AUSTRIAN MARSHALL PLAN FOUNDATION
VIENNA | AUSTRIA

Gene editing of commercially important genes of oil crops

Marshall Plan Fellowship Final Report

Markus Freudhofmaier[1]

Research period:

7th March to 7th June, 2017

[1]Plant Biotechnology Unit, Department of Biotechnology, University of Natural Resources and Life Sciences, Vienna, Austria

[2]Center for Applied Genetic Technologies, University of Georgia, 111 River Bend Road, Athens, GA 30602, USA

Supervisors:

Ao.Univ.Prof. Dr. Margit Laimer[1]

Priv.-Doz. Dr. Fatemeh Maghuly[1]

Prof. Dr. Wayne Parrott[2]

Vienna, September 2017

BOKU VIBT Vienna Institute of BioTechnology

UNIVERSITY OF GEORGIA
UNIVERSITY OF GEORGIA

Acknowledgments

I would first like to acknowledge my supervisors Ao.Univ.Prof. Dr. Margit Laimer and Priv.-Doz. Dr. Fatemeh Maghuly, for giving me the opportunity to work on such an amazing topic and for all their help in theoretically and practically issues. I could not have imagined having better advisors for my Master Thesis.

I would also like to thank the Austrian Marshall Plan Foundation for the financial support of my research work in the USA. The Marshall Plan-Scholarship allowed me to work on current, complex research questions and to gain work experience in the field of plant improvement at an international level.

A special thanks to Prof. Dr. Wayne Parrott from the Center for Applied Genetic Technologies of the University of Georgia who enabled me to work in his renowned research team to learn about CRISPR/Cas9.I would also like to thank Dr. Peter LaFayette for all the troubleshooting. The vector construction could not have been successfully conducted without his help and support.

With a special mention to Jacob Hoyle, Louis Miles, Gary Orr, Tim Chappell, Evan McCoy and Lauren Lail. It was fantastic to have the opportunity to work and have fun with you during my visit in your lab.

Finally, I must express my very deep gratitude to my family and to my girlfriend Claudia for providing me with unfailing support and continuous encouragement throughout my years of study.

Abstract

Jatropha curcas L. is an undomesticated perennial plant belonging to the *Euphorbiaceae* family. It is native to America, but is distributed widely in the tropical and subtropical areas. The high production of oil bearing seeds makes it a potential candidate for a sustainable biofuel crop. To achieve a competitive biofuel cultivar it is necessary to eliminate existing deficits by improving the genetic background. The current work aims at providing the molecular tools to establish a knowledge base for genes involved in the biosynthetic pathways related to commercially important traits by CRISPR/Cas9 knock-out constructs. Three genes related to the fatty acid biosynthesis, two genes involved in allergen content and one gene responsible for toxicity were selected. The gene structures were analyzed and guide RNAs for potential target sites were designed. The NEBuilder Assembly tool was used to construct gRNA-Cas9 binary vectors with the p201N:Cas9 plasmid carrying up to 6 gRNAs.

Contents

Introduction

Jatropha curcas (L.)

General Plant Characteristics

The genus *Jatropha* is belonging to the subfamily *Crotonoideae* within the family *Euphorbiaceae*. The center of origin is Central America and Mexico. Today it is distributed widely in tropical and subtropical areas all over the world. The plant architecture is very variable, single to multiple branched stems build up a big shrub or a small tree with a height of 3 to 5 m. It is a perennial plant producing oil bearing seeds for up to 50 years. The leaves are green to pale green 5 to 7 lobed(Figure 1).The annual dormancy, accompanied with leaf fall in the dry season, is induced by rainfall and temperature fluctuation (Contran et al. 2013).

Figure 1. .*Jatropha curcas* leaf.

The initial root growth of *J. curcas* is very predictable. It starts with a deep taproot and four main secondary roots. Particularly, in (semi)arid environments the taproot continuously supplies the plant with water by providing access to deeper soil horizons. This enables the plant to grow in areas which are susceptible to desertification. The lateral roots and the dense fine

root net in the topsoil prevent effective soil erosion caused by wind and water (Reubens et al. 2011).

From Flower to Seed

J. curcas is a monoecious plant with separate male and female inflorescences. Nevertheless some accessions are dioecious and hermaphrodites. This phenomenon indicates a complex sex determination system like sex chromosomes (Adriano-Anaya et al. 2016).

The flowers are greenish-yellow and composed of five pedals, five sepals and five nectarines (Figure 2). The male flower is 5 to 10 mm in diameter and has ten stamens. The female flower is bigger with 7 to 15 mm and has one ovary with three ovules. The flower morphology is very variable between different accessions according to male flower diameter, female pedal length and male nectary length. The flowers are located on the end of mature branches (Adriano-Anaya et al. 2016).

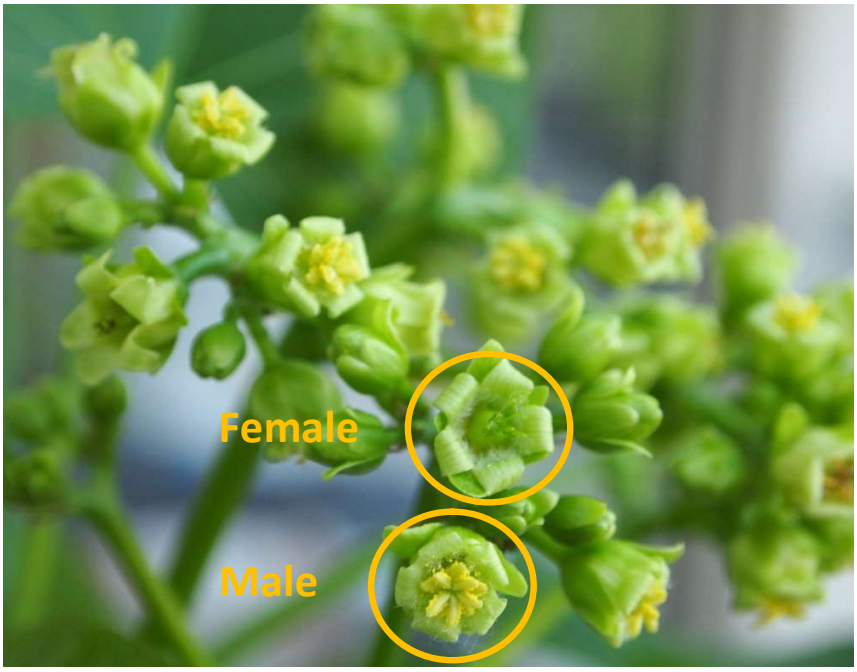

Figure 2. Female and Male flowers of *Jatropha curcas* (PBU, BOKU).

The high male/female flower ratio of about 13:1 – 29:1 is genetically defined and has to be improved for higher seed yields (Ahmad and Sultan 2015). The generally three-lobed capsular

fruits, carrying 2-3 seeds, have a diameter of 3 cm, are 4 cm long and weigh between 2.0-3.5 g (Heller 1996). According to Silip et al. (2010) fruits are physiologically developed within 21 to 35 days after fruit set. After maturity the green fruits (Figure 3) turn to yellow (fully ripe fruits) in 2 to 4 days and to brownish-black (fully senesced fruits) in 3 to 9 days. The highest seed-oil content was found at the brownish-black maturity stage.

Figure 3. Fruits of *Jatropha curcas* (PBU, BOKU).

The seeds are about 2 cm long, 1 cm thick and weigh 0.5-0.7 g. The shape is ellipsoid, triangular-convex.30-40 % of total seed weight is made up of the brown-black seed shell whereas 60-70 % consists of the white kernel. The seed oil (30-35 % per dry mass) is stored mainly in the kernel. Beside the oil content the seeds contain a high amount of proteins with high level of essential amino acids, expect of lysine (Contran et al. 2013).

Genetics and Breeding

J. curcas is a diploid plant with 22 chromosomes and a genome size of about 370 MB (Dahmer et al. 2009). The genetic diversity is very high in accessions from Central America and Mexico

(Maghuly et al. 2015), thus Mexico is supposed to be the center of origin (Pecina-Quintero et al. 2014). The level of heterozygosity is relatively low due to a high self-pollination rate (Maghuly et al. 2015, Montes et al. 2016). Two whole genome sequences from *J. curcas are public available*. The first was published from the Kazusa DNA research institute in Japan (www.kazusa.or.jp/jatropha). The *J. curcas* line "Palawan" from the Palawan Island in the Philippines was used for genome sequencing and eleven further lines from different origins for diversity analysis by Sanger sequencing and new-generation multiplex sequencing technologies (Sato et al. 2011). The second whole genome sequence available was published from the Chinese Academy of Science. The *J. curcas* inbred line GZQX0401 was used for genome sequencing by next generation Illumina sequencing method (Wu et al. 2015).

Economic Importance and Productivity

J. curcas as plant itself as well as its products and byproducts are very versatile in use. The trees are used to defense environmental problems like erosion (caused by wind and water) and desertification (Reubens et al. 2011). The high drought tolerance due to a deep tap root and the possibility to shift between C3- and CAM-metabolism allows the plantation of *J. curcas* in non-arable land which leads to the possibility to produce biofuels without a competition to food production (Maes et al. 2009). The deep tap root leads to circulation of down leached nutrients and minerals. Soil aggregation in the top soil is improved by the secondary root net and the annual leaf shedding leads to increased soil organic matter (Contran et al. 2013).

Beside the high amount of oil (285-656 mg/g) and protein (184-377 mg/g), the seeds of most accessions contain phorbol esters (PEs), the most toxic compounds in *J. curcas*, with tumor promoting and cytotoxic effects(Montes et al. 2013). The high level of PEs (up to 10 mg/g seed) makes the plant unsuitable for human consumption or for animal feed (Montes and Melchinger 2016).

All parts of the plant are used for several human and veterinary afflictions in traditional medicine. Alcoholic extractions are used as antibiotic against different pathogens like *E. coli*. Leaves are used against Malaria and to treat rheumatic pain and the stem sap (latex) is used to

stop bleeding of wounds. However further investigations and scientific proof of *J. curcas* as medicinal plant is needed (Heller 1996).

The seed oil is used for biodiesel production, soap making and some non-toxic varieties from Mexico with low PE-content are also used for human consumption (Contran et al. 2013).

The reported seed yields vary very strongly between 0.4 t and 12 t $ha^{-1}$and are affected by many factors, like genetic background, plant age, environmental conditions, soil type and management. Due to the undomesticated state of *J. curcas,* the yields between single trees within a plantation are strongly heterogeneous (Contran et al. 2013). According to Wani et al. (2016) on wasteland up to 3 t ha^{-1} dry seeds are achievable, but despite its drought tolerance *J. curcas* requires high amounts of annual rainfall (750-1000 mm) for economic productivity.

Biodiesel

The production of biodiesel (mono-alkyl esters) from seed oil is the economically most important use of *J. curcas*. There are several methods to produce biodiesel from vegetable oil, but the most common is transesterification. The four major groups of fatty acids in the seed oil of *J. curcas* (oleic acid, linoleic acid, palmitic acid and stearic acid) can be transesterified to alkyl esters in the presence of an acyl acceptor like alcohol (Jain and Sharma 2009). Compared to petroleum-based diesel, biodiesel has several advantages like reduced pollutants emission, a better performance of diesel engines due to a higher cetane number and the ability to locally produce renewable energy (Juan et al. 2011).Furthermore, there is no increase of CO_2 levels in the atmosphere and it has lower sulphur content as well as a higher biodegradability than conventional diesel (Koh and Ghazi 2011).

Biofuelproperties

The dimension-less cetane number (CN) describes the ignition quality of a fuel, i.e. the higher the cetane number, the higher the quality. An increasing chain length and increasing saturation of fatty acids increase the CN. A high CN also reduces exhaust emissions like nitrogen oxide (NOx) (Knothe 2008).

There are several low-temperature (cold flow) properties for biodiesel like the cold point (CP). When fuel is cooled down, the CP is the temperature at which the first solids are visible. These properties are determined by the melting point of fatty esters, which decrease with an increasing level of unsaturation and a decreasing chain length (Knothe 2008). The CP of fatty acid methyl esters of *J. curcas*is 10°C. The CP of biodiesel from soybean oil is 0°C due to the high amount of polyunsaturated fatty acids and from conventional diesel it is around – 16 °C (Ye et al. 2013).

