

Genome editing in diatoms: achievements and goals

Peter G. Kroth¹  · Atle M. Bones² · Fayza Daboussi³ · Maria I. Ferrante⁴ · Marianne Jaubert⁷ · Misha Kolot^{5,6} · Marianne Nymark² · Carolina Río Bártulos¹ · Andrés Ritter⁷ · Monia T. Russo⁴ · Manuel Serif³ · Per Winge² · Angela Falciaiore⁷

Abstract

Diatoms are major components of phytoplankton and play a key role in the ecology of aquatic ecosystems. These algae are of great scientific importance for a wide variety of research areas, ranging from marine ecology and oceanography to biotechnology. During the last 20 years, the availability of genomic information on selected diatom species and a substantial progress in genetic manipulation, strongly contributed to establishing diatoms as molecular model organisms for marine biology research. Recently, tailored TALEN endonucleases and the CRISPR/Cas9 system were utilized in diatoms, allowing targeted genetic modifications and the generation of knockout strains. These approaches are extremely valuable for diatom research because breeding, forward genetic screens by random insertion, and chemical mutagenesis are not applicable to the available model species *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*, which do not cross sexually in the lab. Here, we provide an overview of the genetic toolbox that is currently available for performing stable genetic modifications in diatoms. We also discuss novel challenges that need to be addressed to fully exploit the potential of these technologies for the characterization of diatom biology and for metabolic engineering.

Keywords Diatom · Genome editing · TALEN · CRISPR · Conjugation · Promoter · Mutant screening

✉ Peter G. Kroth
Peter.Kroth@uni-konstanz.de

✉ Angela Falciaiore
angela.falciaiore@upmc.fr

- ¹ Fachbereich Biologie, Universität Konstanz, 78457 Konstanz, Germany
- ² Cell, Molecular Biology and Genomics Group, Department of Biology, Norwegian University of Science and Technology, 7491 Trondheim, Norway
- ³ LISBP, Université de Toulouse, CNRS, INSA, 135 Avenue de Rangueil, 31077 Toulouse, France
- ⁴ Integrative Marine Ecology, Stazione Zoologica Anton Dohrn, Villa Comunale 1, Naples 80121, Italy
- ⁵ Institute of Plant Sciences and Genetics in Agriculture, Robert H. Smith Faculty of Agriculture, Food and Environment, The Hebrew University of Jerusalem, 76100 Rehovot, Israel
- ⁶ Department of Biochemistry and Molecular Biology, Tel-Aviv University, Tel-Aviv 69978, Israel
- ⁷ Laboratoire de Biologie Computationnelle et Quantitative, Institut de Biologie Paris-Seine, Sorbonne Université, CNRS, 75005 Paris, France

Introduction

Diatoms are unicellular algae which represent an essential part of the planktonic and benthic communities (Malviya et al. 2016) and contribute significantly to global primary production (Armbrust 2009). They also have peculiar cellular and metabolic features, which likely are a result of their evolution by multiple secondary endosymbiotic events (Archibald 2015). During the last two decades, significant efforts have been devoted to develop molecular tools to foster studies of diatom biology and ecology, but also to establish these algae as a significant, renewable and sustainable resource of biomass for feed, food, energy, and other value-added products. After the first reports on successful biolistic transformation of diatoms such as *Cyclotella cryptica* (Dunahay et al. 1995) and *Phaeodactylum tricornutum* (Apt et al. 1996; Falciaiore et al. 1999) had been published, this transformation method quickly became the standard for genetic diatom modification (Fischer et al. 1999, Apt et al. 2002; Buhmann et al. 2014; Ifuku et al. 2015). More recently, highly efficient protocols for introducing transgenes into diatoms via electroporation (Niu et al. 2012; Miyahara

et al. 2013; Zhang and Hu 2014) or bacterial conjugation (Karas et al. 2015) have been added to the diatom transformation portfolio. In addition to standard vectors containing selectable marker and reporter gene constructs, a number of expression vectors for high-throughput protein tagging and overexpression studies in *P. tricornutum* have been generated based on the Gateway technology (Siaut et al. 2007). Furthermore, Golden Gate cloning based on Type IIS restriction enzymes (Engler et al. 2008) has been applied to assemble multiple DNA fragments to be expressed in the diatom *Thalassiosira pseudonana* (Hopes et al. 2016). However, a versatile Modular Cloning (MoClo) tool box with a library of basic components such as promoters, CDS, terminators (Weber et al. 2011) is not yet available for any diatom model species.

Another useful application of diatom transformation is gene silencing (De Riso et al. 2009; Lepetit et al. 2013), enabling the downregulation of gene expression. While gene silencing may reduce the amount of a target protein, only knockout mutations allow the complete elimination or the modification of the respective gene product. In this respect, the generation of knockout strains via genomic engineering recently became a powerful tool for diatoms. Different strategies for targeted genome editing have been developed by a number of laboratories, which are either based on TALENs (Transcription Activator-Like Effector Nucleases) (Christian et al. 2010) or CRISPR (clustered regulatory interspaced short palindromic repeats)/Cas9 (Doudna and Charpentier 2014) technologies. Importantly, these methods allow the modification of both alleles of a gene, permitting a complete loss of a gene or the modification of a specific gene function. This is especially helpful for the diploid diatoms and the molecular model species, such as *P. tricornutum* and *T. pseudonana*, which do not cross sexually in the lab. In this review, both established and developing approaches for genome editing in diatoms are described.

