

Agrigenomics for Microalgal Biofuel Production: An Overview of Various Bioinformatics Resources and Recent Studies to Link OMICS to Bioenergy and Bioeconomy

Namrata Misra,^{1,2} Prasanna Kumar Panda,^{1,2} and Bikram Kumar Parida²

Abstract

Microalgal biofuels offer great promise in contributing to the growing global demand for alternative sources of renewable energy. However, to make algae-based fuels cost competitive with petroleum, lipid production capabilities of microalgae need to improve substantially. Recent progress in algal genomics, in conjunction with other “omic” approaches, has accelerated the ability to identify metabolic pathways and genes that are potential targets in the development of genetically engineered microalgal strains with optimum lipid content. In this review, we summarize the current bioeconomic status of global biofuel feedstocks with particular reference to the role of “omics” in optimizing sustainable biofuel production. We also provide an overview of the various databases and bioinformatics resources available to gain a more complete understanding of lipid metabolism across algal species, along with the recent contributions of “omic” approaches in the metabolic pathway studies for microalgal biofuel production.

Introduction

IN RECENT YEARS, RESEARCH INITIATIVES on renewable bioenergy and biofuels have been gaining momentum, not only due to fast depletion of finite reserves of fossil fuels but also because of the associated concerns for the environment and future energy security. Use of biofuels derived from biomass feedstocks is considered to be promising transport fuels to substitute petroleum products (Ravindranath et al., 2011). According to a recent estimate (IEA, 2012) the world production of biofuels grew from 16 billion liters in 2000 to more than 100 billion liters in 2011, which accounted for 3% of the global transport fuels. In the United States between the period 2009 and 2011, the combined production of two principal liquid biofuels *viz.*, ethanol and biodiesel from non-renewable sources, registered an increase of over 30% from 11454 to 14915 million gallons (EIA, 2012). Presently, biofuels contribute about 10 Mtoe (Million Tonnes of Oil Equivalent) to the European Union road transportation of which 80% is biodiesel, mostly derived from rapeseed, and the remaining is bioethanol obtained from wheat, maize, sugarbeet, and sugarcane (EASAC, 2012). Global bioethanol and biodiesel production are projected to the tune of 180 and 42 billion liters in 2021, respectively. While United States, Brazil, and the European Union are expected to remain as top producers of

ethanol, the European Union is expected to retain its lead position as the largest producer as well as consumer of biodiesel (OECD-FAO, 2012). Domestic policy incentives aiming at energy security and reduction of CO₂ emission, along with mandatory blending targets (varying from 5%–20% of the petroleum fuel) stipulated by many countries are considered to be the driving forces for future higher demand of biofuels. Besides, promotion of biofuel programs by several developing countries to support rural livelihood enhancement programmes, reclamation of waste land, and local energy security are expected to augment the growth of global biofuel industry. Presently, while much of the first-generation biofuel feedstock come from agricultural crops such as maize, sugarcane, sugarbeet, rapeseed, and oil palm, their large-scale adoption for biofuel production is perceived to have adverse impacts on security and agri-biodiversity (Chisti, 2008; Schenk et al., 2008). On the other hand, second-generation biofuel, including lignocellulosic biomass derived from agricultural residues, and non-food dedicated energy crops grown on marginal lands overcome some of the aforementioned limitations of the first-generation crops (Immerzeel et al., 2013). However, the technologies for second-generation biofuels have not fully matured, and despite several successful pilot-scale operations, they are yet to be available commercially (EASAC, 2012).

¹Academy of Scientific and Innovative Research, and ²Bioresources Engineering Department, CSIR-Institute of Minerals and Materials Technology, Bhubaneswar, Odisha, India.

In this context, many recent studies have conclusively advocated the oleaginous fast growing microalgae having multiple advantages over land plants, as the most promising alternative feedstocks for the production of third-, and next-generation biofuels (Chisti, 2007; Lam and Lee, 2012; Mata et al., 2010) (Fig. 1). Algae can be grown on waste land that is unsuitable for cultivation and utilize a wide variety of water sources including waste and saline water, thus mitigating the problem of competition with agricultural resources. Algae efficiently recycle carbon-rich flue emissions and are responsible for more than 40% of global carbon fixation. Algal strains have the potential to accumulate significant amount of neutral lipids, particularly triacylglycerol (TAG), with a magnitude higher than that of potential biofuel crops such as soybean, jatropha, oil palm, and sunflower. In addition to oil accumulation, the rate of microalgal biomass production is quite high as compared to land plants. Furthermore, the low complexity of the cell structure allows easier engineering of algal strains for efficient production of biofuel precursors and other valuable bioactive co-products (Chisti, 2007; Lam and Lee, 2012; Malcata, 2011; Mata et al., 2010; Schenk et al., 2008).