Lower viscosity of biofuels leads to a better atomization in the combustion chamber. Therefore, the reduction of kinematic viscosity is the major reason for the transesterification of oils to biodiesel. Viscosity of the produced biodiesel can be reduced by decreasing fatty acid chain length as well as increasing levels of unsaturation (Knothe and Steidley 2005).

The oxidative stability is reduced by double bounds of unsaturated fatty acids, which increase the susceptibility to reaction with oxygen (Zahira et al. 2014). Linoleic and linolenic acids are very susceptible due to methylene-interrupted double bounds in the fatty acid chains. The mono-unsaturated oleic acid is the most oxidatively stable fatty acid. However, to meet international biodiesel quality standards of ASTM (American Section of the International Association for Testing Materials) as well as European standards (EN 14214) it is necessary to use antioxidant additives for all biodiesel fuels based on unsaturated fatty acids (Knothe 2008).

Jatropha curcas is supposed to be a potential source of biofuel production (Juan et al. 2011). The dominant fatty acids in *J. curcas* seed oil are5-8% stearic acid (18:0), 13-17% palmitic acid (16:0), 28-42% linoleic acid (18:2) and 37-52%oleic acid (18:1)(Nayak and Patel 2010, Juan et al. 2011, Abdullah et al. 2013, Ogbogu et al. 2014, Fernandes et al. 2015, Barros et al. 2015).

The high amount of poly-unsaturated linoleic acid leads to less oxidative stability and a reduced CN. The low-temperature properties are negatively affected by the high amount of saturated palmitic acid. For good fuel properties it is necessary to reduce the level of saturated fatty acids (stearic acid and palmitic acid) and polyunsaturated fatty acids (linoleic acid). There is no fatty acid in *J. curcas* available, which meets all criteria at the same time, so it is necessary to make a compromise between cold flow properties on the one hand and oxidative stability and cetane number on the other hand. High levels (> 70%) of mono-unsaturated oleic acids (18:1) and low

levels of saturated fatty acids (< 10%) in seed oil would lead to ideal biodiesel properties (Ye et al. 2013).

Biosynthesis Pathway of Fatty Acids in Plants

Fats and oils are triglycerides composed of glycerol (three carbon alcohol) and three esterified long-chain fatty acids (Salisbury and Ross 1992).

The biosynthesis of triglycerides takes place in the endoplasmic reticulum (ER).Fatty acids (FAs) are activated to acyl-coenzym A (acyl-CoA) and esterified to glycerol 3-phospate to form triacylglycerol through the intermediate products lysophosphatidic acid, phosphatidic acid and diacylglycerol (Figure 4) (Maghuly and Laimer 2013).

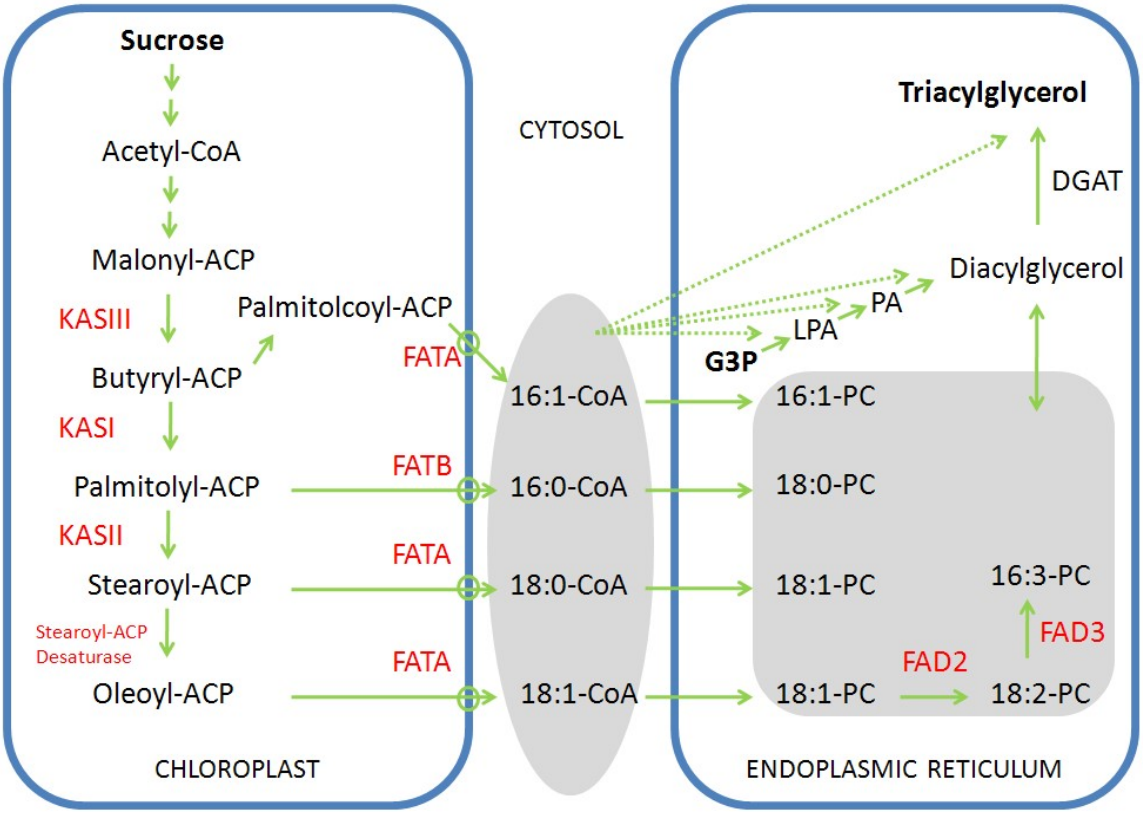

Figure 4. Simplified model of the biosynthesis pathway of triacylglycerol in plants. CoA, coenzyme-A; ACP, acyl carrier protein; KAS, beta-ketoacyl-ACP synthase enzymes; FATA, fatty ACP thioesterases; FATB, fatty ACP thioesterases;thioesterases B; FAD2, 1-acyl-2-oleoyl-sn-glycero-3-phosphocholine delta 12-desaturase, FAD3, linoleoyl-phosphatidylcholine ω-3 desaturase;G3P, glycerol-3-phosphate; LPA, lysophosphatidic acid; PA, phosphatidic acid (Maghuly and Laimer 2013, modified)

The complex biosynthesis pathway of plant lipids involves several fatty acid biosynthetic enzymes. The biosynthesis of fatty acids is well studied but the regulation of the produced amount and type of fatty acids are almost unknown and further investigations are necessary (Gu et al. 2012).

The fatty acid biosynthesis takes place in the plastids. Sucrose serves as carbon source for the production of acetyl-coenzym A (Acetyl-CoA) the precursor of fatty acids. The primary products of the fatty acid biosynthesis are the long-chain fatty acids palmitate (16:0) and stearate (18:0). Secondary FAs can be derived by desaturation, elongation and further modification like oxygenation and hydroxylation (Schopfer and Brennicke 2010).

The beta-ketoacyl-acyl carrier protein (ACP) synthase enzymes I, II and III (KASI, KASII and KASIII) belong to the type II fatty acid synthase (FAS) complex in plastids. FAS enzymes are responsible for fatty acid chain lengths determination as well as the level of saturation. This set of enzymes generates mainly palmitic acid (C16:0), stearic acid (C18:0), palmitoleic acid (C16:1), oleic acid (C18:1) and the polyunsaturated linoleic acid (C18:2) (YU et al. 2015).

A second set of fatty acid chain length termination enzymes in the plastids are the fatty acid acyl carrier protein (ACP) thioesterases A and B (FATA, FATB)(DANI et al. 2010).

The FA desaturases FAD2 and FAD3 in the endoplasmic reticulumare responsible for the production of linoleic acid (18:2)and α-linoleic acid (18:3), respectively (Wu et al. 2013).

The biodiesel quality is mainly determined by the fatty acid composition of *J. curcas* seed oil (Knothe et al. 2008). In order to gain a competitive biofuel crop it is necessary to modify the FA biosynthesis pathway by genetic engineering (Maghuly and Laimer 2013).

Genes involved in Toxicity

Most accessions of *J. curcas* are toxic and contain anti-nutritional compounds (Tanya et al. 2011). Diterpenoids such as phorbol esters (PEs) and ribosome-inactivating proteins (RIPs) like curcin make the press cake unsuitable for the use as by-product for livestock feed (Maghuly and Laimer 2013). The press cake is the solid remaining after oil extraction by pressing. From 1 ton of seeds a press cake of about 264 kg can be obtained. The press cake is rich in proteins

(about60 %), which consist of amino acids with high nutritional values like sulphur amino acids (Makkar and Becker 2009). Moist heating destroys the heat labile trypsin inhibitors, but PEs are insensitive to heat. Different methods to reduce the PE content of *J. curcas* press cake have been developed, but most of them are environmentally unfriendly, expensive or time consuming (Gogoi et al. 2015).

To obtain PE free accessions it is necessary to knock out their biosynthesis pathway (Nakano et al. 2012). The terpenoid biosynthesis starts with the generation of dimethylallyl diphosphate (DMAPP) and isopentenyl diphosphate (IPP). IPP and DMAPP are synthesized during the methylerythritol phosphate (MEP) pathway and the mevalonic acid (MVA) pathway. The MEP pathway operates in plastids and begins with the production of 1-deoxy-Dxylulose 5-phosphate (DXP) by pyruvate and glyceraldehydes phosphate condensation. The MVA pathway operates in the cytosol as well as in the ER and peroxisomes and starts with acetyl CoA condensation to 3-hydrocy-3-methylglutaryl CoA. After IPP and MAPP production the synthesis of diterpens are similar regardless of their production site (Maghuly and Laimer 2013, Chang et al. 2013). Geranylgeranly pyrophosphate (GGPP) is synthesized by the condensation of three IPP molecules with DMAPP (Singh and Sharma 2015). Casbene (monocyclig diterpene), which is synthesized by the enzyme casbene synthase due to conversion of GGPP, is the precursor of tigliane (tetracyclic diterpene).Finally two fatty acids are esterified to the tigliane skeleton to form phorbol esters (Nakano et al. 2012).

Genes involved in Pathogen Response

Genes responsible for the production of pathogenesis related (PR) proteins are expressed at increased levels during infection by pathogens and by environmental stress (Stintzi et al. 1993). The first PR proteins are found in tobacco in the 1970s (van Loon and van Kammen 1970). The PR proteins are very diverse in function and structure so that they are classified into 17 families (Puehringer et al. 2003).

Several classes of PR proteins are involved in allergenicity. The PR-2 class which consist of enzymes responsible for the hydrolysis of β-1,3-clucans. The PR-3 and PR-4 classes which are enzymes responsible for the hydrolysis of chitin. Thaumatin-like Proteins and osmotins,

belonging to the class PR-5. The PR-10 proteins with still unknown function and the non-specific lipid transfer proteins PR-14 (Maghuly et al. 2009).

Allergens are substances, which induce an allergic reaction, are able to bind IgE antibodies and have the property to sensitize (i.e. trigger the production of antibodies by the immune system) susceptible patients (Aalberse 2000).

Beside the role of pathogen defense, PR proteins have several additional functions in developmental processes (Lotan et al. 1989). Genetic engineering seems to be the perfect tool to reduce allergen content in crops. In contrast to conventional breeding the spatial and temporal control of PR genes are feasible (Maghuly et al. 2009).

Genetic Improvement of Crops

Long before the discovery of Mendel's laws people improved plant characteristics by conventional breeding methods. The selection of desirable traits and their combination by sexual crossings resulted in improved crops over numerous generations. In perennial crop species, this strategy needs up to 15 years to gain a new variety (Key et al. 2008). In general breeding methods can be classified into three groups. Selection breeding, combination breeding and hybrid breeding are based on different genetic methods to obtain genetic variation and to select stable variants (Kuckuck 1991). Since the 1990s the molecular marker-assisted breeding with DNA markers has been extensively used in order to accelerate the improvement of novel crops (Guo-Liang 2015).Mutations are induced in plant genomes since several decades to increase the genetic variability. The use of radiation and chemical mutagens result in random changes in the exposed plant genome. Mutation breeding is now a common tool for plant breeders worldwide. Improvement of single traits can be made easily but the reduction of background mutations and the selection of rare mutants may a considerable challenge (Forster and Shu 2011).

