Renewable and Sustainable Energy Reviews 74 (2017) 1080–1107

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

Genome engineering for breaking barriers in lignocellulosic bioethanol production

Kandasamy Ulaganathan*, Sravanthi Goud, Madhavi Reddy, Ulaganathan Kayalvili

Centre for Plant Molecular Biology, Osmania University, Hyderabad 500007, Telangana, India

ARTICLE INFO

Keywords:
Bioethanol
Lignocellulose
Genome engineering
CRISPR/Cas
TALEN
ZFN
Meganuclease
YOGE

ABSTRACT

Lignocellulosic biomass, though available in massive volumes, is not used for production of bioethanol due to existence of several barriers which escalate the cost of production. Microorganisms possess different proteins associated with different stages of lignocellulosic bioethanol production. Though a large number of such proteins have been identified, their specificities and expression levels are not suitable for lignocellulosic bioethanol production. Additionally, the host organism used for bioethanol production may not be tolerant to temperature, pH and ethanol stresses. Hence, the host organisms and the proteins used for bioethanol production needs to be engineered to suit the conditions for ethanol production. Engineering the host strain and altering specificities of proteins employed for bioethanol production can be achieved by genetic engineering techniques, where the gene of interest is isolated first, manipulated *in vitro* and introduced back into the host organism. Recently, a number of precision genome engineering techniques have been developed which facilitate modification of genes / genomic regions directly in the organism of interest without the need for isolating the genes/genomic regions. These techniques include (a). The bacterial immunity based CRISPR/Cas system, (b). *Xanthomonas* transcription-activator-like effector nuclease based TALEN system, (c). Zinc finger domain based ZFN system, (d). Long region recognizing-nuclease based meganuclease system and (e). Oligonucleotide based YOGE system. Protein engineering studies and whole genome sequencing of bioethanol producing strains have shown that alteration of one or more nucleotides can bring out large changes that facilitate increased production of cellulosic bioethanol. These precision engineering techniques can supplement genetic engineering to bring out alteration in specificities of enzymes and change the host's tolerance to various stress levels by specific alteration of genomic regions. In this review, various methods of genome engineering available and their possible application for breaking barriers in lignocellulosic bioethanol production are discussed.

1. Introduction

Climate change and its potential dangerous consequences have accelerated global efforts on finding a renewable environmental friendly replacement to the fossil fuels. Bioethanol is considered to be a viable alternative but the bioethanol produced now is mostly from cereal starch which is in conflict with feeding human and cattle. Ample quantities of lignocellulosic biomass (60 billion tons) are available from terrestrial plants [1] which are renewable, and can be used without disturbing the food, economy and the environment. Economically viable biofuel production from lignocellulosic biomass is yet to be realized due to the existence of several barriers [2]. Efforts are underway to break the barriers utilizing the unprecedented tools made available by the genomic revolution sweeping biology recently. Precision genome engineering is the latest among the tools contributed by the field of genomics. A closer look at the barriers existing in making lignocellulosic bioethanol production economically viable and the powerful tools available with genome engineering methods prompted us to write this review. The objectives of the review are to discuss the genome engineering tools available and explore the possibilities of application of these tools to break the barriers in lignocellulosic bioethanol production. Though various aspects of lignocellulosic bioethanol production [2–6], genome engineering methods and their potential applications in various branches of biology [7–9], have been

Abbreviations: REM, RNA-guided endonuclease mediated; EEM, Engineered endonuclease mediated; CRISPR/Cas, Clustered regularly interspersed short palindromic repeats/CRISPR associated sequence; TALE, Transcription Activator-like Effectors; TALEN, Transcription Activator-like Effector nuclease; ZFN, Zinc finger nuclease; DSB, double strand break; NHEJ, non-homologous end joining; HR, Homologous recombination; DRT, DNA repair template; RVD, repeat variable diresidues; sgRNA, single guide RNA; PAM, Protospacer Adjacent Motif; NADH, Nicotinamide Adenine Dinucleotide; NADP, Nicotinamide Adenine Dinucleotide Phosphate; NAD, Nicotinamide Adenine Dinucleotide

* Corresponding author.
E-mail address: kulaganathan123@gmail.com (K. Ulaganathan).

http://dx.doi.org/10.1016/j.rser.2017.01.028
Received 12 December 2015; Received in revised form 17 November 2016; Accepted 8 January 2017
1364-0321/

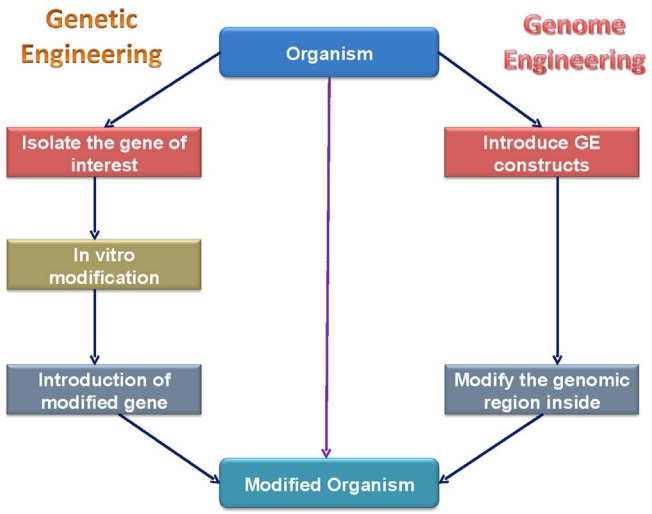

Fig. 1. Genetic Engineering Vs Genome Engineering. Genetic engineering involves isolation of genes, in vitro manipulation and introduction back to the host organism through genetic transformation methods. Genes introduced through genetic engineering remain as plasmids (in microorganisms) or randomly integrated into the host chromosome. Genome engineering involves introduction of genome engineering constructs by genetic transformation and introduction of precision changes into the genome of the host organism directly.