Notwithstanding these many advantages, the relatively high cost of algal biofuels (US \$300–2600 per barrel) compared to petroleum (US \$40–80 per barrel) indicates further improvements in technology are required to make it cost-competitive and commercially scalable (Hannon et al., 2010; Pienkos and Darzins, 2009; Singh et al., 2011; Wijffels and Barbosa, 2010). One of the potential solutions to the above problem is to enhance the lipid production capabilities of microalgae by inducing nutrition-deficient conditions in order to channel metabolic fluxes towards lipid biosynthesis. However, several studies have reported that cultivating algal strains under such controlled stress regimes have resulted in impeded cell growth with low lipid productivity (Courchesne et al., 2009). The recent advancements in molecular biology techniques have led to speculation that lipid accumulation in microalgae can alternatively be augmented without applying the aforementioned stress by appropriately modifying their

genomes through genetic engineering (Radakovits et al., 2010). Significant progress has been made towards the successful overexpression or knock-out of genes from oleaginous microalgal species with high potential for biofuel production (Courchesne et al., 2009; Hu et al., 2008; Radakovits et al., 2010). While such efforts demonstrate the feasibility of genetic engineering in improving algal biofuels, its effectiveness will be dependent on a deep understanding of the target genes and metabolic pathways responsible for lipid accumulation in microalgae (Khozin-Goldberg and Cohen, 2011).

However, compared with higher plants and other eukaryotes, knowledge of lipid biosynthetic pathways in microalgae remains far from complete (Hu et al., 2008). Access to multiple microalgal genome sequences now provides a wealth of opportunities for application of “omic” approaches to unravel algal lipid metabolism and identify gene targets for the development of potentially engineered strains with optimized lipid content (Fig. 2) (Beer et al., 2009; Georgianna and Mayfield, 2012; Mukhopadhyay et al., 2008; Rodriguez-Moya and Gonzalez, 2010; Yu et al., 2011). The term “omics” in biological sciences widely encompasses genomics, transcriptomics, proteomics, and metabolomics which refers to the study of DNA, mRNA, proteins and metabolites, respectively. While “omics” is now indispensable in the search of novel biomarkers for drug and vaccine development (Hu et al., 2011; Ozdemir et al., 2011), its application to microalgae is just beginning to emerge in the last few years. Furthermore, to date, no review attempt has been undertaken to evaluate the successful integration of “omic” approaches in algal biofuel research. This review is mainly focused on the role of “omics” in improving the development of sustainable biofuel production from microalgae. First, a comprehensive overview of the recent advances in algal genomics is presented, followed by a series of bioinformatics resources available for exploring lipid metabolic pathways in microalgae. Some recent studies employing various “omic” approaches for identification and characterization of putative genes responsible for microalgal biofuel production are also discussed.