To be able to transfer novel genes directly to host plants, genetically modification is necessary. Genetically modified (GM) crops are produced by the insertion of genes or gene elements from a donor organism directly to the genome of a host plant. The transferred gene can be originated from a relative plant or from non-related species like microorganisms. GM crops

have been developed and used successfully in many countries. However especially GM crops of the first generation, like herbicide tolerant and pest resistant crops, are fallen into disrepute (Georges and Ray 2017).

Genome edited (GenEd) crops are produced by genome mutations at defined sites without background mutations. This technology provides a fast and controlled approach to alter specific gene functions. The expression of genes can be silenced or enhanced in a temporal and spatial manner (Georges and Ray 2017).

Genome Editing Tools

Different tools have been developed for precise genome editing by inducing double stranded DNA breaks (DSBs) with programmable nucleases and nickases like zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and CRISPR/Cas (clustered regularly interspaced short palindromic repeat/CRISPR-associated nuclease) (Guha et al. 2017).

ZFNs are synthetic proteins with separate DNA-cleavage and DNA–binding domains, connected by a short linker sequence (Figure 5). By designing the recognition domain it is possible to control the site of cleavage. The DNA-binding domain consists commonly of three zinc fingers (ZFs).Each connects with 3 bp of target DNA.The FokI cleavage domain, isolated from *Flavobacterium okeanokoites* (Sugisaki and Kanazaw 1981), must dimerize to become active. So it is necessary to design two ZFNs, one for the complementary and one for the non-complementary DNA strand resulting in an 18 bp recognition site (Carroll 2011).

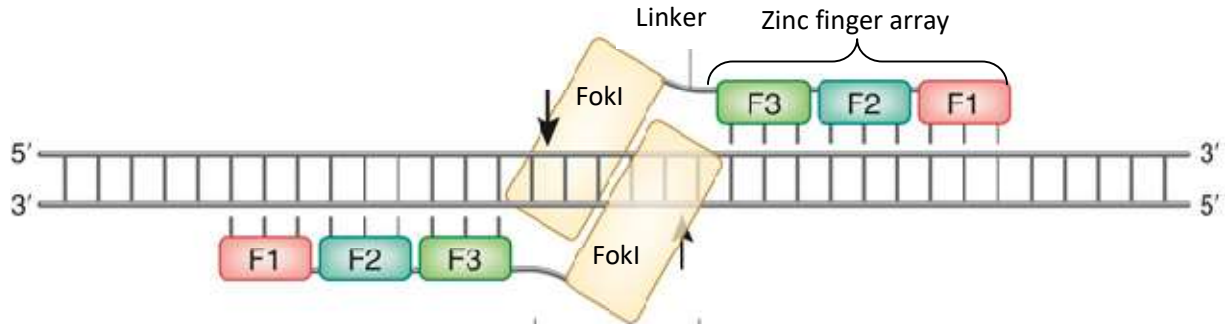

Figure 5.A pair of ZFNs bound to DNA. The arrows symbolize the common cleavage site (Carroll 2011, modified).

TALENs have been developed as an alternative tool to ZFNs. Similar to ZFNs, TALENs consist of a programmable DNA-binding domain and an unspecific FokI cleavage domain (Joung and Sander 2013). The TALENs recognition domain consists of arrays of multiple 34-amino acid repeat sequences. A single DNA nucleotide can be recognized by the amino acids at the 12^{th} and 13^{th} position (Guilinger et al. 2014).

Both, TALENs and ZFNs have been used widely for precise genome editing. However some disadvantages are associated with these methods based on protein-DNA interactions for targeting (Sander and Joung 2014). The construction of large recognition domains are very labor and cost intensive and the rate of failure is relatively high, especially for ZFNs. The CRISPR/Cas system is a cheap, fast and precise alternative to the previous gene editing tools (Belhaj et al. 2015).

CRISPR/Cas system

The Natural CRISPR/Cas Immune System

In the year 1987 clustered repeats in bacterial genomes were first reported without recognizing the huge potential of this versatile system. In 2005 it was found that CRISPR/Cas is involved in bacterial immune defense (Hsu et al. 2014).

With this system bacteria and archaea can incorporate short fragments from foreign DNA (invading phage or plasmid DNA) into the CRISPR loci. Initially, pre-CRISPR RNAs (crRNAs) are transcribed and processed to mature crRNAs. Subsequently the crRNAs form a complex with RNA-guided endonucleases (RGENs) and induce double strand brakes at homologues sequences (protospacer). Cleavage occurs only in the presence of a RGEN specific protospacer adjacent motif (PAM) (Mali et al. 2013) (Figure 6).

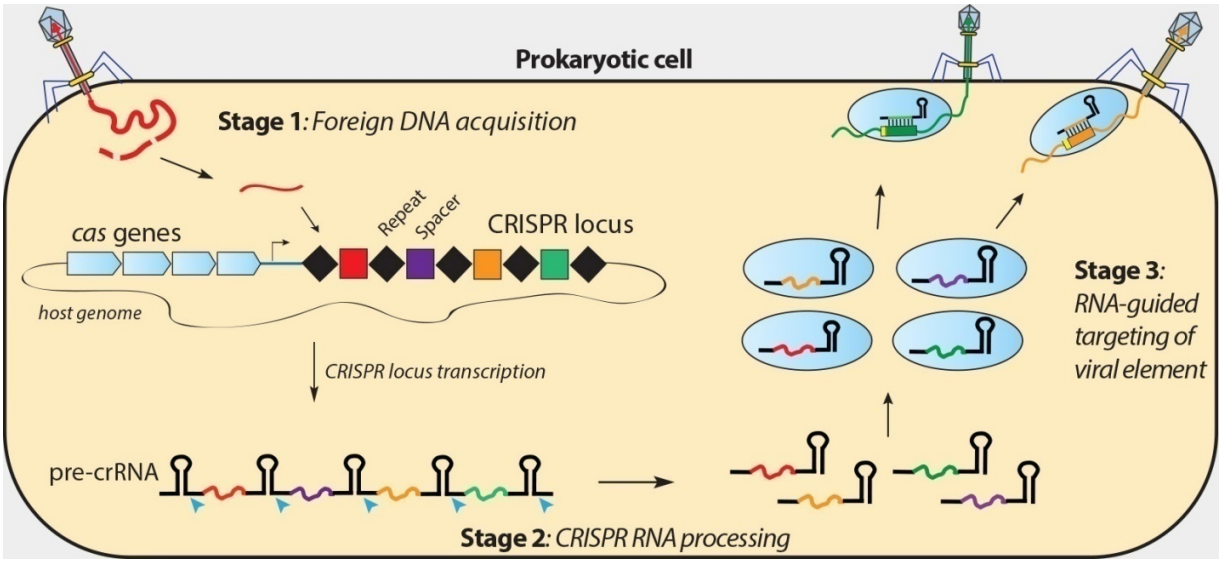

Figure 6. CRISPR/Cas immune system. In the first stage the foreign DNA (spacer sequences) were integrated into the CRISPR locus of the prokaryotic genome. In the second stage the pre-crRNA (pre CRISPR RNA) transcribed from the CRISPR locus will be processed to mature crRNAs by the hybridization of pre-crRNA and tracrRNA (trans-activating crRNA). In the third stage the crRNAs form a complex with the RNA-guided endonucleases (RGENs) which cleave foreign DNA at sequences that match the crRNA adjacent to a PAM (protospacer adjacent sequence) (The Doudna Lab, http://rna.berkeley.edu/crispr.html).

The CRISPR/Cas systems are very diverse and several attempts have been made to reclassify the systems since the first classification proposed by Haft et al. (2005) (Makarova et al. 2011, 2015, Koonin et al. 2017). According to Koonin et al. (2017) the CRISPR/Cas systems are separated into two classes with several types and subtypes. The class 1 CRISPR systems (type I, III and IV) use numerous Cas proteins to form a complex with crRNAs, whereas the class 2 CRISPR systems (type II, V and VI) use only one Cas protein. The most studied type II CRISPR system uses the RGEN Cas9.

Programmable RGENs for Precise Genome Editing

In the natural system the guide RNA (gRNA) is composed of the pre-crRNA, which hybridizes with a trans-activating crRNA (tracrRNA) during crRNA processing. To make the CRISPR-Cas9 system a fast and versatile gene editing tool, the pre-crRNA and tracrRNA were fused to a single guide RNA (sgRNA). By designing the first 20 nt of this engineered sgRNA sequence it is possible

to precisely guide the Cas9 RGEN to induce a double stranded DNA break (DSB) in the gene of interest. Imprecise repair by the subsequently activated non-homologous end joining pathway (NHEJ) leads to insertions or deletions, which result in frameshift mutations and successful knock-out of the target gene (Belhaj et al. 2015). In 2013, the first genome engineering was demonstrated in eukaryotic cells and in 2014 the first genome wide functional screening was carried out with the CRISPR/Cas9 system (Hsu et al. 2014).

Beside gene knock-out the CRISPR/Cas tool can be used for several applications. The deletion of whole gene clusters is possible by the simultaneous expression of two or more gRNAs. Due to the cell's homology directed repair (HDR) of DSBs, genes can be knocked-in by providing template DNA with overlapping flanking regions (Belhaj et al. 2015).

The Cas9 Protein

The Cas9 protein is composed of two connected lobes, a recognition (REC) lobe and a nuclease (NUC) lobe. The NUC lobe has a PAM-interacting (PI) domain and two nickase domains, HNH responsible for cleavage of the complementary strand and RuvC responsible for cleavage the non-complementary strand. The REC lobe is responsible for Cas9-sgRNA complex formation. During gRNA loading the confirmation of Cas9 changes from an inactive to an active form by building a central channel where the RNA-DNA heteroduplex will be positioned. The Cas9-sgRNA complex (Figure 7) scans the double stranded DNA until the PI domain recognizes a PAM sequence. Subsequently the DNA is melted and the complementary strand connects with the guide RNA. After RNA-DNA heteroduplex formation HNH and RuvC cleave the two DNA strands three bases upstream of the PAM site (Belhaj et al. 2015).

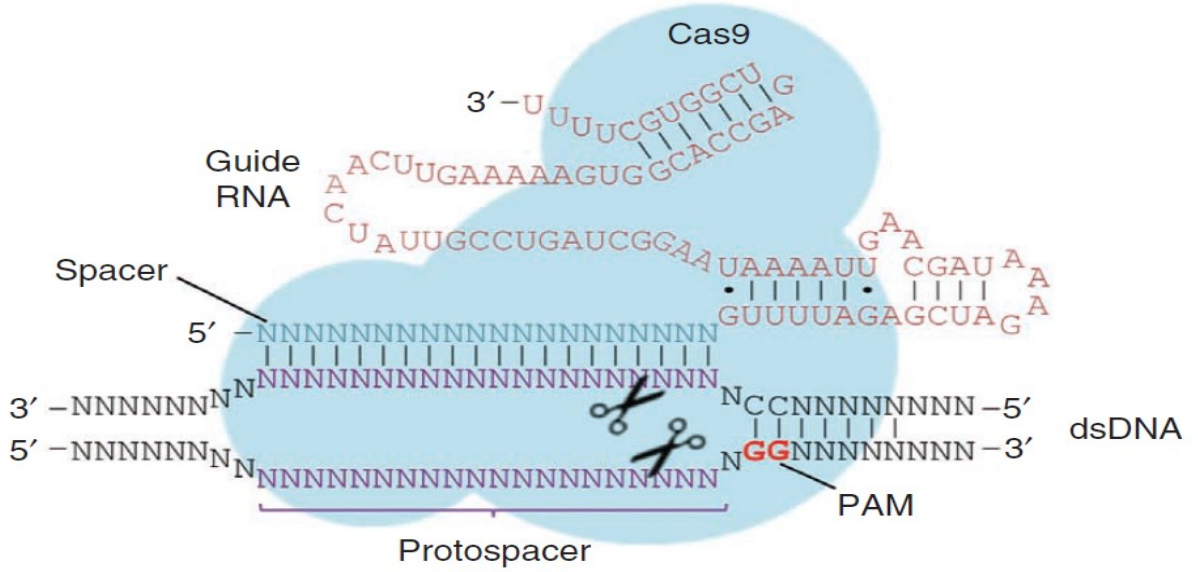

Figure 7. Cas9-sgRNA targeting complex(Mali et al. 2013).

