Plant Archives Vol. 21, Supplement 1, 2021 pp. 202-209 e-ISSN:2581-6063 (online), ISSN:0972-5210

Plant Archives

Journal homepage: http://www.plantarchives.org
doi link : https://doi.org/10.51470/PLANTARCHIVES.2021.v21.S1.035

CRISPR-CAS9: A GENOME EDITING TOOL FOR IMPROVEMENT OF BIOFUEL PRODUCTION IN DIATOMS: A REVIEW

Sarvjeet Kukreja[1], S.M. Savini Gunarathne[2], Tapas Giri[2], Umesh Goutam[2] and Shristy Gautam[2*]

[1]Department of Agronomy, Lovely Professional University, Phagwara, India, 144411.
[2]Department of Molecular Biology and Genetic Engineering, School of Bioengineering and Biosciences, Lovely Professional University, Phagwara, India, 144411.
*Corresponding author: Department of Molecular biology and Genetic Engineering, School of Bioengineering and Biosciences, Lovely Professional University, Punjab, India, 144411, Email: shristy.g@gmail.com

ABSTRACT

Clusters of Regularly Interspersed Short Palindromic Repeats-Cas 9 (CRISPR-Cas 9) protein is a simple but powerful tool that has been introduced for site specific gene editing so far. Crispr technology is a natural phenomenon that occurs in bacteria and archaea as a self-defense mechanism against the attacks by virus and other foreign particles. This simple phenomenon comes into the picture and plays a remarkable role in the improvement of the genome of various organisms, including diatoms by site specific genome editing so as to harness various products of human needs. Due to the increasing human population and the limited availability of fossil fuels and not forgetting to mention the urge to reduce the environmental pollution caused by the burning of fossil fuels, we are constantly looking for ecofriendly, cost friendly, renewable alternative sources of fuels. Diatoms as a great source of lipids play a major role as an alternative source of biofuels. Nuclease guided genome editing tools such as FOKI-I, TALENS, mega-nucleases, CRISPR associated Cas 9 make the improvement of the genome of diatoms to enhance the production of lipid content possible facilitating the production of biofuels. Among them CRISPR-Cas9 technologies are believed to have a great potential, inducing site specific genome editing more effectively and accurately than any other technique. Thus, this review, basically focuses on the importance of biofuels and how the diatoms can be utilized effectively manipulating their genome with the help of genome editing tools.

Keywords: Diatom, CRISPR-Cas 9, Biofuel, Triacylglycerol, gRNA

1. Introduction

Among the most striking issues the man has to deal with at the present moment. The depletion of the fossil fuels takes a prominent place. The simplest reason is the high consumption rate of the fossil fuels that cannot keep up with the rate of formation of the fossil fuels. This is why the production and development of alternative fuels that are sustainable, renewable and ecofriendly has become one of the necessities or the responsibility of the scientific community. The depletion and the shortage of fossil fuels is not the only reason why we are looking for sustainable and ecofriendly fuel sources, but also the environmental pollution caused by these fossil fuels is also one of the factors that drive us towards the other types of fuels (Demirbas & Demirbas, 2010, Voloshin *et al.*, 2016). The burning of fossil fuels releases large amounts of greenhouse gasses such as carbon dioxide and other gasses such as sulfur dioxide and other particles. These gasses ultimately contribute to the harmful effects such as global warming that would affect the lives of each and every living being on the planet. So, among the alternative fuels that have been introduced so far, biofuels made of bio materials are found to be the best option for this crisis of fuels. Biofuels also can be produced through different sources obtained from the environment. Among them biofuel production from diatoms is of interest of many researchers today due to many advantages that can be obtained over any other source (Hildebrand *et al.*, 2012, d'Ippolito *et al.*, 2015, Gautam *et al.*, 2016). The most important thing is that diatoms are a kind of algae that can grow any kind of water including sewage water without any effort and the most interesting property that is possessed by these diatoms is that they produce and store lipids inside the cells which can be use in the form of fuels. These are mostly called the oil droplets or the oil bodies of the diatoms. If this oil can be extracted and processed in large scale it can be the best solution for the fossil fuel crisis. Due to the reduced release of greenhouse gases the environmental pollution is also can be reduced to a great extent and it helps to mitigate the global climate change. The only barrier to the biofuel production from the diatoms is the enhancing the lipid accumulation in the diatom cells so as to match the demand of the humans. Because the diatoms produce only a little amount of lipids per cell that is required for their survival. The most successful solution for this is the genetic manipulation of the diatom genome so as to increase the lipid accumulation in the cell. Various studies have manipulated the genome of diatoms using various genomes editing techniques such as TALENS, RNAi, meganucleases, etc. (Sayanova & A Napier, 2016, Huang & Daboussi, 2017). Among them CRISPR-Cas 9 technology is believed have a great potential in inducing site-specific genome editions most effectively and accurately than any other technique. So, this study basically focuses on the importance of biofuels and

how the diatoms can be utilized effectively manipulating their genome with the help of genome editing tools.