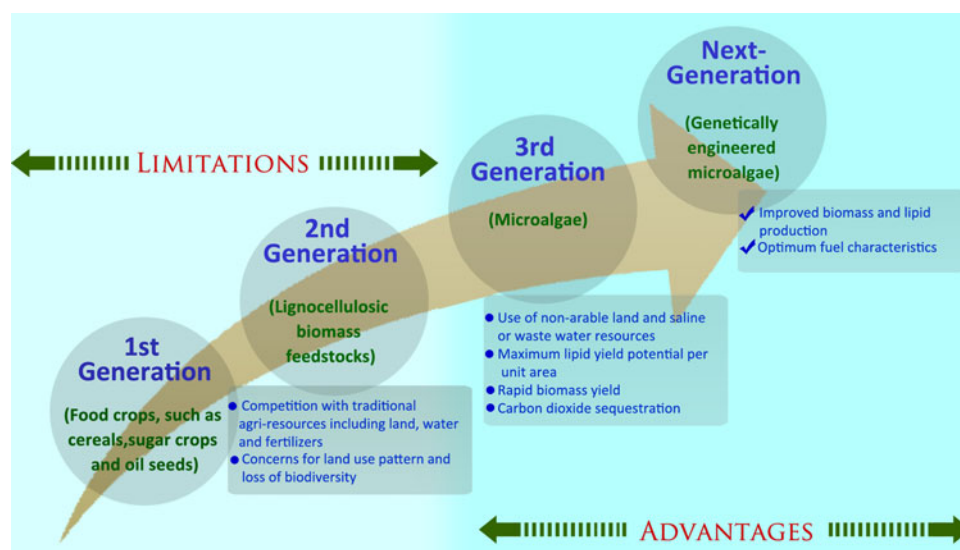


FIG. 1. Schematic representation of the advantages and limitations of different biofuel feedstock derived from renewable resources.

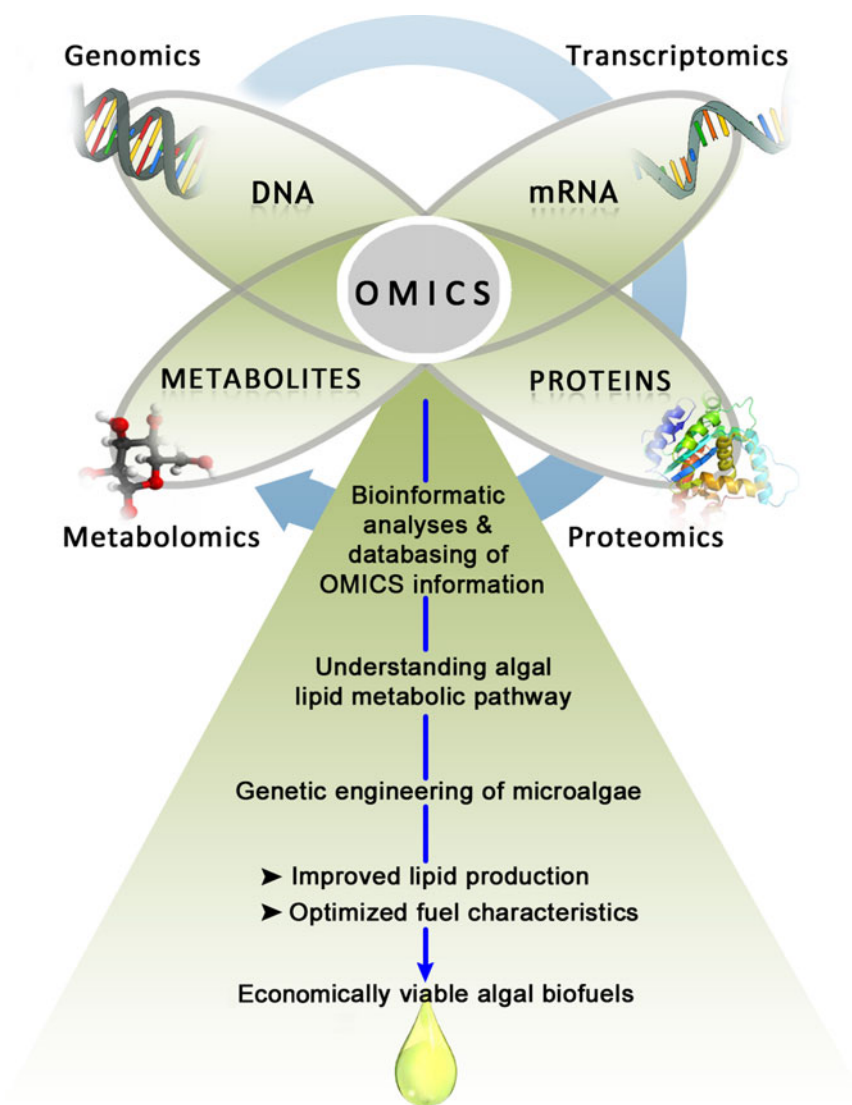


FIG. 2. Schematic overview of the “omic” approaches for improved algal biofuel production.