Genome editing

Genome editing approaches allow the direct modification of one or more copies of a gene within a cell using engineered nucleases. One of the first systems applied in diatoms was based on meganucleases (Daboussi et al. 2014). These endonucleases bind to specific sites in genomes and can be used to replace, eliminate or modify sequences. However, the production of meganucleases is time consuming (several weeks) and not sufficiently flexible to envisage a significant role in microalgae genome engineering (Smith et al. 2006). Instead, TALENs or CRISPR/Cas9 gRNA complexes (Malzahn et al. 2017) can be easily designed to edit most of the genomic regions (Moscou and Bogdanove 2009; Doudna and Charpentier 2014). Those approaches have in

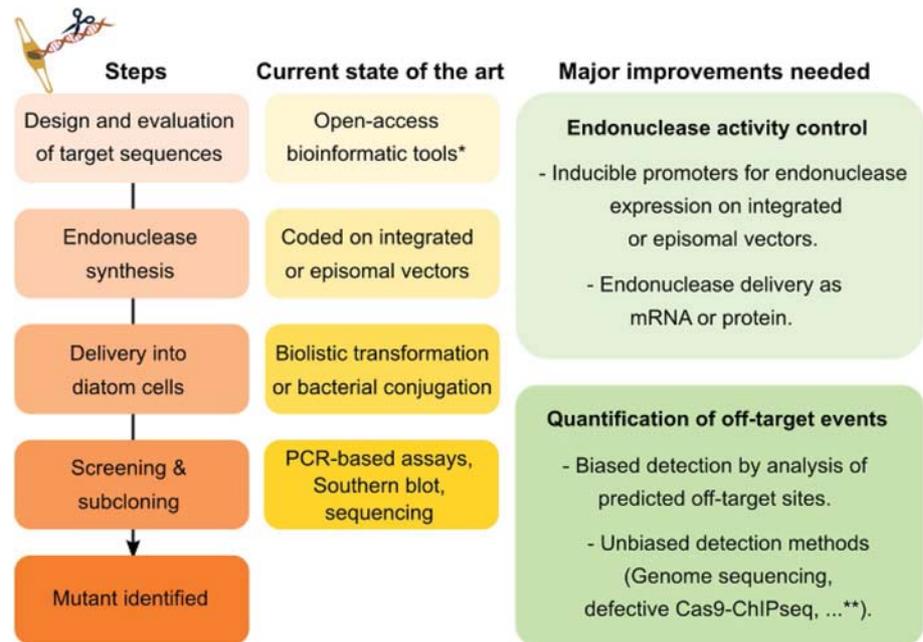
common that they may induce a double-DNA strand break (DSB), which subsequently triggers an error-prone DNA repair system such as non-homologous end-joining (NHEJ) that may introduce insertions or deletions (INDELs), eventually resulting in the inactivation of the targeted gene (Lieber 2010). More rarely, DSBs can also be fixed by homology-directed repair (HDR), which involves homologous recombination with a donor DNA sequence (Steinert et al. 2016). In many organisms, HDR has been used to introduce precise DNA mutations, to insert specific sequences in a target locus, or to achieve targeted gene replacements. In the following paragraphs, we report the key steps and tools already available to perform efficient genome editing in diatoms, including steps that require further improvement (Fig. 1).

TALE nucleases

TALE nucleases are chimeric proteins created through the fusion of a TALE DNA-binding domain designed to recognize and bind a specific sequence with a non-specific catalytic head, the *FokI* endonuclease (Bitinaite et al. 1998). The DNA-binding domain of TALENs is derived from transcription activator-like effectors (TALEs), that are produced by plant pathogenic bacteria of the *Xanthomonas* genus (Zu et al. 2013). This binding domain is composed of 14–24 repeat units of 33–35 amino acids nearly identical to each other, except for two polymorphic amino acids called RVDs (repeat variable di-residue) located at positions 12 and 13, being responsible for the specific recognition of a target nucleotide. Based on this feature, it is possible to customize the DNA-binding domain to recognize virtually any sequence. As *FokI* functions only as a dimer, two independent monomeric TALENs are required to generate a DSB. The target site is then determined by the two DNA-binding domains resulting in high-sequence specificity (Christian et al. 2010). The proof of concept of targeted gene inactivation and targeted sequence insertion induced by TALEN has been established in microalgae recently (Daboussi et al. 2014), demonstrating its usefulness for rewriting diatom metabolism, e.g., knocking out a key gene for storage carbohydrate synthesis in *P. tricornutum* resulted in a strong lipid production (Daboussi et al. 2014). With the same approach, knockout mutants for a phytochrome photoreceptor showing loss of responsiveness to far-red light have been generated (Fortunato et al. 2016). In addition, a urease gene (Weyman et al. 2015) and a nitrate reductase gene (McCarthy et al. 2017) have been knocked out through HDR.

Technical improvements should decrease the TALENs production costs and increase the feasibility of producing these endonucleases in a short time. A protocol by Sanjana et al. (2012) for generating fast, easy, and affordable TALEN constructs, based on a library of monomers, each representing one repeat unit of a TALE, and a two-step Golden-gate

Fig. 1 Current state of the art and further improvements needed for genome editing in diatoms. Suggested articles: *Doyle et al. (2012), Lin et al. (2014), Rastogi et al. (2016), Haeussler et al. (2016), **Zischewski et al. (2017)



reaction, has recently been adopted for diatoms generating knockout strains for blue light Aureochrome photoreceptors in *P. tricornutum* (Serif et al. 2017). Moreover, compact TALENs (cTALENs) (Beurdeley et al. 2013), a new TALEN scaffold where the *FokI* catalytic head has been replaced by the *I-TevI* homing endonuclease, have been designed for diatoms. As *I-TevI* can generate DSB in a monomeric form, only one cTALEN construct is required instead of two as for TALENs. The specificity of cTALEN is mediated both by the DNA-binding domain and the sequence specificity of the *I-TevI* domain itself. The proof of concept of cTALENs has been established recently in diatoms by targeting the *UGPI* gene (Daboussi et al. unpublished data). However, the probability to find an *I-TevI* recognition motif (CNNNGN) adjacent to the TALE binding in the *P. tricornutum* genome is low (e.g., one occurrence every 300 bp-1 kbp), making the use of cTALEN less versatile than classical TALEN scaffolds.