1.1 The increasing thirst for the biofuels and the significance of diatoms in the biofuel production

The traditional fossil fuels derived from the decayed organic matter which takes thousands of years to form have been playing the most prominent role in the lives of humans in terms of source of energy as well as an important component of the world's economy. Even though these conditions prevail this moment we cannot guarantee that the situation will be same few more years later. This is due to many reasons, most of which have been identified as the irreversible harmful effects of the consumption of the fossil fuels. Due to the continually increasing population, no one can deny the fact that the day humans run out of fossil fuels is not that far because simply these fuels are being consumed at a rate that cannot match the rate of the formation of fossil fuels. So, the need of an alternative to fossil fuels is inevitable. In addition to that, the burning of fossil fuels makes a huge contribution to the environmental pollution. The poisonous gases released due to the combustion of fossil fuels, including greenhouse gases such as carbon dioxide results in the most serious environmental issues such as acid rains, global climatic change, etc. Undoubtedly global climate change can affect the lives of not only humans, but also the animals and plants badly. In other words, it can turn the lives of all the living beings on earth upside down if the prevailing condition worsens further. According to the statistics, transportation solely contributes to more than one fifth of the total greenhouse gas emissions in the United Kingdom. This has led the search for an ecofriendly, cost friendly, renewable and sustainable type of fuel become one of the top priorities of the researchers today. Biofuels derived from the crops or any other biological matter come into the picture as the best remedy for this issue man has to tackle with. They are basically renewable, sustainable, biodegradable, nontoxic sources of energy. Also, these biofuels help in mitigating the global climate change, reducing the release of greenhouse gases (Maity *et al.*, 2014). Even the economic wise the production of biofuels help in creating new job opportunities for the people and also reduce the dependence on the fossil fuels that being imported from the foreign countries mostly. Furthermore, the bio fuels also provide a good method to handle the organic waste that is being generated from our day to day life. Biofuels can be produced in various ways. They can be produced from the agricultural or organic waste, also they come as extractions from oil rich crops such as rapeseed, sunflower. Not only that, but also, biofuels can be produced from the fermentation of starch derived from the agricultural products such as corn, sugar cane, wheat etc. In addition to those sources' diatoms, a class of algae plays a significant role as an efficient and most economical source of biofuels than any other source.

Diatoms are an either unicellular or colonial significant group of organisms found in freshwater, oceans or any kind of water, including the sewage water all around the world (Gautam *et al.*). The ability to grow in a wide range of habitats without facing the problem of land limitation makes diatoms more significant and the reason for the current studies. There are approximately 200 diatom genera and more than 200,000 diatom species (Guiry, 2012), but still only 2 diatom genera have been completely sequenced (Armbrust *et al.*, 2004, Kaeriyama *et al.*, 2011). Based on the shape and symmetry diatoms are mainly 2 groups as pennales that move with the gliding motion and have pinnate markings and round non motile centrales with radial markings. The cell wall of diatoms is a silicified shell made of overlapping halves. This wall made of silica being transparent allows the entry of light and its perforated nature facilitates the diffusion and excretion of materials to outside. The cytoplasm of the diatoms contains single nucleus, mitochondria, golgi bodies and chloroplasts. The chloroplasts take different shapes like stellate, H-shaped, discoid etc. Photosynthetic pigments such as chlorophyll a, beta carotene and xanthophylls like fucoxanthin, diatoxanthin and diadinoxanthin are present in diatoms (Conceição *et al.*, 2020). They help in the energy fixation of photosynthesis. The majority of diatoms have a large central vacuole or vacuole pair. The plastids are brown-yellow or brown-greenish in color. Polymer or polysaccharides are usually secreted as locomotive stalks, capsules, tubes, chitin fibers. As a source for the biofuel production, diatoms are of great value. This is due to many reasons. It is a self-generating biomass that can increase in biomass in huge amounts within a short time. They are able to produce up to 300 times more oil than the conventional crops per acre. The other most important property of the diatoms is the ability to grow even in sewage water. Therefore, the diatoms do not have to compete with other plants for land. Triacylglycerols and chrysolaminarin are the main storage compounds of diatoms. When the diatoms are subjected to deprivation of nutrients such as silicon, nitrogen and phosphorus, they tend to produce the triacylglycerols more than chrysolaminarin. These triacylglycerols are stored in oil storage droplets known as oil bodies (Figure 1).

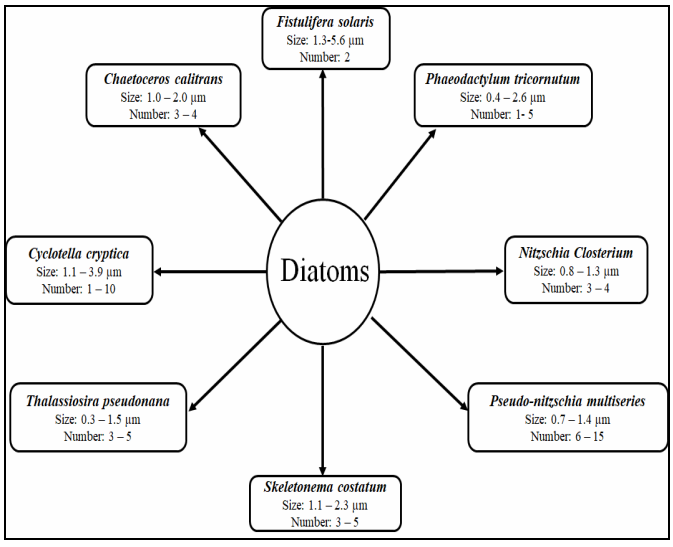

Fig. 1 : The size and the number of oil bodies present in some of the diatom species.

The accumulation of oil droplets is not a rare situation and it is common in animals, plants and other microalgae. But the factors such as the evolutionary history of secondary endosymbiosis, Combination of genes similar to that of animals, plants and bacteria and also diatom specific genes with the complex intracellular structures along with the species diversity among the diatoms have made them more beneficial in the biofuel production and the center of the modern researches (Maeda *et al.*, 2017).

Diatoms can be subjected to genetic manipulations and they facilitate the production of match the existing demand of