Update on Sequenced Microalgal Genomes

Significant advances in next-generation sequencing technology have facilitated rapid accumulation of microalgal genomic sequences along with EST (expressed sequence tag) and transcriptome data sets (Grossman, 2005; Radakovits et al., 2010; Tirichine and Bowler, 2011). To date, the whole genome sequences of more than ten microalgae have been generated (Guarnieri et al., 2011; Liu and Benning, 2012; Table 1). These includes the *Cyanidioschyzon merolae* 10D (Matsuzaki et al., 2004), *Phaeodactylum tricornutum* CCP1055/1 (Bowler et al., 2008), *Thalassiosira pseudonana* CCMP1335 (Armbrust et al., 2004), *Guillardia theta* CCMP2712 (Curtis et al., 2012), *Chlamydomonas reinhardtii* CC-503 (Merchant et al., 2007), *Ostreococcus tauri* OTH95 (Derelle et al., 2006), *Ostreococcus lucimarinus* CCE9901 (Palenik et al., 2007), two strains of *Micromonas pusilla*, RCC299 and CCMP1545 (Worden et al., 2009), *Bathycoccus prasinus* RCC1105 (Moreau et al., 2012), *Volvox carteri* UTEX2908 (Prochnik et al., 2010), *Chlorella vulgaris* NC64A (Blanc et al., 2010), *Coccomyxa subellipsoidea*

C-169 (Blanc et al., 2012), *Ectocarpus siliculosus* EC32 (Cock et al., 2010), *Aureococcus anophagefferens* CCMP1984 (Gobler et al., 2011), *Nannochloropsis gaditana* (Radakovits et al., 2012), and *Bigeloviella natans* CCMP2755 (Curtis et al., 2012). Other algal genomes in the sequencing pipeline are *Ostreococcus sp* RCC809, *Botryococcus braunii* Berkeley strain, *Dunaliella salina* CCAP19/18, *Galdieria sulphuraria*, *Chondrus crispus*, *Fragilariopsis cylindrus* CCMP1102, *Pseudo-nitzschia multiseries* CLN-47, *Emiliana huxleyi* CCMP1516 (Radakovits et al., 2010; Tirichine and Bowler, 2011). Several organelle (mitochondria or/and plastid) genomes in microalgae have also been sequenced, including those for *D. salina* CCAP19/18 (Smith et al., 2010), *Botryococcus braunii* (Weiss et al., 2010; 2011), *Nephroselmis olivaceae* (Turmel et al., 1999), *Chaetosphæridium globosum* (Turmel et al., 2002), *Mesostigma viride* (Lemieux et al., 2000), *Cyanophora paradoxa* (Stirewalt et al., 1995), *Cyanidium caldarium* (Glockner et al., 2000), *Gracilaria tenuistipitata* (Hagopian et al., 2004), *Porphyra purpurea* (Reith and Munholland, 1995), and *Odontella sinensis* (Kowallik et al., 1995).