CRISPR/Cas9

In the last decade, the bacterial CRISPR and *CRISPR-associated protein-9 nuclease* (*Cas9*) system has been demonstrated to mediate gene editing in a large variety of eukaryotic cells (Cong et al. 2013). The CRISPR system requires a guide RNA fragment complementary to the target site to recognize a specified target sequence in the genome, and a nuclease (e.g., *Cas9*) to generate a doublestrand break (DSB). The custom single-guide RNA (sgRNA) contains a targeting sequence (crRNA sequence) homologous to the genomic region to be modified, and a *Cas9* nuclease-recruiting sequence (tracrRNA). The binding specificity is based on the sgRNA and a 3-nucleotide

downstream sequence called the protospacer adjacent motif (PAM; NGG, in the case of *S. pyogenes* *Cas9*). The *Cas9* nuclease carries two nuclease domains (HNH and RuvC) and cleaves both DNA strands generating DSBs at sites defined by the 20-nucleotide guide sequence (Chen et al. 2014). The first paper reporting stable CRISPR/*Cas9*-based gene editing in algae (Nymark et al. 2016) describes a vector encoding both a codon-optimized synthetic *Cas9*, controlled by the strong diatom *PtLHCF2* promoter, and sgRNAs, controlled by the RNA polymerase III *PtU6* snRNA promoter (Nymark et al., 2017). The system is adaptable to any specific target sequence by simply replacing the 20 bp fragment at the 5' end of the sgRNA fragment of the vector. By this approach Nymark et al. (2016) targeted the gene encoding the *P. tricornutum* chloroplast signal recognition particle 54 (CpSRP54), which is involved in insertion of chloroplast proteins into the thylakoid membrane (Kirst and Melis 2014). A high proportion of biallelic modifications and an almost complete absence of wild-type alleles were observed. Increased sensitivity of CpSRP54-modified cell lines to high-quantum blue light exposure provided functional evidence for a successful modification. Hopes et al. (2016) reported successful editing of the urease gene of another diatom, *T. pseudonana*, using a CRISPR system with two sgRNAs. This study also indicates the feasibility of targeting multiple genes in a single transformation event using multiplex sgRNAs.

Expression of endonuclease genes

Until now, genome editing in diatoms has been achieved by expressing the endonucleases as a transgene. In most of the studies, endonuclease expression was driven by the promoter regions of the genes encoding fucoxanthin chlorophyll *a/lc*-binding proteins (FCPs, *Lhcf* genes). These promoters provide a robust expression, but a drawback of these promoters is their light dependency, leading to a potentially oscillating expression when the cells are grown in a day/night cycle (Russo et al. 2015).

A major concern regarding the presence of endonucleases in the cells is the possibility that they can generate off-target DNA cleavages, resulting in undesired permanent side effects. Therefore, it is important to control the expression of these enzymes and to minimize the exposure of the genome to their activity. Since the mid-nineties, a number of studies described the utilization of different promoters for transgene expression in diatoms (for a compilation see Huang and Daboussi 2017). For temporary induction or inhibition of endonuclease expression, the nitrate reductase (NR) promoter has been used, as the expression of the NR gene can be induced or inhibited by exchanging ammonia by nitrate as a nitrogen source or vice versa. The NR promoter construct, originally developed for the diatom *Cylindrotheca fusiformis* (Poulsen and Kröger 2005), and then adapted for *P. tricornutum* and *T. pseudonana* (Poulsen et al. 2006; Miyagawa et al. 2009; Hempel et al. 2011; Chu et al. 2016; Lau et al. 2016), has already been used to drive a strong expression of TALEN (Serif et al. 2017) and Cas9 (Stukenberg et al. 2018) in *P. tricornutum* cells grown in nitrate-containing media. However, studies of NR promoter-driven GFP expression in *P. tricornutum* also indicated that gene expression cannot be completely switched off in the absence of nitrate (Chu et al. 2016). Furthermore, nitrate deprivation may affect the photosynthetic capacity, the chlorophyll content, and the accumulation of neutral lipids, leading to possible secondary phenotypes (Valenzuela et al. 2012; Alipanah et al. 2015; Chu et al. 2016; Shrestha and Hildebrand 2017). An alternative system could be derived from the silicon starvation inducible promoters (SSIPs), that are induced in the absence of silica in *T. pseudonana* and *C. cryptica* (Shrestha and Hildebrand 2017). Under such silica-limited conditions, while both cell division and cell growth are blocked, the energy is channeled into the gene expression process, with little detrimental effect on cellular physiology (Shrestha and Hildebrand 2017).

The detailed characterization of the regulatory regions controlling gene expression in diatoms appear now as a mandatory step to set up inducible systems based on endogenous elements. Due to the limited data on promoter

structures, usually 0.5–1 kbp fragments upstream of the coding regions have been utilized without further analysis, based on the assumption that all necessary regulatory regions to drive expression should be included. Recently, several publications identified diatom transcription factors (TFs) (Rayko et al. 2010; Buitrago-Flórez et al. 2014; Matthijs et al. 2016, 2017; Kroth et al. 2017), but their binding site in the promoters of their target genes are still largely uncharacterized. Information is only available for CO₂-cAMP-responsive cis-elements (Ohno et al. 2012), iron cis-regulatory elements (Yoshinaga et al. 2014), and a transcription enhancer-like sequence in the 5'-flanking region of the *P. tricornutum* *Lhcf2* gene (Russo et al. 2015). To identify promoters that might be similarly active in different diatoms, a selection of highly conserved non-coding elements (CNEs) have recently been identified in a comparative genome study of *Pseudo-nitzschia multistriata* (Basu et al. 2017). However, none of the five CNEs studied proved to be functional in GUS assay experiments in *P. tricornutum* (Ferrante et al. unpublished data). Systematic studies of the transcription terminators are also missing, although a recent study indicates that transcription terminators and 3' UTRs may have an influence on gene expression in diatoms (Slattery et al. 2018). In parallel with endogenous inducible systems, the identification of inducible synthetic promoters and of heterologous gene expression systems working in diatoms, will also permit to better control the expression of endonucleases in these algae.