the population the diatom genomes have to be manipulated to enhance the natural production of oils and to make it economically viable so that it is accessible to everyone. The different techniques of genetic engineering are helpful here to accomplishing this goal. Diatom species such as *Nannochloropsis* (Eustigmatophyceae) (Rodolfi *et al.*, 2009, Radakovits *et al.*, 2012), *Chlorella vulgaris* (Chlorophyceae) (Converti *et al.*, 2009), *Fistulifera solaris* (Bacillariophycea) (Tanaka *et al.*, 2015) and *Cyclotella cryptica* (Bacillariophyceae) (Roessler, 1988, Kwon *et al.*, 2013) have been studied for the enhancement of biofuel production through different genetic engineering techniques. RNAi approaches, TALEN based mutagenesis, CRISPR-*Cas9* and mega-nucleases are some of the genetic engineering techniques that can be used for the site-specific genome editing. The main aim of utilizing such genome editing tools is to enhance the production of lipid content in the diatom cells through various approaches. For example the lipid catabolism in cells was decreased by a gene targeted knockdown of a multifunctional lipase/ phospholipase/ acyltransferase using antisense and RNAi approaches in the diatom *Thalassiosira pseudonana* strains 1A6 and 1B1 (Trentacoste *et al.*, 2013). Moreover, in another study TAGs content was increased by disruption of the gene encoding a Hotdog-fold thioesterase involved in acyl-CoA hydrolysis (ptTES1) using TALEN-based targeted mutagenesis in *Phaeodactylum tricornutum* (Hao *et al.*, 2018). The lipid biosynthesis pathway was modified by over express the enzymes involved diacylglycerol acyltransferase (DGAT) is a key enzyme that catalyzes the last step of triacylglyceride (TAG) biosynthesis. Out of these techniques that are available for the site-specific genome editing tools CRISPR-Cas 9 takes an important place. It has been successfully and efficiently employed at the site-specific editing of the genome or in the induction of stable genome mutations in diatoms. In fact, this technology renders the benefits of easy adaptability, versatility and cost effectiveness (Sharma *et al.*, 2018). Bacterial conjugation and biolistic particle bombardment are employed in the delivery of CRISPR- Cas 9 system while bacterial conjugation allows the transgene free genome editing.

1.2 CRISPR- Cas9 Technology

CRISPR-Cas9, Clusters of Regularly Interspersed Short Palindromic Repeats- Cas9 proteins are a simple but powerful tool that has been introduced for site specific gene editing so far. CRISPR-Cas9 system has been adapted from a natural phenomenon that takes place in bacterial cells. This system consists of a DNA fragment called CRISPR that carries short, direct repeats separated by spacer elements which are unique sequences (Mojica *et al.*, 2005, Pourcel *et al.*, 2005). When virus attack the bacterial cells, the bacteria capture short pieces of dDNA from attacking virus and produce CRISPR arrays using those captured DNA. These CRISPR arrays help to remember the virus, so if the same virus or similar ones that are closely related attack the bacteria again, the bacteria produce target specific RNA segments called crRNAs from those CRISPR arrays. These crRNA invade the foreign genetic material and provide the protection for the bacteria against those foreign viruses. An AT rich leader sequence always follows the CRISPR array and this region in the DNA is flanked by a set of genes called Cas genes that code for Cas protein that cleave the DNA and make the invading virus dysfunctional.

The CRISPR- Cas 9 system that we perform as a genome editing tool in many organisms has been adapted from this concept. Mainly two molecules are involved in the CRISPR-Cas 9 system.

gRNA: It is the segment of RNA, which is called the guided RNA. It is a small piece of RNA, which is predesigned complementary to the nucleotide sequence on the target DNA. This is around 20 nucleotides long and it is located within a long RNA scaffold. This RNA scaffold binds onto the target DNA, while the gRNA guides the Cas 9 enzyme to the target nucleotide segment.

Cas 9 enzyme: This enzyme is responsible for cutting the target DNA sequence at specific locations once guided by the gRNA. So, it mainly acts as a pair of molecular scissors. It enables this system to add or remove small parts of DNA.

Since the function of the guided RNA is to bind with the target DNA segment of the genome, the sequence of the guided RNA has been always complementary to the target DNA segment. So theoretically the guided RNA will not bind with anywhere on the genome other than the target sequence. However, the complementarity of all 20 bases of the guided RNA is not required for the binding to the target DNA.

The mechanism of the CRISPR-Cas 9 technology works in 3 main steps:

Adaptation stage: Part of the invading DNA gets integrated into the CRISPR array. This is called the protospacer sequence.

***Expression stage*:** The pre-crRNA or the precursor crRNA and the Cas proteins are transcribed. Then, the pre-crRNA is converted to matured crRNA that consists of a repetitive sequence and the invading target sequence or the spacer.

Interference stage: crRNA and Cas proteins cleave the target DNA that is complementary to the crRNA spacer sequence.

The CRISPR-Cas9 system is three types according to the mechanism of recognition and cleavage of the target sequence (Makarova *et al.*, 2011). These three systems can be found simultaneously in the same organism.

Type 1: This system consists of the Cas3 enzyme. It has both the domains of the DNA helicase and DNase for the function of DNA degradation. It further consists of six subtypes which have a variable number of Cas genes. Once the CRISPR array is transcribed into pre-crRNA (precursor-crRNA) Cas6 cleaves it into individual repeat units. This is common to third system too (Sinkunas *et al.*, 2011). In this system a complex called cascade is formed by the complexing between the crRNA and the multiprotein Cas complex. This is helpful in the defending against the viruses by base pairing with the complementary sequence of the foreign DNA (Brouns *et al.*, 2008). Cas 3 creates single strand breaks, which are later degraded by Cas-3 (Westra *et al.*, 2012).

Type 2: This shows a huge difference from the Type1 and type 3. This comprises of a Cas9 enzyme with nuclease activity that can produce blunt ended double stranded breaks. This fact is being updated according to the results of the recent studies (Gasiunas *et al.*, 2012, Wei *et al.*, 2015). This Cas 9 cleaves the DNA molecule 3 nucleotides upstream of the PAM sequence (Jinek *et al.*, 2012). This can also contain genes encoding the enzymes Cas2 or Cas4. Cas9 is responsible of crRNA processing. It also has the two domains

having the nuclease property; HNH and RuvC. HNH functions in cleaving the DNA strand that is complementary to the crRNA and the RuvC functions in the cleaving the DNA strand that is opposite to the complementary strand. Another RNA called trans-activating crRNA (tracrRNA) is a trans-encoded upstream of the type-II CRISPR–Cas locus.

Type 3: The type3 CRISPR Cas system has the Cas-9 enzyme which target either DNA or RNA, while type 1 and type 2 are capable of targeting only the DNA strands. Therefore, it has a dual interference activity. This Cas-9 has also comprised of four subtypes. Unlike the type 2 and type1 that use large complexes of Cas proteins RNA-guided recognition and cleavage, this Cas-9 enzyme uses only a single protein for RNA-guided DNA recognition and cleavage (Jinek *et al.*, 2012). This is very beneficial in the genetic engineering purposes. The chances of non-specific targeting and cytotoxicity can be reduced by inserting the Cas-9 into cells as Cas-9 RNPs together with the single guide RNA. Therefore, the Cas9 protein or the type 3 system is more accurate, simpler and efficient than the type 2 and type 1 CRISPR-Cas-9 system.