TABLE 1. AN OVERVIEW ON GENOMIC CHARACTERISTICS OF COMPLETE AND ONGOING MICROALGAL SEQUENCING PROJECTS

Class	Organism	Strain	Organelle	Status	Genome size	No of chromosomes	GC (%)	Protein coding genes	Genes with introns (%)	Mean intron length (bp)	Institution/Reference	Web link	
Chlorophytes	<i>Chlamydomonas reinhardtii</i>	CC-503	Nucleus	complete	121Mb	17	64	15,143	92	373	Merchant et al., 2007	http://genome.jgi-psf.org/Chlre3/Chlre3.home.html	
	<i>Volvox carterii</i>	UTEX2908	Nucleus	complete	138 Mb	14	56	14,520	92	358	Prochnik et al., 2010	http://genome.jgi-psf.org/Volca1/Volca1.home.html	
	<i>Chlorella vulgaris</i>	NC64A	Nucleus	complete	46.2Mb	12	67	9,791		209	Blanc et al., 2010	http://genome.jgi-psf.org/ChlNC64A_1/ChlNC64A_1.home.html	
	<i>Coccomyxa subellipsoidea</i>	C-169	Nucleus	complete	48.8Mb	20	53	9,851		5.0	Blanc et al., 2012	http://genome.jgi-psf.org/Chlvu1/Chlvu1.home.html	
Prasinophytes	<i>Dunaliella salina</i>	CCAP19/18	Plastid	complete	269Kb		32.1	34			Smith et al., 2010		
	<i>Botryococcus braunii</i>	Berkeley	Mitochondria	complete	28.3Kb		34.4	42			Joint Genome Institute		
	<i>Nephrolepis olivaceae</i>	Yamanaka	Nucleus	ongoing	300Mb		54.4				Weiss et al., 2010		
	<i>Mesostigma viride</i>	Songkla	Nucleus	ongoing	166.2 (±2.2)Mb						Weiss et al., 2011		
	<i>Ostreococcus</i>	CCE9901	Nucleus	complete	166.0 (±.4)Mb	21	42.1	127			Turmel et al., 1999	http://genome.jgi-psf.org/Ost9901_3/Ost9901_3.home.html	
	<i>lucinarius</i>	OTH95	Plastid	complete	211.3 (±1.7)Mb		30.1	135			Lemieux et al., 2000		
	<i>Ostreococcus tauri</i>	RCC809	Nucleus	complete	200.7Kb	20	43	7,651			Palenik et al., 2007	http://genome.jgi-psf.org/Ostta4/Ostta4.home.html	
	<i>Ostreococcus sp</i>	RCC299	Nucleus	complete	118.3Kb		58	7,892	39		Derelle et al., 2006		
	<i>Micromonas pusilla</i>	CCMP1545	Nucleus	complete	13.6Mb	17	64	10,056	2.2		Joint Genome Institute	http://genome.jgi-psf.org/OstRCC809_2/OstRCC809_2.home.html	
	<i>Bathycoccus prasinos</i>	RCC1105	Nucleus	complete	20.9Mb	19	65	10,575	2.1		Worden et al., 2009	http://genome.jgi-psf.org/MicpuN3/MicpuN3.home.html	
Rhodophytes	<i>Cyanidioschyzon merolae</i>	10D	Nucleus	complete	15Mb	19	48	7,847			Moreau et al., 2012	http://merolae.biol.s.u-tokyo.ac.jp/	
	<i>Galdieria sulphuraria</i>		Nucleus	complete	16.5Mb	20	55	5,331	0.5	248	Matsuzaki et al., 2004	http://genomics.msu.edu/galdieria/	
	<i>Chondrus crispus</i>		Nucleus	ongoing	15Mb						Genoscope	http://www.genoscope.cns.fr/spip/	
	<i>Cyanidium caldarium</i>		Plastid	complete	164.9Kb		32.7	232			Glockner et al., 2000		
	<i>Gracilaria tenuistipitata</i>		Plastid	complete	183.8Kb		29.1	231			Hagopian et al., 2004		
	<i>Porphyra purpurea</i>		Plastid	complete	191.0Kb		33	251			Reith and Munholland, 1995		
	<i>Phaeodactylum tricornutum</i>	CCP1055/1	Nucleus	complete	27.4Mb		49	10,402	47		123	Bowler et al., 2008	http://genome.jgi-psf.org/Phatr2/Phatr2.home.html
	<i>Thalassiosira pseudonana</i>	CCMP1335	Nucleus	complete	34.5	24	47	11,242			992	Armbrust et al., 2004	http://genome.jgi-psf.org/thaps1/thaps1.home.html
	<i>Fragilaropsis cylindrus</i>	CCMP1102	Nucleus	ongoing	81Mb							Joint Genome Institute	http://genome.jgi-psf.org/Fracy1/Fracy1.home.html
	<i>Pseudo-nitzschia multiseries</i>	CLN-47	Nucleus	ongoing	218Mb							Joint Genome Institute	http://genome.jgi-psf.org/Psemu1/Psemu1.home.html
Cryptophytes	<i>Ectocarpus siliculosus</i>	EC32	Nucleus	complete	195.8Mb		53.6	16,256		703.8	Cock et al., 2010	http://genome.jgi-doe.gov/Auran1/Auran1.home.html	
	<i>Aureococcus anophagefferens</i>	CCMP1984	Nucleus	complete	57Mb			11,501			Gobler et al., 2011	http://nannochloropsis.genomeproject.solutions-databases.com	
	<i>Nannochloropsis gaditana</i>		Nucleus	complete	29Mb		54.2	9,052		1.62	Radakovits et al., 2012		
	<i>Odontella sinensis</i>		Plastid	complete	119.7Kb		31.8	178			Kowallik et al., 1995		
	<i>Guillardia theta</i>	CCMP2712	Nucleus	complete	87.2Mb		53	24,840	80		110	Curtis et al., 2012	
	<i>Bigelovella natans</i>	CCMP2755	Nucleus	complete	94.7Mb		45	21,708	86		184	Joint Genome Institute	http://genome.jgi-psf.org/Emihu1/Emihu1.home.html
Haptophytes	<i>Emiliana huxleyi</i>	CCMP1516	Nucleus	ongoing	167.7Mb								
	<i>Cyanophora paradoxa</i>		Plastid	complete	135.5Kb		30.4	192				Stirewalt et al., 1995	