In addition to the development of very tightly controlled promoters, nuclease expression from an episome that is not integrated into the genome, is a valid option. Seminal to the development of such tools was the finding by Karas et al. (2015), that DNA (plasmids/episomes) can be transferred to diatoms through bacterial conjugation. This approach is based on the functionality of plasmids with a broad host range that possess genetic elements such as a yeast centromeric CEN6-ARSH4-HIS3 sequence that enables episome maintenance in *P. tricornutum* (Karas et al. 2015; Diner et al. 2017). Such episomes can be controlled by standard selection principles such as antibiotic resistance, and thereby can be eliminated by removal of the selection pressure. The conjugation system has recently been refined (Diner et al. 2016) and improved for gene editing of *P. tricornutum* (Slattery et al. 2018). By removing selection shortly after having identified the mutation, the episome carrying CRISPR/Cas9 can be eliminated to limit genome exposure to the Cas9 enzyme. To quantify the extent of off-target mutations, whole-genome re-sequencing of a mutant generated with this approach has been performed. The analysis has indicated absence of mutations in the predicted off-target sites, but a limited number of small-genomic variations were observed in wild-type as well as in the mutant strain, perhaps

due to mutations during culturing over a period of 1 year (Russo, Ferrante, unpublished data).

Screening

Following the genetic delivery of endonucleases, the identification of cells in which both alleles are effectively mutated, is still time-consuming, but an essential process to obtain a fully clonal cell line knocked out for the targeted gene (Zischewski et al. 2017). Transformant colonies can exhibit for the targeted gene: (1) only the wild-type sequence if no editing has taken place, (2) identical biallelic mutations, possibly resulting from DSB induced gene conversion mechanism such as break-induced replication (BIR) (Nymark et al. 2016), (3) different mutations in the two alleles, or (4) a heterogeneous mixture of wild-type and different mutated sequences, because of non-simultaneous action of the nucleases on both alleles following transformation, or repeated action of a nuclease after a minor mutation.

The occurrence and the frequency of mutagenic events within a transgenic colony can be assessed by different PCR-based assays of the targeted gene region: T7 endonuclease I assay (Daboussi et al. 2014; Slattery et al. 2018), High Resolution Melting (HRM) analysis (Nymark et al. 2016), loss of restriction enzyme sites, or a PCR band shift assay (Hopes et al. 2016). Recently, digital PCR technology has become an emerging tool for INDEL detection (Santurtún et al. 2017). The very high number of partitions permits statistical modelling to precisely determine the actual number of target DNA molecules present in the original sample. Once colonies containing mutations are detected and their frequency estimated with one of the approaches described above, Sanger sequencing of the targeted region of the initial transgenic colonies or of their respective subclones, can be done to characterize the nature of the mutations.

Single chromatogram traces corresponding to both allele sequences will indicate a homozygous biallelic mutant (Nymark et al. 2016). Multiple traces with overlapping peaks at the mutation sites will be obtained in case of heterozygous biallelic or monoallelic mutations. To differentiate mutations in the two alleles, allele-specific primers can be generated from genome-sequencing trace files (see Serif et al. 2017).

If PCR reactions should fail because of larger deletions or insertions, a Southern blot might be helpful, as it reveals shifts of fragments compared to the wild-type DNA. Southern blots together with PCR/sequencing was also the method of choice for a TALEN-based homologous recombination approach by Weyman et al. (2015), which was based on co-transformation of *P. tricornutum* with two plasmids, one containing both TALENs and a plasmid containing a homolog recombination fragment including a resistance cassette for selection of the cell lines. In nearly all diatom

studies that were based on NHEJ repair mechanisms, large insertions have been observed in the resulting cell lines (Daboussi et al. 2014; Nymark et al. 2016; Serif et al. 2017). These insertions mostly consist of vector fragments being inserted at the DSB site, which may be due to the use of the biolistic transformation methods that result in shearing of the plasmid DNA.

Future improvements for controlled genome editing in diatoms

The TALEN and CRISPR/Cas9 systems described in this review have been tested and validated in different laboratories independently. The work of the DiaEdit consortium (Daboussi et al. 2014; Fortunato et al. 2016; Nymark et al. 2016; Serif et al. 2017, unpublished data) along with other studies (Weyman et al. 2015; Hopes et al. 2016; McCarthy et al. 2017; Allorent et al. 2018; Slattery et al. 2018; Stukenberg et al. 2018) clearly show that genome editing in diatom is feasible and efficient. Because of the different approaches used to generate transgenic lines and to screen the mutants, at this stage estimates on editing efficiency in diatoms, compared to other systems, are not meaningful.

As described above, so far genome editing in diatoms has been achieved mostly by integrating the genes encoding the different nucleases into the genome of the cells. This process might have negative impacts on genome stability (by potentially interrupting endogenous genes or disrupting regulatory regions). Moreover, the nucleases can later target additional sites in the genome producing additional off-target mutations. Finally, integration of the transgenes encoding the endonucleases into the nuclear genome results in the creation of diatom cells that will be defined as genetically modified organisms (GMOs), thus representing a limitation for commercial use of these strains. In contrast, non-transgenic edited lines might not be regulated as GMOs depending on the individual countries gene technology legislation. Therefore, in addition to the development of very tightly controlled promoters to regulate the expression of nucleases either integrated into the genomes or located on episomes, the transient delivery of nucleases in the form of mRNAs or proteins might be the most efficient way to limit off-target effects and also the problems related to the use of GMO strains. Recently, gene knockout in the green algae *Chlamydomonas reinhardtii* has been achieved by delivering Cas9/sgRNA ribonucleoprotein (RNP) complexes into the cells (Baek et al. 2016; Shin et al. 2016; Greiner et al. 2017). In the same algal system, Ferenczi et al. (2017) also demonstrated targeted DNA editing and replacements by the Cpf1 ribonucleoproteins, which use single-stranded oligodeoxynucleotides (ssODNs) as repair templates. Recent promising

results indeed indicate that, similarly to *Chlamydomonas*, *P. tricornutum* cells can be efficiently edited with the CRISPR/Cas9 system using the RNP directly (Daboussi et al. unpublished data).