PAM Sequences

The most widely used RGEN for gene editing is the Cas9 protein from *Streptococcus pyogenes* (SpCas9) which recognize the PAM sequence 5'-NGG-3'. To be more versatile by searching target sites in the gene of interest it is possible to use Cas9 proteins from other bacteria like *S. aureus, S. thermophilus, Neisseria meningitides* or *Brevibacillus laterosporus,* which have different PAM specificity. However recently, attempts have been made to design artificial spCas9 variants with altered PAM specificity (Andres et al. 2016).

Cas9 Specificity

The CRISPR/Cas9system is very effective, cheap and less labor intensive compared to previous sequence specific nucleases like zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) (Belhaj et al. 2015). However, CRISPR is not an unfailing system. Unexpected mutations, so called off-target effects could be found in several studies (Schaefer et al. 2017, Zhang et al. 2015, Pattayak et al. 2013),which lead to concerns for biomedical and clinical applications. In contrast to human and animal cell systems, the off-target effects in plants are very rare (Peterson et al. 2016). Nevertheless different strategies were developed to

increase the specificity of Cas9 cleavage. The use of double nicking by two Cas9 nickase mutants requires the simultaneous binding of two gRNAs (Ran et al. 2013). By truncating the gRNA to 17-18 nucleotides length the binding energy at the gRNA-DNA interface is reduced and the mismatch sensitivity increased (Fu et al. 2014).

Cas9 Activity

Beside the Cas9 specificity, the activity of the protein is one of the major concerns in the CRISPR/Cas system (Peng et al. 2015). Liang et al. 2016 suggest different criteria for the selection of efficient gRNAs for plant genome editing. The G/C content should be between 30% and 80%, the maximum size of gRNAs should not exceed 12 nucleotides and the secondary structure has to be intact instead of the first stem loop (Figure 8).

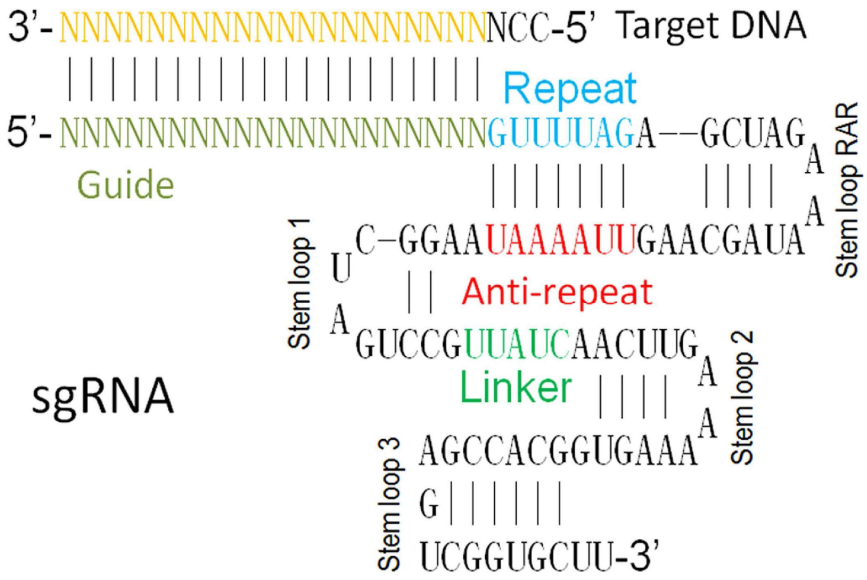

Figure 8. Schematic diagram of gRNA secondary structure (Liang et al. 2016).

Delivering Genes to Plants

In the first step of plant genetic engineering a suitable vector plasmid has to be designed and constructed before it will be delivered to the plant genome (Hansen and Wright 1999). There are several physiological and biological methods available to transfer specific DNA sequences to plants. Particle bombardment with a 'genegun' as well as *Agrobacterium* mediated transformation are the most commonly used systems (Lorence and Verpoorte 2004).

Particle Bombardment

A gene gun is a biolistic device to shoot DNA coated metal particles to plant cells in order to stable integrate DNA to the host chromosomes. For genetic transformation small gold or tungsten particles are coated with DNA by precipitation with spermidine and calcium chloride. The gen gun device consists of an acceleration tube which is filled with helium up to a specific pressure defined by the rupture disk. After the maximum pressure the disk rupture and release the helium to a plastic macrocarrier on which the DNA coated particles (microcarrier) are placed. The macrocarrier is shot down to a stopping screen and release the microcarrier which penetrate the target cells with a high velocity. Inside of the cell the DNA elutes off the microcarrier and can be stable integrated into the host chromosomes (Kikkert et al. 2006).

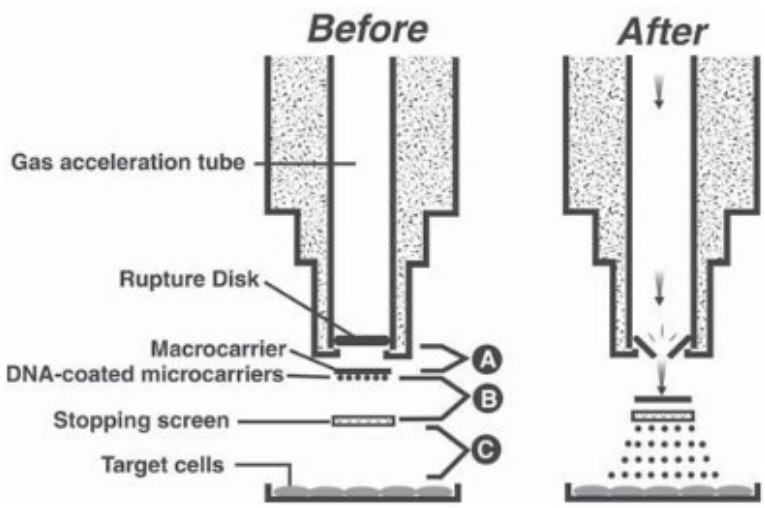

Figure 9. The biolistic bombardment process (Kikkert et al. 2006).

Agrobacterium tumefaciens

The genus *Agrobacterium* contains several plant disease causing soil bacteria. *Agrobacterium rhizogenes* cause hairy root disease, *Agrobacteriumvitis* cause crown gall of grape, *Agrobacterium rubi* cause cane gall disease and *Agrobacterium tumefaciens* is responsible for crown gall disease after infection of wounded plant cells (Escobar and Dandekar 2003). The identification of *Agrobacterium* as plant tumor inducing bacteria was over 100 years ago. Since then, *Agrobacterium* revolutionized plant genetic engineering and is today commonly used as vector to create transgenic plants (Nester 2015).

The transfer and integration of a specific region of a tumor-inducing (Ti) plasmid into a plant cell is the molecular basis of *A. tumefaciens* mediated genetic transformation. Ti plasmids have a size of 200 to 800 kbp and transfer naturally 10 to 20 kbp DNA (T-DNA) from the T-region which is flanked by highly homologous 25 bp border sequences. This transfer is caused by the activation of virulence (*vir*) genes which are located on the Ti plasmid (Glevin 2003). In order to obtain a plant transformation tool, the Ti plasmid of *Agrobacterium* was engineered. Genes responsible for opine biosynthesis and tumor formation in the T-region were replaced by a gene of interest as well as plant selectable marker. As a result, the genes placed between the specific border sequences will be transferred to plant cells (Pacurar et al. 2011). In the 1980s the *Agrobacterium* transformation tool was improved by the separation of the *vir* genes and the T-region by establishing the binary plant vector strategy. A disarmed Ti-plasmid (without T-region) carrying the *vir* genes is responsible for plant infection and T-DNA transfer whereas the second Ti plasmid contains the T-DNA. Due to the reduced size of the Ti plasmid which carries the T-region, the T-DNA can be easily manipulated using *E. coli* as host (Hoekema et al. 1983).

The genetic manipulation of *Agrobacterium* is the first step in *Agrobacterium* mediated plant transformation. There are three methods which are commonly used to transfer engineered plasmids to *Agrobacterium*. The most efficient method is electroporation. Pores in the bacterial lipid membrane are created by an electric shock and large DNA molecules like purified plasmids can easily enter the cell. By removing electrolytes during several washing steps, electro-competent cells can be easily prepared. For electroporation an electroporator and specific cuvettes are necessary. An easy and cheap alternative is the freeze/thaw method. A rapid

change in temperature and exposure to divalent cations change cell membrane fluidity and allow DNA uptake. However the exact mechanism of the freeze/thaw method is not well understood. Another method to transfer DNA to *Agrobacterium* is called triparental mating. Two *E. coli* strains are used to move a plasmid from *E. coli* to *Agrobacterium* (Wise et al. 2006).

Plant Tissue Culture Prerequisites

The technique of growing single cells, complex tissues or whole organs of plants in artificial nutrient medium under aseptic conditions is referred to as plant cell or tissue culture (Kumar 2003). Since the first attempts to cultivate isolated plant cells by Gottlieb Haberlandt (Haberlandt 1902) the techniques have been improved and are used routinely in different disciplines like plant breeding, production of pharmaceutically important compounds, micropropagation, elimination of pathogens, in-vitro conservation and genetic engineering of crop plants (Laimer and Rücker 2003).

To obtain optimal growth as well as morphogenesis of plant tissues in vitro it is necessary to support the cells with nutrient. Different standard media were developed like the frequently used Murashige and Skoog (MS) medium. The main components of most plant tissue media are macronutrients, micronutrients, nitrogen supplements, vitamins and myo-inositol, carbon sources and growth regulators. The optimum concentrations vary among the cultivated plant species (Saad and Elshahed 2012).

Synthetic plant growth regulators and natural phytohormones are classified into five main groups: Auxins, cytokinins, abscisic acid, gibberellins and ethylene. The most important phytohormones for successful regulation of plant growth and organization development are auxins and cytokinins as well as their ratio. Auxins are mainly responsible for cell growth expansion, initiation of cell division and the organization of meristems. In organized tissue, auxins promote root growth, are responsible for the maintenance of apical dominance and delays leaf senescence. Indole-3-acetic acid (IAA) and indole-3-butyric acid (IBA) are the most common natural occurring auxins, whereas 2,4-dichlorophenoxyacetic acid (2,4-D) and 1-naphtaleneacetic acid (NAA) are commonly used synthetic auxins. To induce lateral bud formation and to stimulate cell division in cultures the plant growth regulator group cytokinins

are commonly used. The regulation of cell division is controlled by the combination of auxins and cytokinins by influencing different cell cycle phases. In plant tissue culture zeatin and 2-isopentenyladenine (2-ip) are one of the most used natural cytokinins. There are also synthetic cytokinins like kinetin, benzyladenine (BA) and N-benzyl-9-tetrahydropyranyl-adenine (BPA) (Gaspar et al. 1996).

Beside the optimal medium composition and phytohormone concentration the culture environment is also an important factor. The light intensity, photoperiod and temperature strongly influence the process of differentiation of in vitro explant tissues (Thorpe 1994).

For successful plant genetic engineering it is necessary to have a reliable protocol for the plant transformation and selection of putative transformed cells as well as a proved regeneration system (Laimer et al. 1992). The main pathways to regenerate whole plants from transformed cells are the organogenesis and the somatic embryogenesis. The regeneration of plants by organogenesis involves both, the induction and development of shoots as well as the subsequent root organogenesis controlled by phytohormones. If organs are regenerated directly from the explant it is referred to as adventitious organogenesis in contrast to indirect organogenesis with an initial callus formation with subsequent shoot or root formation. If shoots or roots regenerate from cell clusters it is possible to obtain chimeric plants. Somatic embryos are developed from single somatic cells and have a bipolar structure with shoot and root meristems. Similar to organogenesis, embryogenesis can occur directly or indirectly with an intermediate callus phase (Su 2005).

Aim of the Study

The goal of this study was to provide the molecular foundation and knowledge needed to accelerate the development of novel genotypes by the CRISPR/Cas9 gene editing tool. Six different genes were knocked out to alter metabolic pathways in order to improve economical important traits of *J. curcas*.

The pathways were identified, target genes were selected and initial gRNAs were designed at the PBU, BOKU, Vienna, AT. The gene knock-out constructs carrying up to six gRNAs for a

particular gene family were constructed at the Crop Genetic Transformation Laboratory at the Center for Applied Genetic Technologies on the University of Georgia, USA. The constructed plasmids were validated by diagnostic digest and sequencing.

