* genen en de olie samenstelling in de zaden is bestudeerd in twee primaire transformanten (T0 planten) en twee opeenvolgende generaties (T1 en T2) van deze RNAi-lijnen. 'Silencing' van de verschillende *CaFAD2* genen in deze transgene lijnen is gecorreleerd met de olie samenstelling en dit 'silencing' effect is aanwezig in twee opeenvolgende generaties. Uit deze analyse blijkt een sterke correlatie tussen de expressie van *CaFAD2-C3* en de samenstelling van de zaadolie. Hetzelfde geldt in mindere mate voor *CaFAD2-C1* en *CaFAD2-C2*.

Hoofdstuk 4 beschrijft de ontwikkeling van een mutanten populatie (12,000 planten) van crambe cultivar Galactica. Deze populatie bevat 'ad random' puntmutaties geïnduceerd door behandeling met de chemische stof ethylmethaansulfonaat (EMS). Het opsporen van planten met specifieke mutaties vereist een zeer efficiënte en gevoelige detectiemethode, in het bijzonder in een hexaploid gewas zoals crambe. Met behulp van '454-amplicon sequencing' zijn tien unieke mutaties geïdentificeerd (in 168, 10xDNA pools van de M2-generatie) waarvan er vier ook daadwerkelijk een aminozuur verandering in FAD2 veroorzaken (zogenaamde 'missense' mutaties).

In **Hoofdstuk 5** wordt beschreven hoe de hele mutanten populatie gescreend werd voor mutaties met behulp van Illumina amplicon sequencing (HiSeq2000) in combinatie met twee manieren van sequentie analyse: 1) motief analyse en 2) analyse van alle verschillen. Er zijn met deze efficiënte strategie zeventien nieuwe mutaties gevonden, waaronder 'nonsense' mutaties in de drie *CaFAD2* target genen en 'missense' mutaties in geconserveerde sequenties. De 'nonsense' mutaties creëren niet-functionele allelen. De homozygote 'nonsense' mutatie in *CaFAD2-C3* veroorzaakt een significante toename in het gehalte oliezuur (C18:1) van 25% ten opzichte van het gehalte in de splitsende wildtype, in combinatie met een afname in PUFA van ook ca. 25 %. In heterozygote vorm veroorzaakt de mutatie een intermediair effect (12 % toename in C18:1 en 11% afname in PUFA ten opzichte van het wildtype niveau). Voor *CaFAD2-C1* is alleen de heterozygote 'nonsense' mutatie gevonden. Deze vertoont een vergelijkbaar intermediair effect op de olie samenstelling als de heterozygote mutatie in

CaFAD2-C3. Planten met de ‘nonsense’ mutatie in *CaFAD2* konden niet getraceerd worden. Homozygote en heterozygote ‘missense’ mutanten in geconserveerde gebieden van *CaFAD2-C1* (Alanine naar Valine) en *CaFAD2-C2* (Glycine naar Asparagine) gaven een kleine toename in C22:1 (homozygoot 2% en heterozygoot 1% voor *CaFAD2-C1* en homozygoot 1% en heterozygoot 0.5% voor *CaFAD2-C2*, met alleen voor de mutatie in *CaFAD2-C1* daarnaast ook veranderingen in C18:1 en PUFA gehalte)

In het laatste hoofdstuk (**Hoofdstuk 6**) worden de resultaten besproken met nadruk op de functies van de verschillende *CaFAD2* genen in relatie tot de veranderingen die kunnen optreden in de biosynthese route van zaadolie. Vergelijkbare mutatiedetectie strategieën in andere gewassen worden besproken. Verder wordt gesteld dat door combinatie (‘stapeling’) van mutaties in de verschillende *CaFAD2* genen een verdere verhoging van C18:1 en verlaging van PUFA gehalte verkregen kan worden. Deze veronderstelling wordt ondersteund door vergelijkbare studies in andere gewassen.

Acknowledgements
This work was supported by the EU-project ICON (Industrial crops producing added value oils for novel chemicals) with co-funding of a Dutch R&D program Biobased Economy Kennis Basis Biobased Economy).

Acknowledgements

To end up this dissertation, I would like to express my gratitude to all the people who have contribute to my work and life particularly during the past five years.