Similar to TALEN (Daboussi et al. 2014; Weyman et al. 2015), CRISPR/Cas9 also appears to be an amenable technology for carrying gene replacement via homologous recombination. Such endonuclease-driven homologous recombination could be important in the future to target transgenes to specific genomic sites and to generate engineered landing sites. The latter could be particularly useful to perform comparative gene expression studies between different mutants by facilitating the introduction of different DNA regions of interest into a characterized and clearly defined genomic environment. Recent evidence suggests that such landing sites could also be created by developing a site-specific recombinase (SSR) approach in diatoms. SSRs are extensively used for genome manipulations in a plethora of organisms (Wirth et al. 2007; Nafissi and Slavcev 2014; Meinke et al. 2016). The SSR integrase (Int) of the HK022 bacteriophage catalyzes the integration (between the phage *attP* and the host *attB* sites) and excision (between the recombinant *attR* and *attL* sites) of the phage into and out of the chromosome of its *Escherichia coli* host by site-specific recombination reactions (Azaro and Landy 2002). It has been shown that Int is active in plants, cyanobacteria, and human cells (Gottfried et al. 2005; Harel-Levi et al. 2008; Melnikov et al. 2009). Recent in silico analysis based on the Int *attB* promiscuity (Kolot et al. 2015) have identified native secondary Int *attB* sites on the chromosomes of *P. tricornutum* that may be used for genome manipulations via Int-catalyzed recombinase-mediated cassette exchange reactions. These sites have been proven to be active in an *E. coli* assay (Kolot, unpublished).

Other approaches to improve genome editing in diatoms include the generation of DNA single-strand breaks or nicks (Wu et al. 2014) by fusing a nickase to a single TALE which may mediate high-efficient gene addition, but with marked reduction of random mutagenesis. In perspective, the improvement of genome-editing technologies may also allow to establish loss of function screens via the generation of CRISPR-Cas9 knockout libraries targeting gene families or the whole genome. This may facilitate the unbiased discovery of novel gene function in diatoms, as shown in human cells (Shalem et al. 2014).

Finally, as the generation of mutants now became feasible, an essential issue for future molecular diatom research is the description and storage of stable diatom mutants, as well as their distribution among scientists. Here, deposition sites as well as info databases will have to be created to avoid repetitive and simultaneous creation of mutants in different labs.

Acknowledgements The authors would like to thank the Gordon and Betty Moore Foundation for generously supporting the project GBMF 4966 “DiaEdit—Development of genetic tools for the establishment of routine genome editing in the marine diatom *Phaeodactylum tricornutum*”. Research on genome editing in AF and MF labs is also supported by the European Assemble plus (Association of European Marine Biological Research Laboratories Expanded, H2020-INFRAIA-1-2016-2017) consortium. Several vectors for *P. tricornutum* genome editing generated in the framework of the DiaEdit consortium are available through Addgene: the CRISPR-Cas9 pKS_diaCas9_sgRNA (Addgene ID: 74923, <https://www.addgene.org/74923/>), the pPtPuc3m_diaCas9_sgRNA for episome-based genome editing (Addgene ID: 109219, <https://www.addgene.org/109219/>) (Winge et al. unpublished), and 8 TALEN Vectors (Addgene ID: 90415–90423; https://www.addgene.org/Peter_Kroth/).

References

- Alipanah L, Rohloff J, Winge P et al (2015) Whole-cell response to nitrogen deprivation in the diatom *Phaeodactylum tricornutum*. *J Exp Bot* 66:6281–6296. <https://doi.org/10.1093/jxb/erv340>
- Allorant G, Guglielmino E, Giustini C, Courtois F (2018) Generation of mutants of nuclear-encoded plastid proteins using CRISPR/Cas9 in the diatom *Phaeodactylum tricornutum*. *Methods Mol Biol* 1829:367–378. https://doi.org/10.1007/978-1-4939-8654-5_24
- Apt KE, Kroth-Pancic PG, Grossman AR (1996) Stable nuclear transformation of the diatom *Phaeodactylum tricornutum*. *Mol Gen Genet* 252:572–579. <https://doi.org/10.1007/BF02172403>
- Apt KE, Zaslavkaia L, Lippmeier JC et al (2002) In vivo characterization of diatom multipartite plastid targeting signals. *J Cell Sci* 115:4061–4069. <https://doi.org/10.1242/jcs.00092>
- Archibald JM (2015) Endosymbiosis and eukaryotic cell evolution. *Curr Biol* 25:R911–R921. <https://doi.org/10.1016/j.cub.2015.07.055>
- Armbrust EV (2009) The life of diatoms in the world’s oceans. *Nature* 459:185–192. <https://doi.org/10.1038/nature08057>
- Azaro MA, Landy A (2002) λ Integrase and the Lambda int family. In: Craig NL, Craigie R, Gellert M, Lambowitz AM (eds) *Mobile DNA II*. ASM Press, Washington, DC, pp 118–148. <https://doi.org/10.1128/9781555817954.ch7>
- Baek K, Kim DH, Jeong J et al (2016) DNA-free two-gene knockout in *Chlamydomonas reinhardtii* via CRISPR-Cas9 ribonucleoproteins. *Sci Rep* 6:30620. <https://doi.org/10.1038/srep30620>
- Basu S, Patil S, Mapleson D et al (2017) Finding a partner in the ocean: molecular and evolutionary bases of the response to sexual cues in a planktonic diatom. *New Phytol* 215:140–156. <https://doi.org/10.1111/nph.14557>
- Beurdeley M, Bietz F, Li J et al (2013) Compact designer TALENs for efficient genome engineering. *Nature Commun* 4:1762. <https://doi.org/10.1038/ncomms2782>
- Bitinaite J, Wah DA, Aggarwal AK, Schildkraut I (1998) FokI dimerization is required for DNA cleavage. *Proc Natl Acad Sci USA* 95:10570–10575. <https://doi.org/10.1073/pnas.95.18.10570>
- Buhmann MT, Poulsen N, Klemm J et al (2014) A tyrosine-rich cell surface protein in the diatom *Amphora coffeaeformis* identified through transcriptome analysis and genetic transformation. *PLoS ONE* 9:e110369. <https://doi.org/10.1371/journal.pone.0110369>
- Buitrago-Flórez FJ, Restrepo S, Riaño-Pachón DM (2014) Identification of transcription factor genes and their correlation with the high diversity of stramenopiles. *PLoS ONE* 9:e111841. <https://doi.org/10.1371/journal.pone.0111841>
- Chen X, Xu F, Zhu C et al (2014) Dual sgRNA-directed gene knockout using CRISPR/Cas9 technology in *Caenorhabditis elegans*. *Sci Rep* 4:7581. <https://doi.org/10.1038/srep07581>

