

NUCLEIC ACID SEQUENCES FROM CHLORELLA SAROKINIANA AND USES THEREOF

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is a continuation under 35 U.S.C. § 120 of U.S. application Ser. No. 09/920,953, filed on Aug. 2, 2001 (abandoned), which claims the benefit of U.S. Provisional Application No. 60/224,389, filed Aug. 9, 2000, each of which application is herein incorporated by reference in its entirety, including its sequence listing and table.

FIELD OF THE INVENTION

[0002] The present invention is in the field of molecular biology; more particularly, the present invention relates to nucleic acid sequences from the unicellular green algae, *Chlorella sorokiniana*. The invention encompasses nucleic acid molecules that encode proteins and fragments of proteins. In addition, proteins and fragments of proteins so encoded and antibodies capable of binding the proteins are encompassed by the present invention. The invention also relates to methods of using the disclosed nucleic acid molecules, proteins, fragments of proteins, and antibodies, for example, for gene identification and analysis, and preparation of constructs.

INCORPORATION OF SEQUENCE LISTING

[0003] An electronic equivalent of a paper copy of the Sequence Listing is submitted herewith electronically via EFS web and contains the file named "P30057US02_seqlist.pdf", which is 6,571,155 bytes in size (measured in MS-DOS) and which was created on Feb. 6, 2009. A computer-readable form of the Sequence Listing is also submitted herewith electronically via EFS web and contains the file named "P30057US02_seqlist.txt", which is 6,949,411 bytes in size (measured in MS-DOS) and which was created on Feb. 6, 2009. The sequence listings submitted herewith as electronic equivalents of a paper copy and computer-readable form are herein incorporated by reference.

BACKGROUND OF THE INVENTION

[0004] The present invention relates in part to DNA sequences from cDNA libraries from the unicellular green algae, *Chlorella sorokiniana*. The green algal genus *Chlorella* includes a variety of species (Fott and Novakova, In: *Studies in Phycology: A Monograph of the Genus Chlorella*, Fott, B. (ed.), Prag: Verlag Acad. Sissensch., pp. 10-74 (1969), herein incorporated by reference in its entirety), some of which have long been served as model organisms in plant physiological and biochemical studies (Govindjee and Braun, In: *Algal Physiology and Biochemistry*, W. D. P. Stewart (ed.), University of California Press, Berkeley and Los Angeles, pp. 346-390, herein incorporated by reference in its entirety). *Chlorella* belongs to the eucaryotic cell category of algae and lives in fresh water as a single cell plant. Its size is approximately 2-8 microns in diameter. Species of *Chlorella* have been classified by cell wall sugar composition (Takeda, *Phytochemistry* 27: 3823-6 (1988)) as well as other physiological and biochemical characters (Kessler, *Plant Syst. Evol.* 125:129-38 (1976)). The name *Chlorella* derives from two Latin words meaning 'leaf' (green) and 'small', referring to the unusually high content of chlorophyll which gives

Chlorella its characteristic deep emerald-green color. *Chlorella* is also rich in protein, vitamins, minerals, "C.G.F." (*Chlorella* Growth Factor) and other beneficial substances. Unicellular green algae *Chlorella* are currently being used to produce compounds of commercial value (Behrens et al., *J. Applied Phycology* 6: 113-122 (1994); Running et al., *J. Applied Phycology* 6: 99-104 (1994), both of which are herein incorporated by reference in their entirety).

[0005] It is generally believed that land plants evolved from green algae (Graham, *J. Plant Res.* 109: 241-251 (1996), herein incorporated by reference in its entirety) and that during this revolution, extensive rearrangements occurred within the chloroplast genome. The complete nucleotide sequence of the chloroplast genome (150613 bp) from the unicellular green alga *Chlorella vulgaris*, a species related to *Chlorella sorokiniana*, has been determined (Wakasugi, et al., *Proc. Natl. Acad. Sci. USA* 94:5967-5972 (1997), herein incorporated by reference in its entirety). The chloroplast genome of *Chlorella vulgaris* contains one copy of rRNA gene consisting of 16S, 23S, and 5S rRNA genes; thirty one tRNA gene, sixty-nine protein genes; eight ORFs conserved with those found in land chloroplasts; two adjacent genes homologous to bacterial genes (minD and minE) involved in cell division; genes encoding ribosomal proteins L5, L12, L19 and S9; and two long ORF's related to ycf1 and ycf2 that are exclusively found in land plants (Wakasugi, et al., *Proc. Natl. Acad. Sci. USA* 94: 5967-5972 (1997), herein incorporated by reference in its entirety). *Chlorella* is closer to land plants than the red and brown algae.

[0006] It is advantageous to identify and/or isolate *chlorella* genes for plant genetic engineering to produce plants with agronomically important characteristics or traits. A cDNA (or complementary DNA) library, which is constructed from mRNA purified from *Chlorella* cell culture, can be one valuable source for isolating genes of interest. Construction of cDNA libraries is well-known in the art and a number of cloning strategies exist. Random clones from a cDNA library can be sequenced from both 3' and 5' ends to generate expressed sequence tags (ESTs), which can represent copies of up to the full length transcript (McCombie, et al., *Nature Genetics*, 1: 124-130 (1992); Kurata, et al., *Nature Genetics*, 8: 365-372 (1994); Okubo, et al., *Nature Genetics*, 2: 173-179 (1992)). Typically, only single run sequence data is obtained from the cDNA library (Adams, et al., *Science* 252:1651-1656 (1991)). Automated single run sequencing typically results in an approximately 2-3% error or base ambiguity rate. (Boguski, et al., *Nature Genetics*, 4:332-333 (1993)). Between 150-450 nucleotides of sequence information is usually generated as this is the length of sequence information that is routinely and reliably produced using single run sequence data.

[0007] ESTs have been found to be useful for similarity searches and mapping (Adams, et al., *Science* 252:1651-1656 (1991)). Sequence comparisons and similarity analysis would allow the identification of genes of interest and then full-length cDNA constructs can be obtained using several methods (Land, et al., *Nucleic Acids Res.* 9:2251-2266 (1981); Okayama and Berg, *Mol. Cell. Biol.* 2:161-170 (1982); Coleclough, et al., *Gene* 34:305-314 (1985); Krawinkel, et al., *Nucleic Acids Res.* 14:1913 (1986); Han, et al., *Nucleic Acids Res.* 15:6304 (1987)). Such isolated genes