1.3 Recently revealed facts about the mechanism of the CRISPR-Cas9 technology

1.3.1 Cas9 can create both blunts and staggered ends

The recent CRISPR-Cas9 technology studies have shown that 1-3 nucleotide 5' staggered ends can be achieved via Cas9 which contains two nuclease domains as HNH and RuvC that are responsible for cleaving the target and non-target strands respectively. Earlier, Cas9 enzymes were only expected to create a blunt cut at the location of 3 base pairs upstream from the adjacent PAM sequence, the protospacer (Li *et al.*, 2015, Jiang & Doudna, 2017, Shou *et al.*, 2018). This study (Shou *et al.*, 2018) shows that the HNH is capable of making a cut exactly at the 3 bases upstream of the PAM sequence while RuvC may make the cut either 3,4 or 5 bases or more upstream. This accurate cleavage of the HNH is believed to be due to the hybrid formation between the sgRNA and the target DNA. The flexible nature of the RuvC is due to the presence of single stranded, non-target displaced flexible DNA strands. Therefore, the Cas9 enzyme has the capability to produce both the blunt end as well as the staggered ends. This will ultimately give rise to template independent insertions, wild type or random insertions in case of repair by the blunt ends and the non-homologous end joining mediated repair (NHEJ) and the overhangs by the staggered ends can give rise to predictable template dependent insertions that is more beneficial in the genome editing in organisms.

1.3.2 Non-Random double stranded break repair

Studies recently carried out show that Cas9 induced double stranded breaks can result in precise genome repairs. This is more important in accurate and precise genome editing (Shen *et al.*, 2018; Shou *et al.*, 2018; Allen *et al.*, 2019; Chakrabarti *et al.*, 2019; Leenay *et al.*, 2019). This was first identified in the particular study (Li *et al.*, 2015). It was further shown that the mutation is not arbitrary by the analysis of the 223 double stranded break repairs in the human genome. The nature of the protospacer sequences is the factor that effects on this to the highest degree (van Overbeek *et al.*, 2016). Furthermore, these studies also suggest that the Cas9 induced mutations are greatly affected by the sequence adjacent to the cleavage site (Shou *et al.*, 2018; Allen *et al.*, 2019; Chakrabarti *et al.*, 2019). The single nucleotide insertion was always found to be the most predictable. The base at the 4^{th} position from the PAM sequence was identical to the inserted nucleotide (Allen *et al.*, 2019; Chakrabarti *et al.*, 2019). It has been found that out of the 4 bases T is more likely to prove this concept. In fact, if T is present at the 4^{th} position of the protospacer, insertion of another T is the most likely to happen. This has been confirmed through various studies done on the human genome, mouse, cell lines as well as the yeast (Liu *et al.*, 2012; Taheri-Ghahfarokhi *et al.*, 2018; Gisler *et al.*, 2019; Leenay *et al.*, 2019). But this predictability decreases in the order T>A>C>G when it is present at the 4^{th} position of the protospacer.

2. CRISPR cas9 technology for the genome editing in diatoms

CRISPR-Cas9 technology have been used on *Chlamydomonas reinhardtii, Phaeodactylum tricornutum, Thalasiossira pseudonana* and *Nannochloropsis.* These studies have utilized various Cas9 delivery methods, including vector-based Cas9 stable expression, episomal delivery, CRISPR Cpf1 variant to Cas9 Cre variants, vector-based transient expression, RNPs (Tanwar *et al.*, 2018). In some studies, they have used particle bombardment method to deliver the DNA into diatom cells while in *P. tricornutum* they have tried electroporation (Sinkunas *et al.*, 2011; Yang *et al.*, 2013; Oishi *et al.*, 2016). In every case the transformation efficiency was not obtained as expected. In most cases it was too low. However, they have been able to achieve successful nuclear transformation using multiple electroporation.

2.1. Enhancing the biofuel production in diatoms

Two different approaches are focused in the research on increasing lipid production in diatoms: firstly, over-expressing enzymes, including the enzyme acyl transferases, to increase or improve the supply of lipogenic enzymes to TAG's synthesis pathways, secondly, stunting of pathways to decrease lipid catabolism or inhibiting the genes coding for lipid catabolizing enzymes. So, the enhancement of the biofuel production in diatoms or in other terms the accumulation of lipids should be based on either of the above approaches where they can be accomplished using different genome editing tools which are site specific.

3. Other approaches for the enhancement of the lipid accumulation in diatoms

3.1 Decreasing lipid catabolism in cells using RNAi approach

This study was focused on the targeted lipid catabolism knockdown, especially on lipases that are responsible for the catalyzing free fatty acid (FA) releases from lipids to increase the accumulation of lipids. Although many strategies have been tried for the increment of the lipid levels in the diatom cells, most of them have resulted in the reduced growth of the engineered diatoms (Wang *et al.*, 2009; Li *et al.*, 2010; Work *et al.*, 2010; Radakovits *et al.*, 2011). Because it has found been difficult to maintain higher rates of biomass accumulation on large economic scale for the biofuel production. The study was therefore based on the knockdown of lipid catabolism, so that the primary carbon pathways linked to growth would not have a major impact. It

was performed on the diatom *T. pseudonana,* which is known to be a lipid accumulating diatom. The gene lipase Thaps3_264297 on which the knockdown was targeted was identified using an Affymetrix whole genome tiling microarray of *T. pseudonana* by analyzing the lipase expression under silicon stress conditions. Under silicone depreciated conditions, 8 hours later the silicone depreciated conditions are applied, *T. pseudonana* arrests its cell cycle (Hildebrand *et al.*, 2007) and accumulates TAG. Using antisense and RNAi approaches the Knockdown of Thaps3_264297 was performed on *T. pseudonana*. For this, the two constructs, pA78cgi that encodes for antisense and pI1001cgi, that encodes for RNAi, were designed to target a portion of Thaps3_264297. Fucoxanthin chlorophyll a/b-binding protein (fcp) promoter carried out the expression of pA78cgi, while an inducible nitrate reductase promoter (Poulsen *et al.*, 2006) controlled the expression of pI1001cgi. Then these constructs were transformed into wild type *T. pseudonana*. All the strains engineered using antisense responded similarly and all of them showed comparatively increased lipid levels than the wild type strains.