- Christian M, Cermak T, Doyle EL et al (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 186:757–761. <https://doi.org/10.1534/genetics.110.120717>
- Chu L, Ewe D, Río Bártilos C et al (2016) Rapid induction of GFP expression by the nitrate reductase promoter in the diatom *Phaeodactylum tricoratum*. *Peer J* 4:e2344. <https://doi.org/10.7717/peerj.2344>
- Cong L, Ran FA, Cox D et al (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339:819–823. <https://doi.org/10.1126/science.1231143>
- Daboussi F, Leduc S, Maréchal A et al (2014) Genome engineering empowers the diatom *Phaeodactylum tricoratum* for biotechnology. *Nature Commun* 5:3831. <https://doi.org/10.1038/ncomm54831>
- De Riso V, Raniello R, Maumus F et al (2009) Gene silencing in the marine diatom *Phaeodactylum tricoratum*. *Nucleic Acids Res* 37:e96–e96. <https://doi.org/10.1093/nar/gkp448>
- Diner RE, Schwenck SM, McCrow JP et al (2016) Genetic manipulation of competition for nitrate between heterotrophic bacteria and diatoms. *Front Microbiol* 7:880. <https://doi.org/10.3389/fmicb.2016.00880>
- Diner RE, Noddings CM, Lian NC et al (2017) Diatom centromeres suggest a mechanism for nuclear DNA acquisition. *Proc Natl Acad Sci USA* 114:E6015–E6024. <https://doi.org/10.1073/pnas.1700764114>
- Doudna JA, Charpentier E (2014) The new frontier of genome engineering with CRISPR-Cas9. *Science* 346:1258096–1258096. <https://doi.org/10.1126/science.1258096>
- Doyle EL, Booher NJ, Standage DS et al (2012) TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic Acids Res* 40:W117–W122. <https://doi.org/10.1093/nar/gks608>
- Dunahay TG, Jarvis EE, Roessler PG (1995) Genetic transformation of the diatoms *Cyclotella cryptica* and *Navicula saprophila*. *J Phycol* 31:1004–1012. <https://doi.org/10.1111/j.0022-3646.1995.01004.x>
- Engler C, Kandzia R, Marillonnet S (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE* 3:e3647. <https://doi.org/10.1371/journal.pone.0003647>
- Falcatore A, Casotti R, Leblanc C et al (1999) Transformation of nonselectable reporter genes in marine diatoms. *Mar Biotechnol* 1:239–251. <https://doi.org/10.1007/PL00011773>
- Ferenczi A, Pyott DE, Xipnitou A, Molnar A (2017) Efficient targeted DNA editing and replacement in *Chlamydomonas reinhardtii* using Cpf1 ribonucleoproteins and single-stranded DNA. *Proc Natl Acad Sci USA* 114:13567–13572. <https://doi.org/10.1073/pnas.1710597114>
- Fischer H, Robl I, Sumper M, Kröger N (1999) Targeting and covalent modification of cell wall and membrane proteins heterologously expressed in the diatom *Cylindrotheca fusiformis*. *J Phycol* 35:113–120. <https://doi.org/10.1046/j.1529-8817.1999.3510113.x>
- Fortunato AE, Jaubert M, Enomoto G et al (2016) Diatom phytochromes reveal the existence of far-red-light-based sensing in the ocean. *Plant Cell* 28:616–628. <https://doi.org/10.1105/tpc.15.00928>
- Gottfried P, Lotan O, Kolot M et al (2005) Site-specific recombination in Arabidopsis plants promoted by the Integrase protein of coliphage HK022. *Plant Mol Biol* 57:435–444. <https://doi.org/10.1007/s11103-004-0076-7>
- Greiner A, Kelterborn S, Evers H et al (2017) Targeting of photoreceptor genes in *Chlamydomonas reinhardtii* via zinc-finger nucleases and CRISPR/Cas9. *Plant Cell* 29:2498–2518. <https://doi.org/10.1105/tpc.17.00659>
- Haeussler M, Schönig K, Eckert H et al (2016) Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol* 17:148. <https://doi.org/10.1186/s13059-016-1012-2>
- Harel-Levi G, Goltsman J, Tuby CNJH et al (2008) Human genomic site-specific recombination catalyzed by coliphage HK022 integrase. *J Biotechnol* 134:46–54. <https://doi.org/10.1016/j.jbiotec.2008.01.002>
- Hempel F, Bozarth AS, Lindenkamp N et al (2011) Microalgae as bioreactors for bioplastic production. *Microb Cell Fact* 10:81. <https://doi.org/10.1186/1475-2859-10-81>
- Hopes A, Nekrasov V, Kamoun S, Mock T (2016) Editing of the urease gene by CRISPR-Cas in the diatom *Thalassiosira pseudonana*. *Plant Methods* 12:49. <https://doi.org/10.1186/s13007-016-0148-0>
- Huang W, Daboussi F (2017) Genetic and metabolic engineering in diatoms. *Philos Trans R Soc Lond B Biol Sci* 372:20160411. <https://doi.org/10.1098/rstb.2016.0411>
- Ifuku K, Yan D, Miyahara M et al (2015) A stable and efficient nuclear transformation system for the diatom *Chaetoceros gracilis*. *Photosynth Res* 123:203–211. <https://doi.org/10.1007/s11120-014-0048-y>
- Karas BJ, Diner RE, Lefebvre SC et al (2015) Designer diatom episomes delivered by bacterial conjugation. *Nat Commun* 6:6925. <https://doi.org/10.1038/ncomms7925>
- Kirst H, Melis A (2014) The chloroplast signal recognition particle (CpSRP) pathway as a tool to minimize chlorophyll antenna size and maximize photosynthetic productivity. *Biotechnol Adv* 32:66–72. <https://doi.org/10.1016/j.biotechadv.2013.08.018>
- Kolot M, Malchin N, Elias A et al (2015) Site promiscuity of coliphage HK022 integrase as a tool for gene therapy. *Gene Ther* 22:521–527. <https://doi.org/10.1038/gt.2015.9>
- Kroth PG, Wilhelm C, Kottke T (2017) An update on aureochromes: Phylogeny—mechanism—function. *J Plant Physiol* 217:20–26. <https://doi.org/10.1016/j.jplph.2017.06.010>
- Lau JB, Stork S, Moog D et al (2016) Protein-protein interactions indicate composition of a 480 kDa SELMA complex in the second outermost membrane of diatom complex plastids. *Mol Microbiol* 100:76–89. <https://doi.org/10.1111/mmi.13302>
- Lepetit B, Sturm S, Rogato A et al (2013) High light acclimation in the secondary plastids containing diatom *Phaeodactylum tricoratum* is triggered by the redox state of the plastoquinone pool. *Plant Physiol* 161:853–865. <https://doi.org/10.1104/pp.112.207811>
- Lieber MR (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem* 79:181–211. <https://doi.org/10.1146/annurev.biochem.052308.093131>
- Lin Y, Fine EJ, Zheng Z et al (2014) SAPTA: a new design tool for improving TALE nuclease activity. *Nucleic Acids Res* 42:e47. <https://doi.org/10.1093/nar/gkt1363>
- Malviya S, Scalco E, Audic S et al (2016) Insights into global diatom distribution and diversity in the world's ocean. *Proc Natl Acad Sci USA* 113:E1516–E1525. <https://doi.org/10.1073/pnas.1509523113>
- Malzahn A, Lowder L, Qi Y (2017) Plant genome editing with TALEN and CRISPR. *Cell Biosci* 7:21. <https://doi.org/10.1186/s13578-017-0148-4>
- Matthijs M, Fabris M, Broos S et al (2016) Profiling of the early nitrogen stress response in the diatom *Phaeodactylum tricoratum* reveals a novel family of RING-domain transcription factors. *Plant Physiol* 170:489–498. <https://doi.org/10.1104/pp.15.01300>
- Matthijs M, Fabris M, Obata T et al (2017) The transcription factor bZIP14 regulates the TCA cycle in the diatom *Phaeodactylum tricoratum*. *EMBO J* 36:1559–1576. <https://doi.org/10.15252/emboj.201696392>
- McCarthy JK, Smith SR, McCrow JP et al (2017) Nitrate reductase knockout uncouples nitrate transport from nitrate assimilation and drives repartitioning of carbon flux in a model pennate diatom. *Plant Cell* 29:2047–2070. <https://doi.org/10.1105/tpc.16.00910>