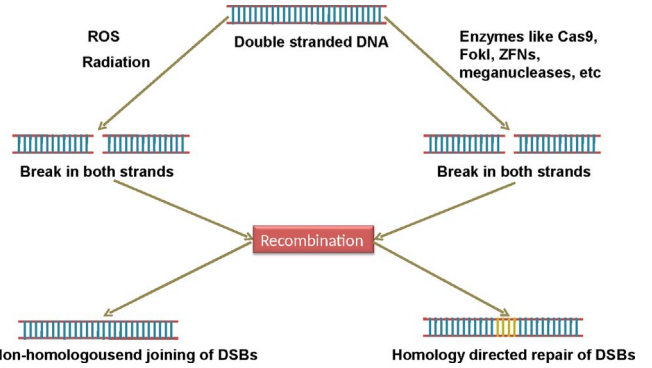

Fig. 2. Schematic representation of double strand break. Double strand break is break in both strands of DNA and double strand breaks are repaired by joining the broken ends of the DSB by non-homologous end joining. When homologous template DNA is available DSB ends are processed and the missing region is copied from the template DNA using homologous recombination.

reviewed extensively, a comprehensive look at the possibilities of utilizing the genome engineering tools for lignocellulosic bioethanol production is lacking. In this review we made an effort to discuss such possibilities to address the barriers associated with lignocellulosic bioethanol production and listed out a large number of potential targets in microorganisms and plants for genome engineering.

2. Genome engineering, the new revolution in genomics

Genome engineering is the process of making precise changes in the native genome to alter the physiology of the organism. It includes gene disruption, deletion and addition in a targeted manner directly inside the organism. This differs from normal genetic engineering where the gene to be engineered is isolated, engineered in *vitro* and reintroduced into the host, or a heterologous gene is introduced by genetic transformation methods to modify the physiology of the organism (Fig. 1). Two types of methods are employed in genome engineering: (a) RNA-guided endonuclease mediated (REM) and (b). Engineered endonuclease mediated (EEM). The bacterial immunity based CRISPR/Cas system [10,11] is the REM based genome engineering method. While the *Xanthomonas* transcription activator like effector nuclease based TALEN system [12], Zinc finger domain based ZFN system [13–15] and the long region recognizing nuclease based Meganuclease system [16,17] are EEM based genome engineering methods.

3. Double strand break in the genome and their repair mechanisms

3.1. Double strand breaks

One of the pre-requisites for genome engineering is the creation of double strand break on specific location of the genome. Originally double strand breaks (DSBs) were identified as errors in the genome caused by internal and external agents. A DSB is a break in both strands of DNA [18,19]. Both endogenous agents like reactive oxygen species and external agents like ionizing radiations can cause double strand breaks (Fig. 2) [20–22]. Double strand breaks also occur during replication [23]. It is also introduced deliberately in the genome by eukaryotes during meiosis [24]. There are different types of DSBs which include blunt DSB, DSBs with variable overhangs, DSBs with base damages and DSBs with sugar damages [25]. As double strand breaks are dangerous to the organism, they are repaired immediately [26].

3.2. Enzymes capable of creating double strand breaks

Both prokaryotes and eukaryotes produce number of proteins capable of inducing double strand breaks in DNA. Among them, Fok1 nuclease, Cas9 protein and Spo11 protein are extensively studied.

3.2.1. FokI nuclease

It is an endonuclease isolated from *Flavobacterium okeanokoites*. It consists of a C-terminal DNA-cleavage domain and an N-terminal DNA-binding domain. FokI recognizes 5′-GGATG-3′, 5′-CATCC-3′ in duplex DNA and cleaves 9 bp away on one strand and 13 bp away on the complementary strand downstream of the recognition site [27]. Though FokI enzyme exists as a monomer with a single catalytic centre, dimerization is essential for catalytic activity [28,29]. After recognition of target by the DNA-binding domain, allosteric alterations lead to cleavage of DNA by the nuclease domain. Binding specificity of this enzyme can be altered by modifying the DNA binding domain [30–32].

3.2.2. Cas9 protein

Cas9 is a protein produced by bacteria as part of its immune system against invading DNA [33,34]. Cas9 protein is made of multiple domains (Fig. 6), a HNH domain, a RuvC domain and an arginine rich domain. The arginine rich domain is involved in nucleic acid binding [35], while HNH and RuvC domains together bring out the double strand break in the incoming foreign DNA. The HNH domain brings out a break in the DNA strand complementary to the guide RNA sequence. The break in the second strand (which is not complementary to guide RNA) is made by the RuvC domain [33,36]. Cas9 protein contains an arginine rich conserved domain involved in nucleic acid binding [37,38] and aids in self and non-self-discrimination [39,40].

3.2.3. Spo11

It is a meiosis-specific protein produced by eukaryotes and is involved in making double strand breaks that initiates meiotic recombination [41]. DSB formation by Spo11 enzyme involves a series of steps and dimerization of Spo11 results in interaction between the active site tyrosine residues of one monomer with metal binding site of another monomer. Nucleophilic attack by the two tyrosine residues on both DNA strands generates a DSB [42]. The DSB ends are further processed by 5′→3′ single-strand resection. The recombination protein