3.2 Enhancing the triacylglycerol production using TALEN-based targeted mutagenesis

Most microalgae, in particular diatoms, are able to synthesize, accumulate and store significant quantities of triacyl glycerols (TAGs) as lipid droplets. These occur more frequently in the event of stress, like deprivation of nutrients. TAG biosynthesis in diatoms has two major metabolic pathways: the acyl-CoA dependent and the acyl-CoA independent pathways. In the Acyl-CoA-dependent pathway, which is also known as the Kennedy pathway, three acyl-CoA molecules are used to form a single TAG molecule as substrates for acylation of the glycerol backbone at sn-1, 2 and 3 positions. Acyl-CoA:diacylglycerol acyltransferase (DGAT) enzyme is the final step in the biosynthesis of the TAG. The fatty acyl moiety from the sn-2 position of phosphatidylcholine (PC), to the sn-3 position of DAG, to form TAG and lyso-PC to the Sn-1 position, is catalyzed by the hospholipid:diacylglycerol acyltransferase enzyme (PDAT). DGAT is an effective enzyme to increase lipid production (Guschina & Harwood, 2009). Several studies have demonstrated the ability to improve the lipid content by using diatom DGATs from *Phaeodactylum* and *Nannochloropsis* (Li *et al.*, 2011; Chauton *et al.*, 2013). A plastid-localized Hotdog-fold thioesterase ptTES1 was identified in the diatom, *P. tricornutum*, showing in-vitro acyl-CoA hydrolytic activity in the saturated and polyunsaturated acyl-CoA range of C3 to C20 chain lengths. The emphasis of this particular study was the deactivation of the gene ptTES1, which has increased the TAG content considerably. The development of the mutant strains has however been affected. The ptTES1 knock-out mutant shows, however, that TAG productivity has increased by 1.7 times over wild strain. This study demonstrates the ability to efficiently increase the TAG development of diatoms by TALEN-mediated knockout of target genes (Hao *et al.*, 2018) (Table-1).

Table 1: Different techniques used in the strain improvement of the strains of diatoms for biofuel production (Butler *et al.*, 2020)

Species /Strain	Engineering technique used	Improved features
***Phaeodactylum tricornutum* CCAP 1055/1**	Targeted mutagenesis with meganucleases, TALENS, and CRISPR/Cas9	Production of various heterogeneous biochemicals such as TAGs for biodiesel, fucoxanthin, EPA, chrysolaminarin, sterols, recombinant proteins, vanillin, PHB
***Thalassiosira pseudonana* CCMP 1335**	Conjugation, Transformation by biolistics, gene knockout, gene silencing	Silica biomineralization
***Fistulifera* sp. JPCC DA058**	Biolistic transformation is done involving; endogenous and heterologous promoters	Enhancement of EPA accumulation and biomass productivity

4. Development in CRISPR-Cas9 (optimization) process in diatoms

4.1 Designing an "easy to handle" vector

A common problem that has to be dealt with the use of CRISPR-Cas9 system is that the occurrence of undesired off target mutations. This easy to handle vector was designed as a solution for this. This vector includes all CRISPR-Cas9 components and markers for the selection of transformants (Stukenberg *et al.*, 2018). The pPha-DUAL-(2xNR) vector was used as a basis for integrating the Cas9 and gRNA-coding genes (guide RNA) in the *Phaeodactylum tricornutum* model organisms. It allows the selection of transformed cells through an antibiotic ampicillin or zeocin resistance. The resulting plasmid is called the ptCC9 plasmid. The CRISPR-Cas9 system targeted the vtc2 or pho4 genes of the diatom. The selected gRNA was cloned into the sgRNA of ptCC9. Analysis of transformants showed that the use of this vector results in the easy induction of mono- and biallelic mutations. More significantly, no off-target mutations were identified at two predicted off-target sites of a single CRISPR-Cas9 construct (Stukenberg *et al.*, 2018).

4.2 Introducing Cas12a (Cpf1) in place of Cas9 enzyme

For genome engineering, previously, Cas9 from *Streptococcus pyogenes* known as SpCas9 and similar Cas9 variants were used. Instead of the Cas9 enzyme, Cas12a (Cpf1) was recently published. The Cas12a enzyme significant in the process of genetic engineering due to some special features it possesses compared to Cas9.

(i) For the purposes of Cas12a, it uses a single crRNA, which is shorter crRNA, at least twice as short as the crRNA used for Cas9 instead of a hybrid combination of crRNA and trans-active crRNA (tracrRNA) on Cas9.

(ii) Cas12a catalyzes the development of its own crRNA that makes it easier to efficiently edit genome, while the Cas9 RNA maturation is based on the RNase III non-Cas ribonuclease.

(iii) The Cas9 uses G-rich PAM (5′-NGG-3′) located downstream, but Cas12a needs an upstream protospacer to use T-rich PAM (5′-TTN-3′).

(iv) Cas9 's activity contributes to the development of blunt ends, but Cas12a contributes to staggered ends with five nucleotide overhangs.

(v) Cas9 is an enzyme with two domains of nucleases (RuvC and HNH), while the Cas12a has one single domain of nucleases (RuvC) which cuts the 18-23 bp downstream from the PAM-proximal sequence (Naduthodi *et al.*, 2018).