- Meinke G, Bohm A, Hauber J et al (2016) Cre recombinase and other tyrosine recombinases. *Chem Rev* 116:12785–12820. <https://doi.org/10.1021/acs.chemrev.6b00077>
- Melnikov O, Zaritsky A, Zarka A et al (2009) Site-specific recombination in the cyanobacterium *Anabaena* sp. strain PCC 7120 catalyzed by the integrase of coliphage HK022. *J Bacteriol* 191:5879–5879. <https://doi.org/10.1128/JB.00813-09>
- Miyagawa A, Okami T, Kira N et al (2009) Research note: high efficiency transformation of the diatom *Phaeodactylum tricornerutum* with a promoter from the diatom *Cylindrotheca fusiformis*. *Phycol Res* 57:142–146. <https://doi.org/10.1111/j.1440-1835.2009.00531.x>
- Miyahara M, Aoi M, Inoue-Kashino N et al (2013) Highly efficient transformation of the diatom *Phaeodactylum tricornerutum* by multi-pulse electroporation. *Biosci Biotechnol Biochem* 77:874–876. <https://doi.org/10.1271/bbb.120936>
- Moscou MJ, Bogdanove AJ (2009) A simple cipher governs DNA recognition by TAL effectors. *Science* 326:1501. <https://doi.org/10.1126/science.1178817>
- Nafissi N, Slavcev R (2014) Bacteriophage recombination systems and biotechnical applications. *Appl Microbiol Biotechnol* 98:2841–2851. <https://doi.org/10.1007/s00253-014-5512-2>
- Niu Y-F, Yang Z-K, Zhang M-H et al (2012) Transformation of diatom *Phaeodactylum tricornerutum* by electroporation and establishment of inducible selection marker. *Biotechniques* 52:1–3. <https://doi.org/10.2144/000113881>
- Nymark M, Sharma AK, Sparstad T et al (2016) A CRISPR/Cas9 system adapted for gene editing in marine algae. *Sci Rep* 6:24951. <https://doi.org/10.1038/srep24951>
- Nymark M, Sharma AK, Hafskjold MCG et al (2017) CRISPR/Cas9 gene editing in the marine diatom *Phaeodactylum tricornerutum*. *BioProtocols* 7:15. <https://doi.org/10.21769/BioProtoc.2442>
- Ohno N, Inoue T, Yamashiki R et al (2012) CO₂-cAMP-responsive cis-elements targeted by a transcription factor with CREB/ATF-like basic zipper domain in the marine diatom *Phaeodactylum tricornerutum*. *Plant Physiol* 158:499–513. <https://doi.org/10.1104/pp.111.190249>
- Poulsen N, Kröger N (2005) A new molecular tool for transgenic diatoms: control of mRNA and protein biosynthesis by an inducible promoter-terminator cassette. *FEBS J* 272:3413–3423. <https://doi.org/10.1111/j.1742-4658.2005.04760.x>
- Poulsen N, Chesley PM, Kröger N (2006) Molecular genetic manipulation of the diatom *Thalassiosira pseudonana* (Bacillariophyceae). *J Phycol* 42:1059–1065. <https://doi.org/10.1111/j.1529-8817.2006.00269.x>
- Rastogi A, Murik O, Bowler C, Tirichine L (2016) PhytoCRISP-Ex: a web-based and stand-alone application to find specific target sequences for CRISPR/CAS editing. *BMC Bioinformatics* 17:261. <https://doi.org/10.1186/s12859-016-1143-1>
- Rayko E, Maumus F, Maheswari U et al (2010) Transcription factor families inferred from genome sequences of photosynthetic stramenopiles. *New Phytol* 188:52–66. <https://doi.org/10.1111/j.1469-8137.2010.03371.x>
- Russo MT, Annunziata R, Sanges R et al (2015) The upstream regulatory sequence of the light harvesting complex Lhcf2 gene of the marine diatom *Phaeodactylum tricornerutum* enhances transcription in an orientation- and distance-independent fashion. *Mar Genom* 24(Pt 1):69–79. <https://doi.org/10.1016/j.margen.2015.06.010>
- Sanjana NE, Cong L, Zhou Y et al (2012) A transcription activator-like effector toolbox for genome engineering. *Nat Protoc* 7:171–192. <https://doi.org/10.1038/nprot.2011.431>
- Santurtún A, Riancho JA, Arozamena J et al (2017) Indel analysis by droplet digital PCR: a sensitive method for DNA mixture detection and chimerism analysis. *Int J Legal Med* 131:67–72. <https://doi.org/10.1007/s00414-016-1422-4>