Material and Methods

Gene Structure Analysis

All relevant sequences for the target genes were assembled from previous publications (Agarwal et al. 2012, Dani et al. 2010, Gu et al. 2012, Jiang et al. 2012, King et al. 2014, Li et al. 2008, Li et al. 2015, Nakano et al. 2012, Sato et al. 2011, Wei et al. 2012, Wu et al. 2009, Wu et al. 2013, Wu et al. 2015,Ye et al. 2009, Yu et al. 2015)and different databases (Kazusa and databases related to NCBI). Keyword-search and the Basic Local Alignment Search Tool (BLAST) were used to find similar sequences within the genus *Jatropha* in both databases. Additionally whole seed transcript sequences were provided by PBU. Two reference genomes were obtained from the Kazusa DNA research institute in Japan (Sato et al. 2011) and from the Chinese Academy of Sciences (Wu et al. 2015). The sequences were downloaded in FASTA format and archived in the bioinformatics software Genious (www.geneious.com) for further analysis. All transcript data from a particular gene copy were aligned by using the MUSCLE (Edgar 2004) or website tool to find conserved regions. Additional information of phylogenetic analysis was used to verify different gene copies or isoforms. Phylogenetic trees by using the Unweighted Pair Group Method with Arithmetic mean (UPGMA) method were created by Geneious. Subsequently the gene structure was determined and the occurrence and position of introns identified. For this purpose the collected sequences were aligned to reference genomic DNA. After extraction of exon sequences it was possible to find conserved sequences within the exons across different gene copies by using the MUSCLE alignment tool. The exon sequences were translated into protein sequences and used as queries for BLAST search in the protein database UniProt (www.uniprot.org/blast/) to discover the position of regulatory protein domains.

Guide RNA Design

Guide RNAs were designed by the CRISPR site finder tool of Geneious (http://assets.geneious.com/manual/9.0/GeneiousManualse72.html). This tool is able to predict the off-target score by searching in a provided off-target database using an algorithm based on the scoring system proposed by Hsuet al. (2013) which takes into account the number and position of mismatches as well as the number of off-target sites. Two whole genome sequences,"JatCur_1.0" and "JAT_r4.5" of *Jatropha curcas* L. from the Chinese Academy of Sciences and the KAZUSA DNA Research Institute respectively, were imported into Genious to enable scoring against an off-target database. For genes with multiple copies the software could not find appropriate gRNAs due to the assumed off-targets in the different genes within the gene family. For these genes the CRISPR site finder was performed without scoring against an off-target database and the off-target scoring was done manually by NCBI BLAST search. The on-target score was obtained from the CRISPR site finder which use an algorithm based on the method proposed by Doench et al. (2014). Guide RNAs were created by searching with the query $GN_{(19)}$ adjacent to a NGG PAM sequence. Alternative gRNAs with the sequence $GN_{(20)}$-NGG or $GN_{(18)}$-NGG were used. The guanine at the 5' end is required for transcription by the U6 promoter (Ran et al. 2013).Subsequently the possible gRNAs were evaluated regarding to performance (on-target activity), off-target activity, secondary structure and position. The on-target activity of the selected gRNAs was ranging between 0.248 and 0.826. Whenever it was possible gRNAs with an on-target score ≥ 0.3 were used. The predicted off-target score ranged between 99.99 % and 100 %. For manual off-target prediction the occurrence of a PAM sequence adjacent to the gRNA and the similarity especially in the seed region were taken into account. The selected gRNAs have at least 10 mismatches or a lacking PAM at potential off-target sites. Uncharacterized sequences predicted by automated computational analysis provided by NCBI were not taken into account. The secondary structure of gRNAs in combination with their scaffold was evaluated according to Liang et al. 2016. The folding was predicted by using the "RNA Fold" function from Geneious with the energy model according to Andronescu et al. (2007). For most target regions, a gRNA with the tetraloop and all three stem loops could be found (Figure 10).

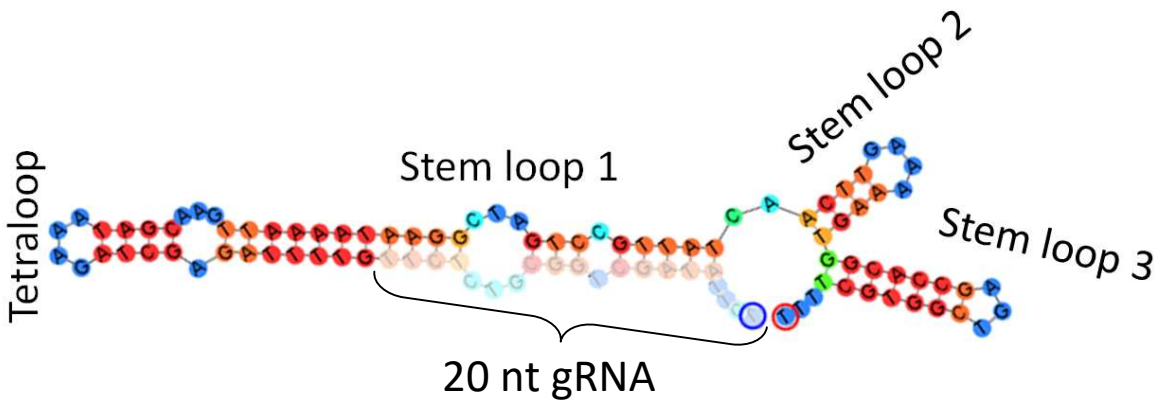

Figure 10. RNA folding of a 20 nt gRNA with the 83 bp gRNA scaffold sequence. The tetraloop and three different stem loops are required for efficient Cas9 cleavage.

Vector Construction

Up to 6 gRNA cassettes (MtU6 promoter + gRNA + Scaffold) were cloned into the linearized vector backbone by a two step NEBuilder assembly according to Jacobs et al. (2015).

gRNA cassettes construction

The gRNAs were ordered with a 20 nt MtU6 overhang sequence on the 5' end as well as a 20 nt scaffold overhang sequence on the 3' end (5'-AAGCGAACCAGTAGGCTTGNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATA-3') as DNA oligonucleotides from SIGMA-ALDRICH® Life Science.

The pUC gRNA shuttle (Addgene plasmid 47024) was used as template for the amplification of the *Medicago truncatula* U6.6 (MtU6.6) promoter which drives the gRNA as well as for the amplification of the gRNA scaffold sequence (Jacobs et al. 2015).The pUC gRNA shuttle plasmid containing *E. coli* strainwas cultivated overnight on a shaker (250rpm) at 37 °C on 3 ml LB media with100 µg ml^{-1}ampillicin. Plasmids were extracted and purified by the Plasmid Miniprep Kit according to the manufacturer's protocol (GenElute™). A spectrophotometer was used to determine the quantity. A working solution in the concentration of 1 ng plasmid DNA per µl was produced. The PCR (95°C for 3 min, 25 cycles [98°C for 20 sec, 62 °C for 20 sec, 72 °C for 1 min] and 72 °C for 1 min) was carried out by using the KAPA HiFi hotstart polymerase (KAPA HiFi)

from Kapa Biosystems, Inc. and the flanking primers MtU6.6F (ATGCCTATCTTATATGATCAATGAG), MtU6.6R (AAGCCTACTGGTTCGCTTGAAG) andScaffoldF (GTTTTAGAGCTAGAAATAGCA), ScaffoldR (AAAAAAAGCACCGACTCGGTG) respectively.For validation of amplification the PCR products were run on a 1% agarose gel (Figure 11)and visualized by a UV transilluminator.

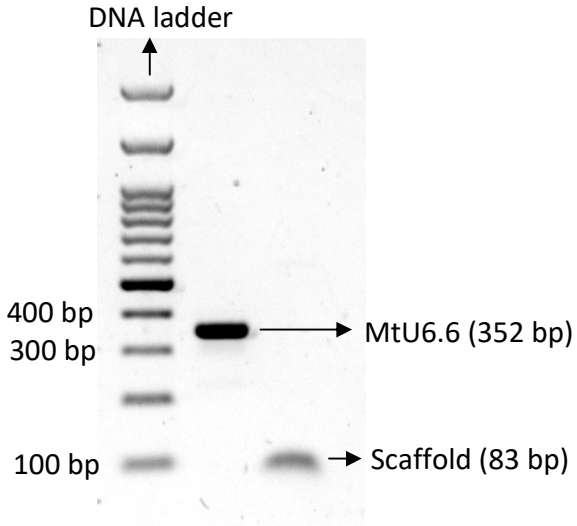

Figure 11.PCR product verification by gel electrophoresis (1 % TBE gel, 120 V, 30 min).
100 kb DNA ladder, MtU6.6 PCR product, gRNA scaffold PCR product.

After purification with the ZYMO DNA Clean & Concentrator Kit the amplicons were diluted to a 0,2 pmol working solution according to the formula *mass of dsDNA (g) = moles of dsDNA (mol) x ((length of dsDNA (bp) x 617.96 g/mol) + 36.04 g/mol)* provide by the NEBuilder Assembly manual (New England BioLabs GmbH).

In a subsequent NEBuilder assembly reaction the gRNA cassettes were assembled. The MtU6.6 promoter, the gRNA and the scaffold (0.2 pmol each) were incubated with the NEBuilder Assembly master mix from New England BioLabs® Inc. (NEB) for 60 min at 50 °C (Figure 12).

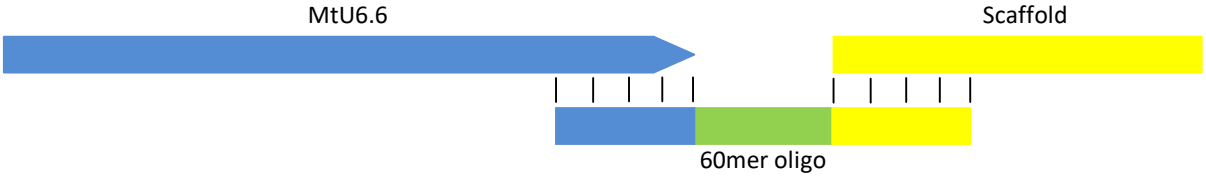

Figure 12. Principle of gRNA cassette assembly. MtU6.6 promoter DNA (blue arrow);60 mer gRNA oligo containing$G(N)_{19}$ nt target sequence (green bar), 20 nt MtU6 overhang (blue bar) and 20 nt scaffold overhang (small yellow bar); gRNA Scaffold DNA (big yellow bar). The 60mer oligo overlaps 20 bp with the adjacent pieces. The NEBuilder Assembly reaction combines all three pieces to a gRNA cassette.

The unpurified NEBuilder products were used as templates for a PCR (95°C for 3 min, 25 cycles [98°C for 20 sec, 62 °C for 20 sec, 72 °C for 1 min] and 72 °C for 1 min)with KAPA HiFi and the flanking primers containing individual overhangs for each cassette. The first cassette starts with a SwaI overhang on the 5' end and has a unique sequence (UNS) on the 3' end. The last cassette has a UNS on the 5' end and ends with a SpeI overhang on the 3' end (Figure 13).

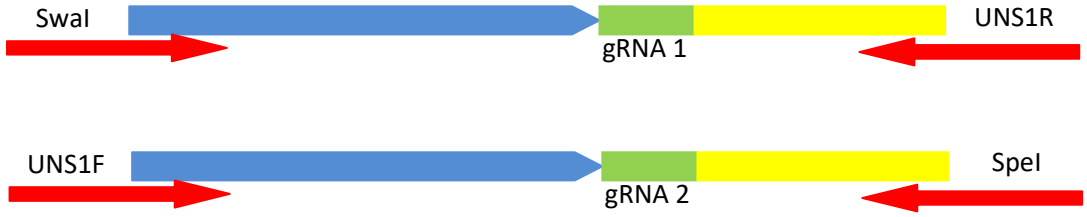

Figure 13. Principle of creating individual overhangs for two gRNA cassettes by PCR. The red arrows indicates the used primers. The first cassette starts with the SwaI overhang on the 3' end and has an UNS1 overhang on the 5' end. The second cassette has an UNS1 overhang on the 3' end and a SpeI overhang on the 5' end.

Amplifications of cassettes were validated by gel electrophoresis.