TABLE 2. NUMBER OF ESTs AVAILABLE FOR VARIOUS MICROALGAL SPECIES

Class	Organism	No of ESTs	Reference
Chlorophytes	<i>Chlamydomonas incerta</i>	5,124	O'Brien et al., 2007
	<i>Mesostigma viride</i>	5,615	O'Brien et al., 2007
	<i>Nephroselmis olivaceae</i>	126	O'Brien et al., 2007
	<i>Scenedesmus obliquus</i>	6,615	O'Brien et al., 2007
	<i>Acetabularia acetabulum</i>	941	Henry et al., 2004
Stramenopiles	<i>Laminaria digitata</i>	500	Crepineau et al., 2000
	<i>Phaeodactylum tricornutum</i>	130,000	Maheswari et al., 2005; 2008
	<i>Thalassiosira pseudonana</i>	77,000	Maheswari et al., 2005; 2008
Rhodophytes	<i>Porphyra yezoensis</i>	10,154	Nikaido et al., 2000
	<i>Galdieria sulphuraria</i>	5,270	Weber et al., 2004
	<i>Gracilaria gracilis</i>	200	Lluisma and Ragan, 1997
Glaucophytes	<i>Cyanophora paradoxa</i>	9,867	O'Brien et al., 2007
	<i>Glaucocystis nostochinearum</i>	8,475	O'Brien et al., 2007
Haptophytes	<i>Emiliania huxleyi</i>	3,000	Wahlund et al., 2004
	<i>Pavlova lutheri</i>	7,590	O'Brien et al., 2007
	<i>Isochrysis galbana</i>	12,205	O'Brien et al., 2007
Chlorachniophytes	<i>Bigelowiella natans</i>	3,472	Archibald et al., 2003
Dinoflagellates	<i>Amphidinium carterae</i>	2,143	Bachvaroff et al., 2004
	<i>Lingulodinium polyedrum</i>	1,012	Hackett et al., 2005
	<i>Alexandrium tamarense</i>	6,723	Hackett et al., 2005

In addition to the assembled genomic sequences, a cornucopia of ESTs is currently available in public databases to explore the functional characterization of microalgal genomes (Table 2). ESTs represent tags of the expressed region of a genome and are extensively used for identification of protein coding genes, prediction of gene structure, alternate splicing analysis, and characterization of putative SNPs (Nagaraj et al., 2006). There are a number of public resources that store the sequenced ESTs from various algal species. A recently developed TBestDB provides complete information on ESTs obtained from the genomes of *Chlamydomonas incerta*, *Mesostigma viride*, *Nephroselmis olivaceae*, *Cyanophora paradoxa*, *Glaucocystis nostochinearum*, *Pavlova lutheri*, *Isochrysis galbana*, and *Scenedesmus obliquus* (O'Brien et al., 2007). Another EST database portal dedicated only for diatom species presently contains 130,000 and 77,000 ESTs (as of November 2008) for *P. tricornutum* and *T. pseudonana*, respectively (Maheswari et al., 2005; 2008). In addition to these databases, ESTs for several other microalgae are freely available in the NCBI (National Center for Biotechnology Information) (<http://www.ncbi.nlm.nih.gov/dbEST>) (Archibald et al., 2003; Bachvaroff et al., 2004; Crepineau et al., 2000; Hackett et al., 2005; Henry et al., 2004; Lluisma and Ragan, 1997; Nikaido et al., 2000; Scala et al., 2002; Wahlund et al., 2004; Weber et al., 2004).