Discussion

CRISPR –Cas9 technology is one of the most effective site-specific genomes editing tools that are available for the manipulation of the genomes of any kind of organism including humans. However, this can be used for the genome editing of the diatoms for purposes such as enhancing the biofuel production or lipid accumulation which is a need in the world right now. These studies and works which have been done till now proves the potential of accomplishing this task successfully. Not only CRISPR-Cas9 but also other techniques such as TALENS, meganucleases, RNAi antisense also can be utilized for this purpose. Every technique has its own drawbacks and advantages. All of these techniques concentrate basically on either inhibiting the diatom's lipid catabolism, silencing the genes coding for those lipids catabolizing enzymes or enhancing the over expression of genes coding for enzymes such as acyltransferases that increase the acyl flux to the TAG biosynthesis pathway. However, the attempts that have been taken previously to accomplish this also indicates that these approaches have not been successful 100% all time. Because in some cases the genetic manipulations that were induced in these organisms in order to increase the lipid accumulations resulted in the decrease in the biomass accumulation. The approach for decreasing the lipid catabolism by the RNAi approach was able to overcome this barrier. In this particular effort, the lipid level in the mutant strains was increased than in the wild. It required the construction of the DNA coding for antisense and RNAi and then expression of them under the control of some promoter sequences. Fucoxanthin chlorophyll a/b-binding protein (fcp) promoter carried out the expression of pA78cgi, while an inducible nitrate reductase promoter (Roessler, 1988) controlled the expression of pI1001cgi in the diatom cells. In the TALEN based genetic manipulation, the lipid accumulation was enhanced by the inactivation of the plastid-localized Hotdog-fold thioesterase ptTES1 gene, which codes for a hydrolytic enzyme that functions in the catabolism of the lipids that is produced in the diatom cells. The lipid accumulation in the mutant strains increased 1.7 times more than the wild type strains. But the growth of the diatoms was also affected here, which indicates that this needs to be further optimized. The diatoms showed increased levels in lipid accumulation specifically in the conditions of nutrient deprivation. When they are made to starve for nutrients such as silicone, increased levels of lipids were observed.

Conclusion

Diatom can be used as the most reliable and the advantageous source for the production of biofuels. Not only as a solution to the fossil fuel depletion, but also as an environment friendly source it is the best option we have. The genome editing tools that have been discussed in this review prove that the large-scale production can be made possible by inducing the diatom cells to accumulate the lipids in higher amounts. The processing is the only part left for the humans. But still the procedures of genetic engineering have to be optimized in order not to comprise the growth of the diatoms and induce most specific genome editing. Along with the genetic engineering techniques other factors such as growing media of the diatoms and the other conditions also should be optimized according to the manipulation technique that is used to get the best results in terms of lipid accumulation as well as the biomass accumulation. Among various editing tools available, CRISPR-Cas 9 shows distinct characteristics to deliver promising beneficial results than any other technique due to its specific mechanism of action. Therefore, CRISPR-Cas9 technology can be optimized for the accumulation of lipids ultimately for the production of biofuels more effectively.

Acknowledgement: Authors are thankful to Lovely Professional University for providing necessary facilities for the research study.

References

Allen, F.; Crepaldi, L.; Alsinet, C.; Strong, A.J.; Kleshchevnikov, V.; De Angeli, P.; Páleníková, P.; Khodak, A.; Kiselev, V. and Kosicki, M. (2019). Predicting the mutations generated by repair of Cas9-induced double-strand breaks. Nature biotechnology 37: 64-72.

Armbrust, E.V.; Berges, J.A.; Bowler, C.; Green, B.R.; Martinez, D.; Putnam, N.H.; Zhou, S.; Allen, A.E.; Apt, K.E. and Bechner, M. (2004). The genome of the diatom Thalassiosira pseudonana: ecology, evolution, and metabolism. Science, 306: 79-86.

Brouns, S.J.; Jore, M.M.; Lundgren, M.; Westra, E.R.; Slijkhuis, R.J.; Snijders, A.P.; Dickman, M.J.; Makarova, K.S.; Koonin, E.V. and Van Der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science, 321: 960-964.

Butler, T.; Kapoore, R.V. and Vaidyanathan, S. (2020). Phaeodactylum tricornutum: A Diatom Cell Factory. Trends in Biotechnology, 38: 606-622.

Chakrabarti, A.M.; Henser-Brownhill, T.; Monserrat, J.; Poetsch, A.R.; Luscombe, N.M. and Scaffidi, P. (2019). Target-specific precision of CRISPR-mediated genome editing. Molecular cell, 73: 699-713. e696.

Chauton, M.S.; Winge, P.; Brembu, T.; Vadstein, O. and Bones, A.M. (2013). Gene regulation of carbon fixation, storage, and utilization in the diatom Phaeodactylum tricornutum acclimated to light/dark cycles. Plant physiology, 161: 1034-1048.

Conceição, D.; Lopes, R.G.; Derner, R.B.; Cella, H.; do Carmo, A.P.B.; D'Oca, M.G.M.; Petersen, R.; Passos, M.F.; Vargas, J.V.C. and Galli-Terasawa, L.V. (2020). The effect of light intensity on the production and accumulation of pigments and fatty acids in Phaeodactylum tricornutum. Journal of Applied Phycology, 1-9.

Converti, A.; Casazza, A.A.; Ortiz, E.Y.; Perego, P. and Del Borghi, M. (2009). Effect of temperature and nitrogen concentration on the growth and lipid content of Nannochloropsis oculata and Chlorella vulgaris for

biodiesel production. Chemical Engineering and Processing: Process Intensification, 48: 1146-1151.

d'Ippolito, G.; Sardo, A.; Paris, D.; Vella, F.M.; Adelfi, M.G.; Botte, P.; Gallo, C. and Fontana, A. (2015). Potential of lipid metabolism in marine diatoms for biofuel production. Biotechnology for biofuels, 8: 28.

Demirbas, A. and Demirbas, M.F. (2010). Energy from algae. Algae Energy,p.^pp. 97-138. Springer.

Gasiunas, G.; Barrangou, R.; Horvath, P. and Siksnys, V. (2012). Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proceedings of the National Academy of Sciences, 109: E2579-E2586.

Gautam, S.; Arya, A. and Vinayak, V. Protocol to Establish Axenic Cultures for Diatoms of Fresh Water.

Gautam, S.; Kashyap, M.; Gupta, S.; Kumar, V.; Schoefs, B.; Gordon, R.; Jeffryes, C.; Joshi, K.B. and Vinayak, V. (2016). Metabolic engineering of tio 2 nanoparticles in nitzschia palea to form diatom nanotubes: an ingredient for solar cells to produce electricity and biofuel. RSC advances, 6: 97276-97284.