Vector assembly

The 201N:Cas9 (Addgene plasmid 59175) was used as vector backbone (Figure 14). It contains a *Streptococcus pyogenes*Cas9 (SpCas) gene driven by double 35S promoter from califlower mosaic virus (CaMV35S) and terminated by a nopaline synthase (NOS) terminator from

Agrobacterium tumefaciens. As plant selectable marker (kanamycin resistance) it contains a neomycin phosphotransferase II (*npt*II) gene under the control of the Ubi3 promoter and terminator derived from *Solanum tuberosum*. The bacterial selection on kanamycin is enabled by an aminoglycoside 3-N-acetyltransferase (aph) gene (Jacobs et al. 2015).

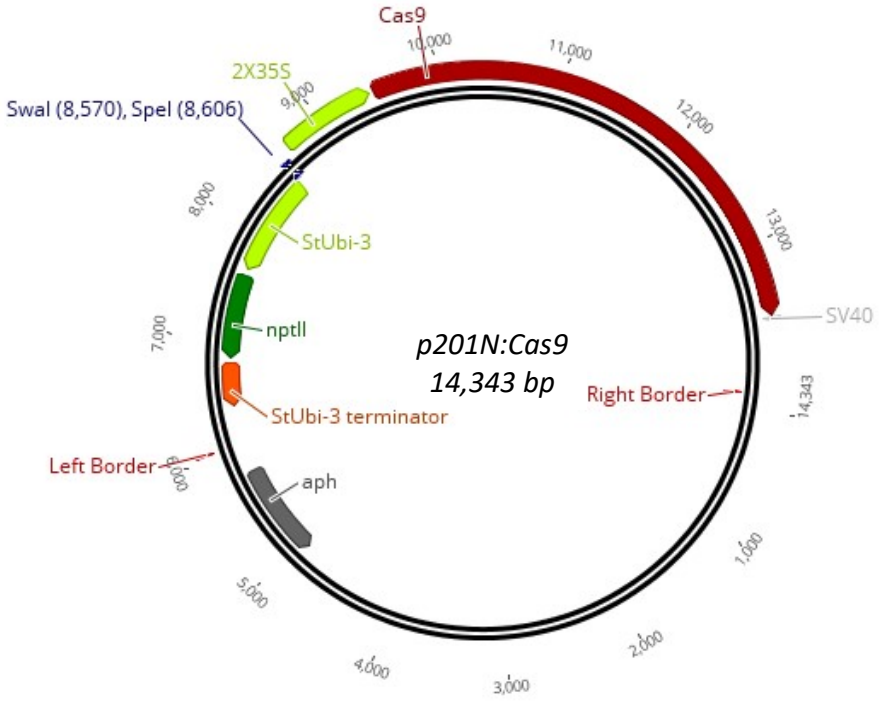

Figure 14. The p201N:Cas9 vector backbone. The Cas9 gene is driven by the 2X CaMV35S promoter and the StUbi3:nptII cassette for plant selection are within the left and right border. The aph gene is responsible for bacterial selection.

The vector backbone was linearized by a two step restriction digestion. The first restriction digest was performed with the SwaI restriction enzyme from NEB in a 10X NEBuffer 3.1 buffer solution for 3 h at 25 °C. The digestion product and an undigested control plasmid were run on a 1% agarose gel to validate digestion. The digestion product was purified with the DNA Clean &Concentrator kit from Zymo Research Corporation according to the manufacturer's protocol and quantified by spectrophotometry. Subsequently the final digestion was performed with the SpeI restriction enzyme from NEB in a 10X NEB-CutSmart buffer for 2 h at 37 °C. In a second NEBuilder reaction the linearized vector backbone and the inserts (gRNA cassettes) were

ligated. For each reaction 100 ng (~ 0.011 pmol) linearized plasmid DNA (unpurified restriction-digest product) were used. For up to three gRNA cassettes 0.2 pmol each were used. For more than 3 inserts 0.1 pmol were used for each gRNA cassette, to reduce total amount of fragments (NEBuilder manual recommends a maximum of 0.5 pmol). The mixture was incubated for 1 h at 50 °C.

Results

Vector Construction

The completeness of all vectors was verified by diagnostic digestions with the restriction enzymes AscI and NcoI. Two fragments with invariable size (approximately 4750 bp and 7950 bp) are consisting only of the vector backbone sequence. The third fragment contains the gRNA cassettes and vary in size from 2072 bp (one gRNA cassette) to 4447 bp (six gRNA cassettes), which is shown in figure 15.

gRNAs	size (bp)
0	1617
1	2072
2	2547
3	3022
4	3497
5	3972
6	4447

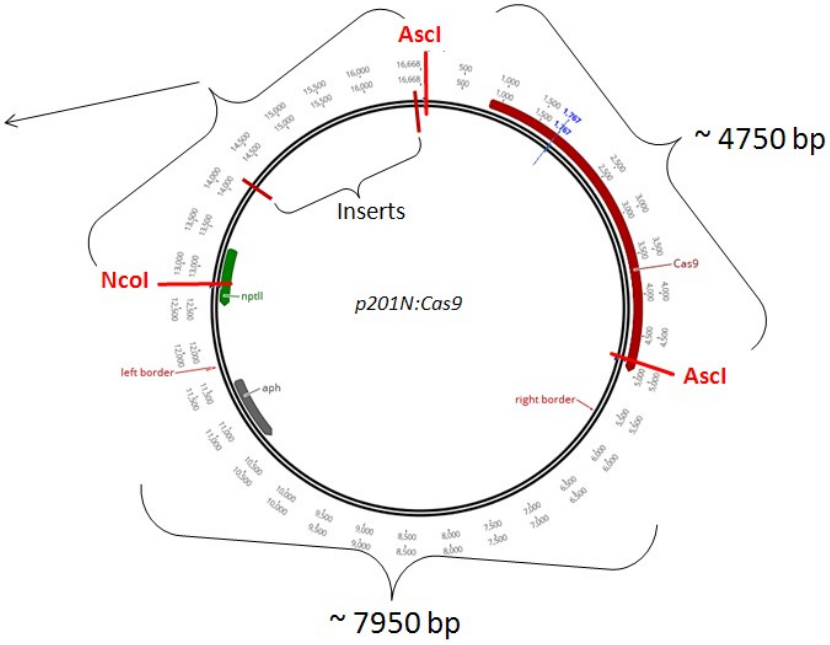

Figure 15. The diagnostic digest of the p201N:Cas9 vector containing the specific gRNA cassettes result in a fragmentation to three linearized sequences. The size of one fragment variesaccording to the amount of gRNA cassettes and can be validated by gel electrophoresis.

All vector fragments separated on an agarose gel showed the expected sizes (Figure 16). The vector backbone fragment between the two AscI restrictions sites (4750 bp) and the fragment containing the six gRNA cassettes from the vector Allergenicity 2 (4447 bp) are very close together and can be distinguished only with difficulty.

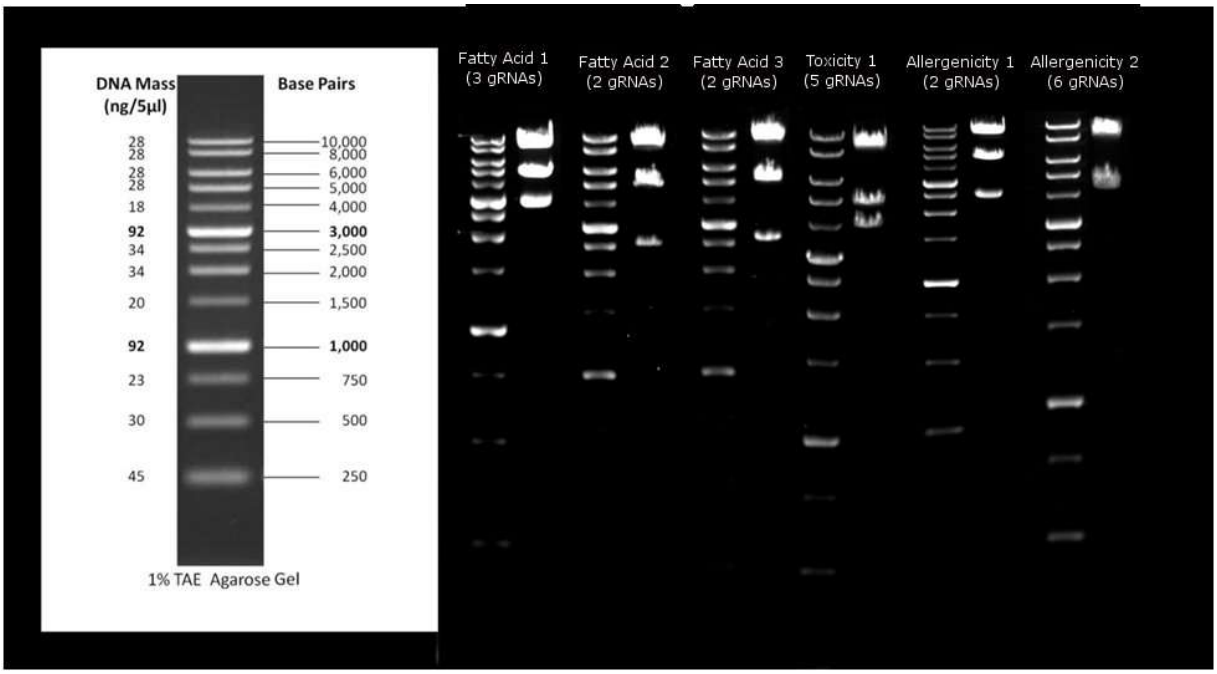

Figure 16. Diagnostic digest of gene knock-out vectors.

The sequencing results (57 single reads) confirmed an error-free sequence of all gRNAs. Two types of INDEL mutations were accepted: one additional or one lacking thymine base in the terminator region of gRNA scaffold sequences (Figure17).

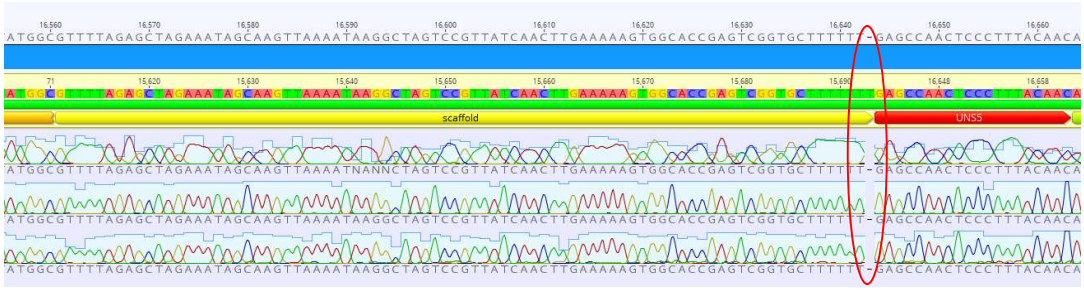

Figure 17. Sequencing indicates the lack of a thymine base of the terminator sequence of the gRNA scaffold.

Additionally a single nucleotide polymorphism (SNP) was found in each MtU6 promoter sequence. At the nucleotide position -45 the promoter multiplied from the pUC gRNA shuttle has a thymine base instead of a guanine. According to Kim and Nam (2013) the SNP is not in a functional motif sequence (Figure 18).

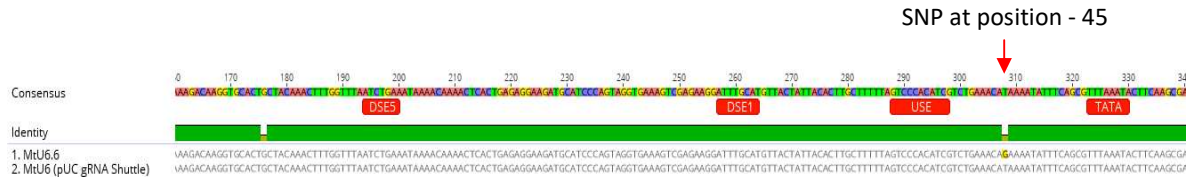

Figure 18. MtU6.6 promoter. Red boxes show conserved DNA elements according to Kim and Nam (2013).

Discussion

The main topic of this report is the precise editing of commercially important genes of *J. curcas*. The CRISPR/Cas system seems to be the perfect tool to fine tune plant metabolic pathways. It is able to knock-out specific gene families simultaneously with relatively simple designed plasmids. However trans-genes like the *Cas9* gene and plant selectable marker are stable integrated into the plant genome. The removal of transgenes will be an issue for further studies.