- Serif M, Lepetit B, Weißert K et al (2017) A fast and reliable strategy to generate TALEN-mediated gene knockouts in the diatom *Phaeodactylum tricornerutum*. *Algal Res* 23:186–195. <https://doi.org/10.1016/j.algal.2017.02.005>
- Shalem O, Sanjana NE, Hartenian E et al (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 343:84–87. <https://doi.org/10.1126/science.1247005>
- Shin S-E, Lim J-M, Koh HG et al (2016) CRISPR/Cas9-induced knockout and knock-in mutations in *Chlamydomonas reinhardtii*. *Sci Rep* 6:27810. <https://doi.org/10.1038/srep27810>
- Shrestha RP, Hildebrand M (2017) Development of a silicon limitation inducible expression system for recombinant protein production in the centric diatoms *Thalassiosira pseudonana* and *Cyclotella cryptica*. *Microb Cell Fact*. <https://doi.org/10.1186/s12934-017-0760-3>
- Siaut M, Heijde M, Mangogna M et al (2007) Molecular toolbox for studying diatom biology in *Phaeodactylum tricornerutum*. *Gene* 406:23–35. <https://doi.org/10.1016/j.gene.2007.05.022>
- Slattery SS, Diamond A, Wang H et al (2018) An expanded plasmid-based genetic toolbox enables Cas9 genome editing and stable maintenance of synthetic pathways in *Phaeodactylum tricornerutum*. *ACS Synth Biol* 7:328–338. <https://doi.org/10.1021/acssynbio.7b00191>
- Smith J, Grizot S, Arnould S et al (2006) A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Res* 34:e149–e149. <https://doi.org/10.1093/nar/gkl720>
- Steinert J, Schiml S, Puchta H (2016) Homology-based double-strand break-induced genome engineering in plants. *Plant Cell Rep* 35:1429–1438. <https://doi.org/10.1007/s00299-016-1981-3>
- Stukenberg D, Zauner S, Dell'Aquila G, Maier UG (2018) Optimizing CRISPR/Cas9 for the diatom *Phaeodactylum tricornerutum*. *Front Plant Sci* 9:740. <https://doi.org/10.3389/fpls.2018.00740>
- Valenzuela J, Mazurie A, Carlson RP et al (2012) Potential role of multiple carbon fixation pathways during lipid accumulation in *Phaeodactylum tricornerutum*. *Biotechnol Biofuels* 5:40. <https://doi.org/10.1186/1754-6834-5-40>
- Weber E, Engler C, Gruetzner R et al (2011) A modular cloning system for standardized assembly of multigene constructs. *PLoS One* 6:e16765. <https://doi.org/10.1371/journal.pone.0016765>
- Weyman PD, Beerli K, Lefebvre SC et al (2015) Inactivation of *Phaeodactylum tricornerutum* urease gene using transcription activator-like effector nuclease-based targeted mutagenesis. *Plant Biotechnol J* 13:460–470. <https://doi.org/10.1111/pbi.12254>
- Wirth D, Gama-Norton L, Riemer P et al (2007) Road to precision: recombinase-based targeting technologies for genome engineering. *Curr Opin Biotechnol* 18:411–419. <https://doi.org/10.1016/j.copbio.2007.07.013>
- Wu Y, Gao T, Wang X et al (2014) TALE nickase mediates high efficient targeted transgene integration at the human multi-copy ribosomal DNA locus. *Biochem Biophys Res Commun* 446:261–266. <https://doi.org/10.1016/j.bbrc.2014.02.099>
- Yoshinaga R, Niwa-Kubota M, Matsui H, Matsuda Y (2014) Characterization of iron-responsive promoters in the marine diatom *Phaeodactylum tricornerutum*. *Mar Genom* 16:55–62. <https://doi.org/10.1016/j.margen.2014.01.005>
- Zhang C, Hu H (2014) High-efficiency nuclear transformation of the diatom *Phaeodactylum tricornerutum* by electroporation. *Mar Genom* 16:63–66. <https://doi.org/10.1016/j.margen.2013.10.003>
- Zischewski J, Fischer R, Bortesi L (2017) Detection of on-target and off-target mutations generated by CRISPR/Cas9 and other sequence-specific nucleases. *Biotechnol Adv* 35:95–104. <https://doi.org/10.1016/j.biotechadv.2016.12.003>
- Zu Y, Tong X, Wang Z et al (2013) TALEN-mediated precise genome modification by homologous recombination in zebrafish. *Nat Methods* 10:329–331. <https://doi.org/10.1038/nmeth.2374>