My deepest gratitude is to my two supervisors, Dr. Robert van Loo and Dr. Elma Salentijn. I still remember the day I arrived in Wageningen. That is a typical Dutch day for weather and Robert has arranged everything done for my coming so that I can start my study immediately. Robert, thank you for trusting me with this project. In the mid-term, the disappointed results almost led to the end of the project but your confidence convinced the coordinate committee to continuously support our project. This is the reason why here we are. Your guidance and encouragement were of great importance in my scientific career. I love to hear that "Jihua, you did a very nice work" from you. It is my honor to be supervised by two supervisors. Elma, I fell much lucky to be supervised by you from the first day to now. I am greatly grateful for your patience and being confident to me. You are always available for discussions that came up with a lot of constructive advices. You endured my constant visits to your office, my irregular emails and questions. Particularly at the time of your illness, you are still in charge of the correction of my manuscript. Wish you all the best. Elma, from the conversation to you I got extensive improvements as well as much fun that I really enjoy. Also, on behalf of my family, acknowledge you for your frequent regards to them.

The utmost gratitude extends to my promotor, Prof. Richard Visser. Richard, your leadership and hard work left an indelible impress on my mind. Whenever I contact you for help and comments to the manuscripts, I can always get back a quick response. Without your quick correction and critical comments, I would not know when this thesis can be completed.

Our project is an internationally cooperated one. It would be impossible to start and finalize this project without cooperation with Prof. Bangquan Huang in Hubei University, China. Bangquan, I greatly appreciate you for offering me the opportunity to study in Wageningen. Thanks for your recommendation and support over the four years. In Wageningen, I am not the only one who was working on the crop, crambe. Frans, Iris and Weicong are the three other members in the team of crambe. I would like to say thank you for your direct involvement in my research.

As one of the members in Biobased economy group, I would like to thank all the biobasers for your kindly help and suggestions at group meetings as well as the sharing at group lunch. Dr. Luisa Trindade, our group leader, you are leading a vigorous group. I am grateful for your kindness particularly at that time Elma was not present due to illness and for the Christmas cards you signed every year, those it is hard to forget. Christel, thank you for making the crambe mutant population that is the fundament of my project. The biochemical lab is my second home lab where I grew up for GC analysis with the help of Luc, Annemarie and Heleen. Particularly Annemarie was in charge of all GC data collection, thank you! Our group meeting is one of the best places where I can find solutions and advices; I appreciate all of you for sharing experiences on the experiments. I would especially like to thank Peter and Xuan for being my paranymphs. I am looking forward to sharing the special day with you.

Have to say that I work in a wonderful lab. Danny, you are my first teacher in the lab and you also initialized my project. Your laughter made the lab work not a boring pipetting. Smoothly running such a large lab needs the efforts from all the labmates and technicians. Doret, Gert, Hanneke, Johan, Patrick, Marjon, Marian, Linda, Fien, Jos, Brian and other technicians, thanks to all of you. Alex and Jan, Elma and I have many conversations in your office; I am thankful for your enduring the 'noise' I made. Also, our department runs in order thanks to the efficient work of the secretariats, Nicole, Janneke, Letty and retired Annie.

At the last few months, I changed my desk. I would like to say thank you to my new officemates, Dirkjan and Gerard, you provided a quiet atmosphere that is of importance for thesis writing. A short talk with you every morning is a good start for a workday and these conversations taught me much about Dutch culture that is an unforgettable memory.

Many Chinese students are based on Wageningen. I benefited a lot of kind help and friendship from some of you. Weicong and Xi, you are my first and the best friends in WUR. I know where to go at weekend: your home where nice foods and drinks were awaiting me. Now you have a baby, wish your family happy forever! Pingping and Freddy, like me you two like to hang out in Weicong's apartment. The fun of talking and game-playing with you I will never forget. Ningwen, Ke Lin, Zhe Yan, Kaile, Juan, Dongli, Suxian, Wei, Zhao, Zheng, Tingting, Junwei, Dong Xiao, Shuhang, Ying Wang and many other Chinese friends, I am glad to meet you and the time with you is an unforgettable memory.

There is no word to completely express my gratitude to my family.

敬爱的父母，姐姐姐夫，谢谢你们对我一直以来的默默支持；亲爱的妻子，能够与你一起分享喜悦，分担疼痛是这一生最大的幸福，谢谢你带给我一个健康可爱的儿子。感谢一路有你们！

谢谢！

Thank you.