The present report contributes to make *Jatropha curcas* a competitive biofuel crop. Particularly for developing countries locally produced biofuels have a huge potential to improve the country`s energy security. Additionally the promoting of *Jatropha curcas* plantation will increase employment opportunities in the agro-sector. Beside the further improvement of oil composition and reduction of toxicity and allergenicity many other traits have to be taken into account to improve *Jatropha* as biofuel crop. The developed molecular tool can be used to eliminate further deficits of *Jatropha* and to obtain more knowledge about genetic basis of important qualitative and quantitative traits.

Conclusion and Perspectives

Future efforts will involve the transformation of the constructed CRISPR/Cas9 gene knock out plasmids into *E. coli* (in order to multiply the plasmids for further use and to be able to create *E. coli* freezer stocks for long term storage). After multiplication and purification the plasmids can be transferred to *Agrobacterium tumefaciens* for transformation of *J. curcas* leaf discs following a protocol established at PBU. Putative transformed cells will be selected on media containing

selection agents. Sequencing will enable the detection of either insertions or deletions induced by the CRISPR/Cas9 system in the genes of interest. Whole plants will be regenerated from transformed plant cells and the successful gene knockout evaluated by phenotypic analyses.

This work provides the fundamental knowledge of gene structure to construct gene knock out constructs for *J. curcas* based on the CRISPR/Cas9 system. The designed gRNAs make it possible to fine-tune the oil composition and to reduce the amount of toxins, as well as the content of allergens. Furthermore, the established tool set can lead to a better understanding and improvement of commercially important traits of *J. curcas*.

References

Aalberse RC (2000) Structural biology of allergens. Journal of Allergy and Clinical Imunology 106(2), 228-238.

Abdullah BM, Yusop RM, Salimon J, Yousif E and Salih N (2013) Physicochemical Properties Analysis of *Jatropha curcas* Seed Oil for Industrial Applications. World Academy of Science, Engineering and Technology 7 (12), 893-896.

Adriano-Anaya Mde L, Pérez-Castillo E, Salvador-Figueroa M, Ruiz-González S, Vázquez-Ovando A, Grajales-Conesa J, Ovando-Medina I (2016) Sex expression and floral diversity in *Jatropha curcas*: a population study in its center of origin. PeerJ. 4:e2071.

Ahmad S and Sultan SM (2015) Physiological Changes in the Seeds of *Jatropha curcas* L. at Different Stages of Fruit Maturity. Brazilian Archives of Biology and Technology 58 (1), 118-123.

Andres C, Bargsten K, Jinek M (2016) Structural Plasticity of PAM Recognitionby Engineered Variantsof the RNA-Guided Endonuclease Cas9. Mol Cell 61 (6), 895-902.

Barros TFS, Arriel NHC, Queiroz MF, Fernandes PD, Mendonca S, Ribeiro JAA and Medeiros EP (2015) Fatty acid profiles of species of*Jatropha curcas* L., *Jatropha mollissima*(Pohl) Baill. and *Jatropha gossypiifolia* L. Industrial Crops and Products 73, 106–108.

Baumert S, Khamazina A and Vlek P (2016) Soil organic carbon sequestration in *Jatropha curcas* systems in Burkina Faso. Land Degrad. Develop. 27, 1813-1819.

Belhaj K, Chaparro-Garcia A, Kamoun S, Patron NJ and Nekrasov V. (2015) Editing plant genomes with CRISPR/Cas9. Current Opinion in Biotechnology 32, 76-84.

Carroll D (2011) Genome engineering with zinc-finger nucleases. Genetics188(4), 773-82.

Chang W, Song H, Liu H and Liu P (2013) Current Development in Isoprenoid Precursor Biosynthesis and Regulation. Curr Opin Chem Biol. 17 (4), 571-579.

Contran N, Chessa L, Lubino M, Bellavite D, Roggero PP, Enne G (2013) State-of-the-art of the *Jatropha curcas* productive chain: From sowing to biodiesel and by-products. Inustrial Crops and Producrs 42, 202-215

Dahmer N, Wittmann MTS and Dias LAS (2009) Chromosome numbers of *Jatropha curcas* L.: an important agrofuel plant. Crop Breeding and Applied Biotechnology 9, 386-389.

Dani KG, Hatti KS, Ravikumar P, Kush A (2010) Structural and functional analyses of a saturated acyl ACP thioesterase, type B from immature seed tissue of *Jatropha curcas*. Plant Biol 13(3),453-61.

Deus de Oliveira N and Lima Tavares Machado O (2011) Allergens and Toxins from Oleaginous Plants: Problems and Solutions. In: Dos Santos Bernardes MA (Ed) Environmental Impact of Biofuels. Chapter 3, pp 41-66. InTech. Rijeka, Croatia. September 2011.

Edgar RC (2004) MUSCLE: multiple sequence alignment with highaccuracy and high throughput. Nucleic Acids Research32 (5), 1792-1797.

Escobar MA and Dandekar AM (2003) Agrobacterium tumefaciens as an agent of disease. Trends in Plant Science 8 (8), 380-386.

Fernandes AMAP, El-Khatib S, Cunha IBS, Porcari AM, Eberlin MN, Silva MJ, Silva PR, Cunha VS, Daroda RJ and Alberici RM (2015) Chemical Characterization of *Jatropha curcas* L. Seed Oil and its Biodiesel by Ambient Desorption/Ionization Mass Spectrometry. *Energy Fuels*29 (5), 3096–3103.

Forster BP and Shu QY (2011) Plant Mutagenesis in Crop Improvement: Basic Terms and Applications. In: Plant Mutation Breeding and Biotechnology. Shu QY, Forster BP and Nakagawa H (Eds). CABI, Rome, 9-20.

Fu Y, Sander JD, Reyon, Cascio VM, Joung JK (2014) Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat Biotechnol. 32(3), 279-284.

Gaspar T, Kevers C, Penel C, Greppin H, Reid DM and Thorpe TA (1996) Plant hormones and plantgrowth regulators in plant tissue culture. In Vitro Cellular & Developmental Biology – Plant 32 (4), 272-289.

Georges F and Ray H (2017) Genome editing of crops: A renewed opportunityfor food security. GM Crops & Foods 8 (1), 1-12.

Glevin SB (2003) Agrobacterium-Mediated Plant Transformation: the Biology behind the "Gene-Jockeying" Tool. Microbiology and Molecular Biology Reviews 67 (1), 16-37.

Gogoi R, Niyogi UK and Tyagi AK (2015) Methods of detoxification of *Jatropha curcas* L seed cake for its use as protein supplement in animal feed - An overview Indian. Journal of Natural Products and Resources6 (3), 176-182.

Gu K, Yi C, Tian D, Sangha JS, Hong Y, and Yin Z. (2012) Expression of fatty acid and lipid biosynthetic genes in developing endosperm of *Jatropha curcas*. Biotechnology Biofuels 5 (1), 47.

Guha TK, Wai A and Hausner G (2017) Programmable Genome Editing Tools and their Regulation for Efficient Genome Engineering. Computational and Structural Biotechnology Journal 15, 146–160.

Guilinger JP, Pattanayak V, Reyon D, Tsai SQ, Sander JD, Joung JK and Liu DR (2014) Broad Specificity Profiling of TALENs Results in EngineeredNucleases With Improved DNA Cleavage Specificity. Nat Methods11(4), 429–435.

Guo-Liang J (2015) Molecular Marker-Assisted Breeding: A Plant Breeder's Review. In: Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools. Al-Khayri JM, Jain SM and Johnson DV (Eds.). Springer, New York, 431-472.

Haft DH, Selengut J, Mongodin EF, Nelson KE (2005) A Guild of 45 CRISPR-Associated (Cas)Protein Families and Multiple CRISPR/CasSubtypes Exist in Prokaryotic Genomes. PLoS Comput Biol. 1(6):e60.

Haberlandt G (1902) Culturversuche mit isolierten Pflanzenzellen. Sitz Akad. Wiss. Wien 111, 69–92.

Hansen and G and Wright MS (1999) Recent advances in the transformation of plants. Trends in Plant Science 4 (6), 226-231.

Heller J (1996) Physic nut. *Jatropha curcas* L. Promoting the conservation and use of underutilized and neglected crops 1. Institute of Plant Genetics and Crop Plant Research, Rome 1996.

Hoekema A, Hirsch PR, Hooykaas PJJ and Schilperoort RA (1983) A binary plant vector strategy based on separation of vir- and T-region of the Agrobacterium tumefaciens Ti-plasmid. Nature 303, 179-180.

Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y,Fine EJ, Wu X, Shalem O, Cradick TJ, Marraffini LA, Bao G and Zhang F (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. Nature Biotechnology 31, 827–832.

Hsu PD, Lander ES, Zhang F (2014) Development and Applications of CRISPR-Cas9forGenomeEngineering. Cell. 157(6), 1262-78.

Jiang H, Wu P, Zhang S, Song C, Chen Y, Li M, Jia Y, Fang X, Chen F, and Wu G (2012) Global analysis of gene expression profiles in developing physic nut (*Jatropha curcas* L.) seeds. PLoS One 7 (5).

Jain S and Kumar A (2015) The Pathogenesis Related Class 10 proteins in PlantDefense against Biotic and Abiotic Stresses. Advances in Plants & Agriculture Research 3 (1).

Jain S and Sharma MP (2009) Prospects of biodiesel from *Jatropha* in India: A review. Renewable and Sustainable Energy Reviews 14, 763–771.

Joung JK and Sander JD (2013) TALENs: a widely applicable technology for targeted genome editing. Nat Rev Mol Cell Biol. 14(1), 49–55.

Juan JC, Kartika DA, Wu TY and Hin TY (2011) Biodiesel production from *jatropha* oil by catalytic and non-catalytic approaches: An overview. Bioresource Technology 102, 452-460.

Kim GB and Nam YW (2013) Isolation and Characterization of *Medicago truncatula* U6 Promoters for the Construction of Small Hairpin RNA-Mediated Gene Silencing Vectors. Plant Molecular Biology Reporter 31, 581-593.

Key S, Ma JKC and Drake PMW (2008) Genetically modified plants and human health. Journal of the Royal Socieety of Medicine 101 (6), 290-298.

Kikkert JR, Vidal JR and Reisch BI (2006) Stable Transformation of Plant Cellsby Particle Bombardment/Biolistics. In: Methods in molecular biology, volume 286. Transgenic plants. Methods and protocols. PeñaP (ed). Humana Press, New Jersey, 61-78.

King A, Brown G, Gilday A, Larson T and Graham I (2014) Production of Bioactive Diterpenoids in the Euphorbiaceae Depends on Evolutionarily Conserved Gene Clusters.Plant Cell 26 (8), 3286-3298.

Knothe G and Steidley KR (2005) Kinematic viscosity of biodiesel fuel components and related compounds.Influence of compound structure and comparison topetrodiesel fuel components. Fuel 84, 1059-1065.

Knothe G (2008) "Designer" Biodiesel: Optimizing Fatty Ester Composition toImprove Fuel Properties. Energy & Fuels 22, 1358-1364.

Koh MY and Ghazi TIM (2011) A review of biodiesel production from *Jatropha curcas* L. Oil. Renewable and Suitable Energy Reviews 15, 2240-2251.

Koonin EV, Makarova KS and Zhang F (2017) Diversity, classification and evolution of CRISPR-Cas systems. Curr Opin Microbiol. 37, 67-78.

Kuckuck H, Kobade G, Wenzel G (1991) Fundamentals of Plant Breeding. Springer, New York, 3-80.

Kumar U (2003) Methods in Plant Tissue Culture. 2nd Edition. Agrobios, India.

Laimer M and Rücker W (2003) Plant Tissue Culture: 100 years since Gottlieb Haberlandt. Springer, Wien.

Laimer M, da Camara Machado A, Hanzer V, Weiß H, Regner F, Steinkellner H, Mattanovich D, Plail R, Knapp E, Kalthoff B and Katinger H (1992) Regnereration of transgenic plants of Prunus armeniaca containing the coat protein gene of Plum Pox Virus. Plant Cell Reports 11 (1), 25-29.

Li J, Li MR, Wu PZ, Tian CE, Jiang HW and Wu GJ (2008) Molecular cloning and expression analysis of KASIII from *J. curcas*. Tree Physiology 28, 921-927.