Gisler, S.; Gonçalves, J.P.; Akhtar, W.; de Jong, J.; Pindyurin, A.V.; Wessels, L.F. and van Lohuizen, M. (2019). Multiplexed Cas9 targeting reveals genomic location effects and gRNA-based staggered breaks influencing mutation efficiency. Nature communications, 10: 1-14.

Guiry, M.D. (2012). How many species of algae are there? Journal of phycology, 48: 1057-1063.

Guschina, I.A. and Harwood, J.L. (2009). Algal lipids and effect of the environment on their biochemistry. Lipids in aquatic ecosystems, pp. 1-24. Springer.

Hao, X.; Luo, L.; Jouhet, J.; Rébeillé, F.; Maréchal, E.; Hu, H.; Pan, Y.; Tan, X.; Chen, Z. and You, L. (2018). Enhanced triacylglycerol production in the diatom Phaeodactylum tricornutum by inactivation of a Hotdog-fold thioesterase gene using TALEN-based targeted mutagenesis. Biotechnology for biofuels, 11: 312.

Hildebrand, M.; Frigeri, L.G. and Davis, A.K. (2007). Synchronized growth of thalassiosira pseudonana (bacillariophyceae) provides novel insights into cell-wall synthesis processes in relation to the cell cycle 1. Journal of Phycology, 43: 730-740.

Hildebrand, M.; Davis, A.K.; Smith, S.R.; Traller, J.C. and Abbriano, R. (2012). The place of diatoms in the biofuels industry. Biofuels 3: 221-240.

Hoffmann M, Kleine-Weber H, Krüger N, Mueller MA, Drosten C and Pöhlmann S (2020) The novel coronavirus 2019 (2019-nCoV) uses the SARS-coronavirus receptor ACE2 and the cellular protease TMPRSS2 for entry into target cells. BioRxiv.

Huang, W. and Daboussi, F. (2017). Genetic and metabolic engineering in diatoms. Philosophical Transactions of the Royal Society B: Biological Sciences, 372: 20160411.

Jiang, F. and Doudna, J.A. (2017). CRISPR–Cas9 structures and mechanisms. Annual review of biophysics, 46: 505-529.

Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J.A. and Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. science, 337: 816-821.

Kaeriyama, H.; Katsuki, E.; Otsubo, M.; Yamada, M.; Ichimi, K.; Tada, K. and Harrison, P.J. (2011). Effects of temperature and irradiance on growth of strains belonging to seven Skeletonema species isolated from Dokai Bay, southern Japan. European journal of phycology, 46: 113-124.

Kwon, J-H.; Bernát, G.; Wagner, H.; Rögner, M. and Rexroth, S. (2013). Reduced light-harvesting antenna: consequences on cyanobacterial metabolism and photosynthetic productivity. Algal Research, 2: 188-195.

Leenay, R.T.; Aghazadeh, A.; Hiatt, J.; Tse, D.; Roth, T.L.; Apathy, R.; Shifrut, E.; Hultquist, J.F.; Krogan, N. and Wu, Z. (2019). Large dataset enables prediction of repair after CRISPR–Cas9 editing in primary T cells. Nature biotechnology 37: 1034-1037.

Li, Y.; Han, D.; Sommerfeld, M. and Hu, Q. (2011). Photosynthetic carbon partitioning and lipid production in the oleaginous microalga Pseudochlorococcum sp.(Chlorophyceae) under nitrogen-limited conditions. Bioresource Technology, 102: 123-129.

Li, Y.; Han, D.; Hu, G.; Sommerfeld, M. and Hu, Q. (2010). Inhibition of starch synthesis results in overproduction of lipids in Chlamydomonas reinhardtii. Biotechnology and bioengineering, 107: 258-268.

Li, Y.; Park, A.I.; Mou, H.; Colpan, C.; Bizhanova, A.; Akama-Garren, E.; Joshi, N.; Hendrickson, E.A.; Feldser, D. and Yin, H. (2015). A versatile reporter system for CRISPR-mediated chromosomal rearrangements. Genome biology, 16: 111.

Liu, W.; Huang, Z.; Li, P.; Xia, J. and Chen, B. (2012). Formation of triacylglycerol in *Nitzschia closterium* f. minutissima under nitrogen limitation and possible physiological and biochemical mechanisms. Journal of Experimental Marine Biology and Ecology, 418: 24-29.

Maeda, Y.; Nojima, D.; Yoshino, T. and Tanaka, T. (2017). Structure and properties of oil bodies in diatoms. Philosophical Transactions of the Royal Society B: Biological Sciences, 372: 20160408.

Maity, J.P.; Bundschuh, J.; Chen, C-Y. and Bhattacharya, P. (2014). Microalgae for third generation biofuel production, mitigation of greenhouse gas emissions and wastewater treatment: Present and future perspectives–A mini review. Energy, 78: 104-113.

Makarova, K.S.; Haft, D.H.; Barrangou, R.; Brouns, S.J.; Charpentier, E.; Horvath, P.; Moineau, S.; Mojica, F.J.; Wolf, Y.I. and Yakunin, A.F. (2011). Evolution and classification of the CRISPR–Cas systems. Nature Reviews Microbiology, 9: 467-477.

Mojica, F.J.; García-Martínez, J. and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. Journal of molecular evolution, 60: 174-182.

Naduthodi, M.I.S.; Barbosa, M.J. and van der Oost, J. (2018). Progress of CRISPR-Cas based genome editing in photosynthetic microbes. Biotechnology journal, 13: 1700591.

Oishi, I.; Yoshii, K.; Miyahara, D.; Kagami, H. and Tagami, T. (2016). Targeted mutagenesis in chicken using CRISPR/Cas9 system. Scientific reports, 6: 23980.

Poulsen, N.; Chesley, P.M. and Kröger, N. (2006). Molecular genetic manipulation of the diatom Thalassiosira pseudonana (Bacillariophyceae) 1. Journal of Phycology, 42: 1059-1065.