About the author

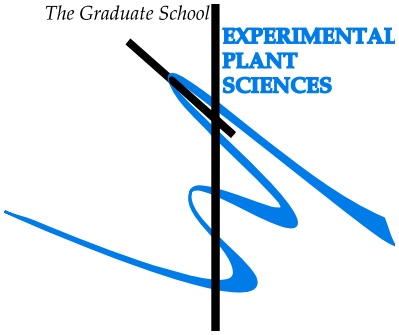
Jihua Cheng was born on 30th of April 1980 in Hubei province, China. He started his BSc study in the field of Agronomy at Hubei Agricultural College. He then obtained a master degree in Plant Breeding and Genetics at Huazhong Agricultural University in 2007. As of March 2009, he started his PhD in the Labortory of Plant Breeding at Wageningen University.

Cover design by Jihua Cheng

Printed at Wohrmann Print Service in Zutphen, the Netherlands

Education Statement of the Graduate School

Experimental Plant Sciences



Issued to: **Jihua Cheng**
Date: **12 June 2014**
Group: **Laboratory of Plant Breeding, Wageningen University & Research Centre**

1) Start-up phase	*date*
► **First presentation of your project**	
Modification of Erucic Acid level by TILLING	Dec 01, 2009
► **Writing or rewriting a project proposal**	
Modification of Erucic Acid and Glucosinolate levels by TILLING in Crambe abyssinica	Jul 30, 2009
► **Writing a review or book chapter**	
MSc courses	
► **Laboratory use of isotopes**	
Subtotal Start-up Phase	*7.5 credits**

2) Scientific Exposure	*date*
► **EPS PhD student days**	
EPS PhD student day, Wageningen	May 20, 2011
EPS PhD student day, Amsterdam	Nov 30, 2012
► **EPS theme symposia**	
EPS Theme 4 "Genome and Plasticity", Wageningen University	Dec 10, 2010
EPS Theme 3 "Metabolism and Adaptation", Wageingen University	Feb 10, 2011
EPS Theme 3 "Metabolism and Adaptation", Utrecht University	Apr 26, 2012
► **NWO Lunteren days and other National Platforms**	
ALW meeting "Experimental Plant Science", Lunteren	Apr 19-20, 2010
ALW meeting "Experimental Plant Science", Lunteren	Apr 04-05, 2011
ALW meeting "Experimental Plant Science", Lunteren	Apr 02-03, 2012
► **Seminars (series), workshops and symposia**	
Plant breeding Research Day (WICC, Wageningen)	Mar 08, 2011
Minisymposium" How to write a world class paper"	Apr 19, 2011
Minisymposium" Plant Breeding in the Genomics Era"	Nov 25, 2011
Plant Breeding Research Day (Arnhem)	Feb 28, 2012
► **Seminar plus**	
► **International symposia and congresses**	
Euro Fed Lipid Congress, Gdansk, Poland	Jul 09-13, 2011
Next Generation Plant Breeding	Nov 12-14, 2012
► **Presentations**	
Poster presentation at ALW meeting Lunteren	Apr 19-20, 2010
Poster presentation at International symposium of plant lipids,	Jul 11-16, 2010

Cairns, Austrialia	
Poster presentation at Euro Fed Lipid Congress, Gdansk, Poland	Jul 09-13, 2011
Oral presentation at EPS Theme 3 symposium, Utrecht University	Apr 26, 2012
► **IAB interview**	
Meeting with a member of the International Advisory Board	Feb 18, 2011
► **Excursions**	
Subtotal Scientific Exposure	*11.6 credits**

3) In-Depth Studies	*date*
► **EPS courses or other PhD courses**	
Bioinformatics-A User's Approach	Aug 29-Sep 02, 2011
Spring School: RNAi & The world of small RNA molecules	Apr 14-16, 2010
Introduction to R for statistical analysis	Oct 22-23, 2012
Basic statistics	Jun 22-23, 27-29, 2011
► **Journal club**	
member of literature discussion group at PBR	2009-2012
► **Individual research training**	
Training in Genetics laboratory of Hubei University, China	Nov 01-14, 2010
Subtotal In-Depth Studies	*10.5 credits**

4) Personal development	*date*
► **Skill training courses**	
TOEFL-iBT course	Feb 27, 2009
Techniques for Writing and Presenting a Scientific Paper	Oct 18-21, 2011
Information Literacy including Endnote Introduction	Jun 12-13, 2012
Career Orientation	Mar 01, 08, 15, 22, 28, 2013
► **Organisation of PhD students day, course or conference**	
► **Membership of Board, Committee or PhD council**	
Subtotal Personal Development	*3.3 credits**

TOTAL NUMBER OF CREDIT POINTS*	**32.9**
Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS which comprises of a minimum total of 30 ECTS credits	
** A credit represents a normative study load of 28 hours of study.*	