Li C, Ng A, Xie L, Mao H, Qiu C, Srinivasan R, Yin Z and Hong Y (2015) Engineering low phorbol ester *Jatropha curcas* seed by intercepting casbene biosynthesis. Plant Cell Rep. 35(1), 103-114.

Liang G, Zhang H, Lou D and Yu D (2015) Selection of highly efficient sgRNAs for CRISPR/Cas9-based plant genome editing. Scientific Reports 6, 21451.

Liu JJ and Ekramoddoullah AKM (2006) The family 10 of plant pathogenesis-related proteins: Their structure, regulation, and function in response to biotic and abiotic stresses. Physiological and Molecular Plant Pathology 68, 3-13.

Lorence A, Verpoorte R (2004) Gene transfer and expression in plants. In: Recombinant Gene Expression.Balbás P and Lorence A (eds).Methods Mol Biol. 267, 329-350.

Lotan T, Ori N and Fluhr R(1989) Pathogenesis-Related Proteins Are Developmentally Regulated in Tobacco Flowers. The Plant Cell 1, 881-887.

Maes WH, Achten WMJ, Reubens B, Raes D, Samson R and Muys B (2009) Plant–water relationships and growth strategies of *Jatropha curcas* L. seedlings under different levels of drought stress. Journal of Arid Environments 73, 877-884.

Maghuly F and Laimer M (2013) *Jatropha curcas*, a biofuel crop: functional genomics for understanding metabolic pathways and genetic improvement. Biotechnol. J. 8, 1172–1182.

Maghuly F, Marzban G and Laimer M (2009) Functional Genomics of Allergen Gene Families in Fruits. Nutrients 1 (2), 119-132.

Maghuly F, Vollmann J and Laimer M (2015): Biotechnology of Euphorbiaceae (*Jatropha curcas*, *Manihot esculenta, Ricinus communis*). Applied Plant Genomics and Biotechnology 6, 87-114.

Maghuly F, Jankowicz-Cieslak J, Pabinger S, Till BJ and Laimer M (2015) Geographic origin is not supported by the genetic variability found in a large living collection of *Jatropha curcas* with accessions from three continents. Biotechnology Journal 10, 536–551.

Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, van der Oost J, Koonin EV (2011) Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol. 9(6), 467-77.

Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou R, Brouns SJ, Charpentier E, Haft DH, Horvath P, Moineau S, Mojica FJ, Terns RM, Terns MP, White MF, Yakunin AF, Garrett RA, van der Oost J, Backofen R, Koonin EV (2015) An updated evolutionary classification of CRISPR–Cas systems. Nat Rev Microbiol. 13(11), 722-36.

Makkar and Becker (2009) *Jatropha curcas*, a promising crop for the generationof biodiesel and value-added coproducts. Eur. J. Lipid Sci. Technol. 111, 773–787.

Mali P, Esvelt KM, Church GM (2013)Cas9 as a versatile tool for engineering biology. Nat Methods. 10(10), 957-963.

Montes JM, Technow F, Bohlinger B and Becker K (2013) Seed quality diversity, trait associations and grouping of accessions in *Jatropha curcas* L. Industrial Crops and Products 51, 178-185.

Montes J and Melchinger A (2016) Domestication and Breeding of Jatropha curcas L. Trends in Plant Science 21 (12), 1045-1057.

Moreno FJ and Clemente A (2008) 2S Albumin Storage Proteins: What Makes them Food Allergens? The Open Biochemistry Journal 2, 16-28.

Nakano Y, Ohtani M, Polsri W, Usami T; Sambongi K and Demura T (2012) Characterization of the casbene synthase homolog from Jatropha (*Jatropha curcas* L.). Plant Biotechnology 29, 185-189.

Nester EW (2015) *Agrobacterium*: nature`s genetic engineer. Frontiers in Plant Science 5:730.

Pacurar DI, Thordal-Christensen H, Pacurar ML, Pamfil D, Botez C and Bellini C (2011) *Agrobacterium tumefaciens*: From crown gall tumors to genetic transformation. Physiological and Molecular Plant Pathology 76, 76-81.

Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA, Liu DR (2013) High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat Biotechnol. 31(9), 839-43.

Pecina-Quintero V, Anaya-L ópez, Zamarripa-Colmenero A, Núñez-Colín CA, Montes-García N, Solís-Bonilla JL and Jiménez-Becerril MF (2014) Genetic structure of *Jatropha curcas* L. in Mexico and probable centre of origin. Biomass and Bioenergy 60, 147-155.

Peng R, Lin G and Li J (2015) Potential pitfalls of CRISPR/Cas9-mediated genome editing. FEBS J. 283(7), 1218-1231.

Peterson BA, Haak DC, Nishimura MT, Teixeira PJ, James SR, Dangl JL, Nimchuk ZL (2016) Genome-Wide Assessment of Efficiency and Specificity in CRISPR/Cas9 Mediated Multiple Site Targeting in Arabidopsis. PLoS One. 11(9):e0162169.

Pühringer HM, Zinöcker I, Marzban G, Katinger H and Laimer M (2003) MdAP, a novel protein in apple, is associated with the majorallergen Mal d 1. Gene 321, 173–183.

Qi X, Zhang J, Zhao Y, Chen T, Xiang Y, Hui J, Cai D, Liu Y, Xia L, Yu T, Li G (2016) The applications of CRISPR screen in functional genomics. Briefings in Functional Genomics16(1), 34-37.

Qu J, Mao HZ, Chen W, Gao SQ, Bai YN, Sun YW, Geng YF, Ye J (2012) Development of marker-free transgenic Jatropha plants with increased levels of seed oleic acid. Biotechnol. Biofuels 5 (1), 10.

Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, Trevino AE, Scott DA, Inoue A, Matoba S, Zhang Y and Zhang F (2013) Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. Cell. 154(6), 1380-1389.

Ran FA, Hsu P, Wright J, Agarwala V, Scott DA and Zhang F (2013) Genome engineering using the CRISPR-Cas9 system. Nat Protoc. 8(11), 2281–2308.

Reubens B., Achten W.M.J., Maes W.H., Danjon F., Aerts R., Poesen J., Muys B. (2011): More than biofuel? Jatropha curcas root system symmetry and potential for soil erosion control. Journal of Arid Environments 75 (2), 201-205.

Saad AIM and Elshahed (2012) Plant Tissue Culture Media. In: Recent Advances in Plant in vitro Culture. Leva A and Rinaldi LMR (eds). InTech. DOI: 10.5772/52760

Salisbury FB and Ross CW (1992) Plant Physiology. 4thEdition. Wadsworth Publishing Company, Belmont California 1992.

Sander JD and Joung JK (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol. 32(4), 347-55.

Sato S, Hirakawa H, Isobe S, Fukai E, Watanabe A, Kato M, Kawashima K, Minami C, Muraki A, Nakazaki N, Takahashi C, Nakayama S, Kishida Y, Kohara M, Yamada M, Tsuruoka H, Sasamoto S, Tabata S, Aizu T, Toyoda A, Shin-i T, Minakuchi Y, Kohara Y, Fujiyama A, Tsuchimoto S, Kajiyama S, Makigano E, Ohmido N, Shibagaki N, Cartagena JA, Wada N, Kohinata T, Atefeh A, Yuasa S, Matsunaga S and Fukui K (2011) Sequence analysis of the genome of an oil-bearing tree, *Jatropha curcas* L. DNA Res 18 (1), 65-76.

Schaefer KA, Wu WH, Colgan DF, Tsang SH, Bassuk AG and Mahajan VB.(2017) Unexpected mutations after CRISPR–Cas9 editing in vivo. Nat Methods. 14(6), 547-548.

Schopfer P, Brennicke A (2010) Pflanzenphysiologie. 7. Auflage. Spektrum Akademischer Verlag Heidelberg 2010. 357pp.

Silip JJ, Tambunan AH, Hambali H, Sutrisno, Surahman M (2010) Lifecycle Duration and Maturity Heterogeneity of *Jatropha curcas* Linn. J Sust Dev. 3(2), 291-295.
Stintzi A, Heitz T, Prasad V, Wiedemann-Merdinoglu S, Kaufmann S, Geoffroy P, Legrand M and Fritig B (1993) Plant 'pathogenesis-related' proteins and their role in defense against pathogens. Biochimie 75, 687-706.

Singh B and Sharma RA (2015) Plant terpenes: defense responses, phylogenetic analysis, regulation and clinical applications. Biotech 5, 129-151.

Su WW (2005) Cell Culture and Regeneration of Plant Tissues. In: Transgenic Plants and Crops. Khachatourians GG, McHughen A, Scorza R and Nip WK (eds). Taylor & Francis e-Library. 175-196.

Sugisaki H and Kanazawa S (1981) New restriction endonucleases from *Flavobacteriumokeanokoites* (FokI) and *Micrococcus luteus* (MluI). Gene. 16(1-3), 73-78.

Tanya P, Dachapak S, Tar MM and Srinives P (2011) New microsatellite markers classifying nontoxic and toxic *Jatropha curcas*. J Genet. 90 (3),76-78.

Thorpe TA (1994) Morphogenesis and Regeneration. In: Plant Cell and Tissue Culture. Vasil IK and Thrope TA (eds). Springer Science. 17-36.

Tsai SQ, Wyvekens N, Khayter C, Foden JA, Thapar V, Reyon D, Goodwin MJ, Aryee MJ and Joung JK (2014) Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. Nat Biotechnol. 32(6), 569-76.

Ugbogu AE, Akubugwo EI, Uhegbu FO, Chinyere CG, Ugbogu OC and Oduse KA (2014) Quality assessment profile of *Jatropha curcas*(L) seed oil from Nigeria. International Food Research Journal 21(2), 735-741.

van Loon LC and van Kammen A (1970) Polyacrylamide disc electrophoresis of the soluble leaf proteins from Nicotiana tabacum var. "Samsun" and "Samsun NN". II. Changes in protein constitution after infection with tobacco mosaic virus. Virology. 40(2), 190-211.

Wani SP, Grarg KK and Chander G (2016) Water needs and productivity of Jatropha curcas in India: myths and facts. Biofuels, Bioprod. Bioref. 10, 240-254.

Waugh R and Brown JWS (1991) Plant gene structure and expression. In: Grierson D (Ed) Plant Genetic Engineering. Volume 1. Chapter 1. pp 1-31. Blackie, London 1991.

Wei Q, Li J, Zhang L, Wu P, Chen Y, Li M, Jiang H and Wu G. (2012) Cloning and characterization of a beta-ketoacyl-acyl carrier protein synthase II from *Jatropha curcas.* Journal of Plant Physiology 169, 816– 824.

Wise AA, Liu Z and Binns AN (2006) Three Methods for the Introduction of Foreign DNA into Agrobacterium. Methods Mol Biol. 343, 43-53.

Wu PZ, Li J, Wei Q, Zeng L, Chen YP, Li MR, Jiang HW and Wu GJ (2009) Cloning and functional characterization of an acyl-acyl carrier protein thioesterase (JcFATB1) from *Jatropha curcas*. Tree Physiology 29 (10), 1299-1305.

Wu P, Zhang S, Zhang L, Chen Y, Li M, Jiang H, Wu G (2013) Functional characterization Of two microsomal fatty acid desaturases from *Jatropha curcas* L. Journal of Plant Physiology 170, 1360–1366.

Yaakob Z, Narayanan BN, Padikkaparambil S, Unni S and Akbur M (2014) A review on the oxidation stability of biodiesel. Renewable and Sustainable Energy Reviews 35, 136-153.

Ye J, Hong Y, Qu J and Wang C (2013) Improvement of J. Curcas oil by genetic transformation. In: Bahadur B, Sujatha M, Carels N (Eds), Jatropha, Challenges for a New Energy Crop, Volume 2. Springer, New York 2013, pp 547-562.

Ye J, Qu J, Bui HT and Chua NH (2009) Rapid analysis of *Jatropha curcas* gene functions by virus-induced gene silencing. Plant Biotechnology Journal 7 (9), 964-976.

Yu N, Xiao WF ,Zhu J, Chen XY and Peng CC (2015) The *Jatropha curcas* KASIII gene alters fatty acid composition of seeds in *Arabidopsis thaliana*. Biologia Plantarum 59 (4), 773-782.

Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, van der Oost J, Regev A, Koonin EV and Zhang F (2015) Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. Cell. 163(3), 759-771.

Zhang XH, Tee LY, Wang XG, Huang QS, Yang SH (2015) Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. Mol Ther Nucleic Acids. 4, e264.