Pourcel, C.; Salvignol, G. and Vergnaud, G. (2005). CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology, 151: 653-663.

Radakovits, R.; Eduafo, P.M. and Posewitz, M.C. (2011). Genetic engineering of fatty acid chain length in Phaeodactylum tricornutum. Metabolic engineering, 13: 89-95.

Radakovits, R.; Jinkerson, R.E.; Fuerstenberg, S.I.; Tae, H.; Settlage, R.E.; Boore, J.L. and Posewitz, M.C. (2012). Draft genome sequence and genetic transformation of the oleaginous alga Nannochloropsis gaditana. Nature communications, 3: 1-11.

Rodolfi, L.; Chini Zittelli, G.; Bassi, N.; Padovani, G.; Biondi, N.; Bonini, G. and Tredici, M.R. (2009). Microalgae for oil: Strain selection, induction of lipid synthesis and outdoor mass cultivation in a low□cost photobioreactor. Biotechnology and bioengineering, 102: 100-112.

Roessler, P.G. (1988). Effects of silicon deficiency on lipid composition and metabolism in the diatom Cyclotella cryptica 1. Journal of phycology, 24: 394-400.

Sayanova, O. and Napier, A.J. (2016). Metabolic engineering of microalgae for sustainable production of omega-3 long chain polyunsaturated fatty acids. Current Biotechnology, 5: 198-212.

Sharma, A.K.; Nymark, M.; Sparstad, T.; Bones, A.M. and Winge, P. (2018). Transgene-free genome editing in marine algae by bacterial conjugation–comparison with biolistic CRISPR/Cas9 transformation. Scientific reports, 8: 1-11.

Shen, M.W.; Arbab, M.; Hsu, J.Y.; Worstell, D.; Culbertson, S.J.; Krabbe, O.; Cassa, C.A.; Liu, D.R.; Gifford, D.K. and Sherwood, R.I. (2018). Predictable and precise template-free CRISPR editing of pathogenic variants. Nature, 563: 646-651.

Shou, J.; Li, J.; Liu, Y. and Wu, Q. (2018). Precise and predictable CRISPR chromosomal rearrangements reveal principles of Cas9-mediated nucleotide insertion. Molecular cell, 71: 498-509. e494.

Sinkunas, T.; Gasiunas, G.; Fremaux, C.; Barrangou, R.; Horvath, P. and Siksnys, V. (2011). Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. The EMBO journal, 30: 1335-1342.

Stukenberg, D.; Zauner, S.; Dell'Aquila, G. and Maier, U.G. (2018), Optimizing CRISPR/Cas9 for the diatom Phaeodactylum tricornutum. Frontiers in plant science, 9: 740.

Taheri-Ghahfarokhi, A.; Taylor, B.J.M.; Nitsch, R.; Lundin, A.; Cavallo, A-L.; Madeyski-Bengtson, K.; Karlsson, F.; Clausen, M.; Hicks, R. and Mayr, L.M. (2018). Decoding non-random mutational signatures at Cas9 targeted sites. Nucleic acids research, 46: 8417-8434.

Tanaka, T.; Maeda, Y.; Veluchamy, A.; Tanaka, M.; Abida, H.; Maréchal, E.; Bowler, C.; Muto, M.; Sunaga, Y. and Tanaka, M. (2015). Oil accumulation by the oleaginous diatom Fistulifera solaris as revealed by the genome and transcriptome. The Plant Cell, 27: 162-176.

Tanwar, A.; Sharma, S. and Kumar, S. (2018). Targeted genome editing in algae using CRISPR/Cas9. Indian Journal of Plant Physiology, 23: 653-669.

Trentacoste, E.M.; Shrestha, R.P.; Smith, S.R.; Glé, C.; Hartmann, A.C.; Hildebrand, M. and Gerwick, W.H. (2013). Metabolic engineering of lipid catabolism increases microalgal lipid accumulation without compromising growth. Proceedings of the National Academy of Sciences, 110: 19748-19753.

van Overbeek, M.; Capurso, D.; Carter, M.M.; Thompson, M.S.; Frias, E.; Russ, C.; Reece-Hoyes, J.S.; Nye, C.; Gradia, S. and Vidal, B. (2016). DNA repair profiling reveals nonrandom outcomes at Cas9-mediated breaks. Molecular cell, 63: 633-646.

Voloshin, R.A.; Rodionova, M.V.; Zharmukhamedov, S.K.; Veziroglu, T.N. and Allakhverdiev, S.I. (2016). Biofuel production from plant and algal biomass. International journal of hydrogen energy, 41: 17257-17273.

Wang, Z.T.; Ullrich, N.; Joo, S.; Waffenschmidt, S. and Goodenough, U. (2009). Algal lipid bodies: stress induction, purification, and biochemical characterization in wild-type and starchless Chlamydomonas reinhardtii. Eukaryotic cell, 8: 1856-1868.

Wei, Y.; Terns, R.M. and Terns, M.P. (2015). Cas9 function and host genome sampling in Type II-A CRISPR–Cas adaptation. Genes and development, 29: 356-361.

Westra, E.R.; van Erp, P.B.; Künne, T.; Wong, S.P.; Staals, R.H.; Seegers, C.L.; Bollen, S.; Jore, M.M.; Semenova, E. and Severinov, K. (2012). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. Molecular cell, 46: 595-605.

Work, V.H.; Radakovits, R.; Jinkerson, R.E.; Meuser, J.E.; Elliott, L.G.; Vinyard, D.J.; Laurens, L.M.; Dismukes, G.C. and Posewitz, M.C. (2010). Increased lipid accumulation in the Chlamydomonas reinhardtii sta7-10 starchless isoamylase mutant and increased carbohydrate synthesis in complemented strains. Eukaryotic cell, 9: 1251-1261.

Yang, Z-K.; Niu, Y-F.; Ma, Y-H.; Xue, J.; Zhang, M-H.; Yang, W-D.; Liu, J-S.; Lu S-H.; Guan, Y. and Li, H-Y. (2013). Molecular and cellular mechanisms of neutral lipid accumulation in diatom following nitrogen deprivation. Biotechnology for biofuels, 6: 67.