It is pertinent to note here that, despite the aforementioned progress in algal genomics, the rationale for sequencing these algal strains was not in the context of their lipid synthesis and hence, there still remains a paucity of biofuel-relevant genomic information (Blankenship, 2010; Rismani-Yazdi et al., 2011). The recently reported endeavours in algal transcriptome profiling have successfully demonstrated the feasibility of bypassing the bottleneck of genome sequencing and have improved our understanding of lipid metabolism in some unsequenced oleaginous microalgal species (Guarnieri et al., 2011; Rismani-Yazdi et al., 2011). Transcriptomics is an excellent tool to quantify the relative abundance of mRNA levels in a single cell or a population of cells. Unlike a genomic approach, it offers a more exhaustive view of the genes that

are being actively expressed in response to particular environmental conditions and thus can be employed to identify putative gene targets that can be engineered to augment lipid content in microalgae (Liu and Benning, 2012; Zhang et al., 2010). In this regard, the first *de novo* transcriptomics study on a potential biofuel algal feedstock, *D. tertiolecta*, revealed a repertoire of enzymes involved in the biosynthesis and catabolism of fatty acids, TAG, and starch (Rismani-Yazdi et al., 2011). Furthermore, the analyses revealed that the reconstructed metabolic pathways of this particular organism were similar to *C. reinhardtii* (Hu et al., 2008; Riekhof et al., 2005) and higher plants (Durrett et al., 2008), and demonstrated its inherent genetic ability of linking starch metabolism with ethanol fermentation through the glycolysis pathway.

Subsequently, the same approach was employed to unravel the TAG biosynthetic pathway in one more unsequenced oleaginous microalga, *C. variabilis* UTEX395 under Nitrogen-replete and -deplete conditions (Guarnieri et al., 2011). In addition to these studies, a plethora of transcriptome information is readily available for the model alga, *C. reinhardtii* under various genetic and physiological stress (Boyle et al., 2012; Castruita et al., 2011; Fang et al., 2012; Gonzalez-Ballester et al., 2010; Miller et al., 2010). For example, Boyle et al. (2012) conducted genome-wide expression analysis to explore the molecular mechanisms underlying the induction of TAG accumulation in *C. reinhardtii*. They identified three vital acyltransferase genes including DGAT1, DGATT1 and PDAT1 to be likely responsible for TAG accumulation. Interestingly, they also observed that following N-deprivation, *Chlamydomonas* cells switch their metabolism from converting acetate to glucose into a more direct incorporation of acetate into fatty acids by down-regulating glyoxylate cycle activity and gluconeogenesis. Likewise, Miller et al. (2010) confirmed the direct involvement of NRR1 and PDAT genes in nitrogen assimilation and TAG accumulation in *C. reinhardtii*. Taken together, these analyses corroborate the significance of transcriptome

sequencing as an emerging powerful approach for identification of genes responsible for algal biofuel production.

Databases and Other Bioinformatics Resources for Studying Lipid Metabolic Pathways in Microalgae

The growing compendium of genome sequences for an increasing number of organisms have resulted in a concomitant need for the development of bioinformatics resources that will not only serve as a knowledge base of various metabolic pathways but also facilitate comparative genomic analysis for functional annotation and interpretation of newly identified genes at proteomic and metabolic levels. Metabolic pathways represent a programmed sequence of molecular events underlying a specific biological activity. As discussed earlier, elucidation of biological pathways, in particular lipid metabolic pathways is useful for genetic engineering efforts directed towards augmenting lipid accumulation in microalgae (Georgianna and Mayfield, 2012; Khozin-Goldberg and Cohen, 2011; Yu et al., 2011). While several such databases specific for lipid pathway analyses have been constructed for plants (Beisson et al., 2003; Caspi et al., 2008; Child et al., 2012; Duvick et al., 2007; Mao et al., 2009; Mekhedov et al., 2000; Sucaet and Deva, 2011), very few to date are available for the study of microalgae (Lopez et al., 2011; May et al., 2009). It is likely that the dearth of algal genomic data as compared to higher plants has contributed to the present status of limited pathway databases and web resources. However, integration of plant lipid pathway databases in comparative genomic analyses might reveal the homologous enzymes responsible for lipid biosynthesis and regulation in microalgae (Beisson et al., 2003; Misra et al., 2012). In this section, we provide a brief overview of the various databases and bioinformatics tools with particular reference to explore lipid metabolic pathways in microalgae, along with few selected plant-specific databases relevant for microalgal biofuel research.

KEGG (Kyoto Encyclopedia of Genes and Genomes) (<http://www.genome.jp/kegg/>) is one of the most widely used comprehensive resource of metabolic pathways including for several organisms (Kanehisa et al., 2010). The database also provides reference pathways that serve as a framework for construction of organism-independent biochemical pathways from the user input whole genome. To further enrich the KEGG pathway information, KEGG ortholog prediction tool is offered to categorize orthologous and paralogous gene groups from evolutionarily related organisms. Recent genome-wide studies have employed KEGG pathway database to identify genes and reconstruct major lipid biosynthetic pathways in various oleaginous microalgal species (Hashimoto et al., 2008; Misra et al., 2012; Rismani-Yazdi et al., 2011; Smith et al., 2012). MetaCyc (<http://metacyc.org>) is also a popular database that contains experimentally verified metabolic pathway information for more than 600 organisms including species ranging from lower microorganisms, to plants and human (Caspi et al., 2008). MetaCyc is also suitable as a reference database to enrich functional characterization of newly sequenced algal putative genes involved in lipid biosynthesis.

ChlamyCyc is the only available algae-specific web-based database (<http://chlamycyc.mpimp-golm.mpg.de>), specialized for the comprehensive analyses of metabolic pathways and fundamental cellular processes in *C. reinhardtii* (May et al.,

2009). It was developed as a part of the GoFORSYS (German Systems Biology research initiative) (<http://www.goforsys.de>) project and currently provides curated experimental information on a total of 253 pathways, 2851 enzymes, 1419 enzymatic reactions, 1416 compounds, and 928 literature citations. The web-portal also houses the known and predicted biochemical pathways from other widely recognized pathway databases such as PlantGDB (Duvick et al., 2007), PlnTFDB (Riano-Pachon et al., 2007), Quantprime (Arvidsson et al., 2008), ProMEX (Hummel et al., 2007), and MapMan (Usadel et al., 2005). To facilitate phylogenomic analyses, Inparanoid (Remm et al., 2001) and OrthoMCL-DB (Chen et al., 2006) databases are also integrated into the tool. Further, to contribute towards the functional annotation of uncharacterized genes and gene products, a web version of the standard BLAST software (Altschul et al., 1990) is provided within the ChlamyCyc.

Another web-based analysis suite known as Algal Functional Annotation Tool (<http://pathways.mcdb.ucla.edu>) has been recently developed for functional annotation of multiple algal genome sequences (Lopez et al., 2011). Currently the tool provides annotation for *C. reinhardtii*, the model green microalgae, with a provision to include additional algal genomes. Annotation includes assigning pathways, ontology, and protein family terms to the predicted proteins by integrating several bioinformatics resources including KEGG (Kanehisa et al., 2010), MetaCyc (Caspi et al., 2008), Pfam (Finn et al., 2010), Reactome (Matthews et al., 2009), Panther (Thomas et al., 2003), Gene Ontology (Ashburner et al., 2000), InterPro (Hunter et al., 2009), MapMan Ontology (Thimm et al., 2004), and KOG (Tatusov et al., 2003). High-throughput gene expression data for various environmental conditions is also provided along with an integrated search tool to identify the functionally related genes. Additionally, the tool enables dynamic visualization of genes on KEGG pathway maps.

Although research over the past few years has led to significant advancement in our understanding of lipid metabolism in higher plants, much remains to be learned about these processes in microalgae. Nevertheless, as the core metabolic pathways are presumed to be conserved in microalgae and higher plant species, particularly in the model plant *Arabidopsis thaliana* (Hu et al., 2008), the homologous sequences and the corresponding pathways responsible for lipid accumulation in microalgae can be accurately inferred to the greatest extent using the metabolic pathway information of plants as reference. Towards this end, the Arabidopsis Lipid Gene (ALG) database (<http://www.plantbiology.msu.edu/lipids/genesurvey/index.html>) that provides a comprehensive catalog of the *Arabidopsis* lipid biosynthetic genes will be highly beneficial (Beisson et al., 2003). The ALG database is an updated version of the previously determined repository of plant lipid biosynthetic genes (<http://www.canr.msu.edu/lgc>) (Mekhedov et al., 2000), and currently includes 600 encoded proteins that have been classified according to biological function, subcellular localization, and alternate splicing. The database uses TargetP (Emanuelsson et al., 2000) for subcellular localization prediction. Additionally, information of on 3700 ESTs stored in the database will certainly assist in designing gene silencing or disruption experiments. The potential of ALG database has been successfully demonstrated in several comparative genomic studies where the putative genes involved in lipid biosynthesis in *Arabidopsis* were used to identify corresponding homologous genes involved in TAG

biosynthesis in several oleaginous microalgal and plant species for biofuel implications (Li et al., 2010; Misra et al., 2012; Sharma and Chauhan, 2012). Other databases developed to facilitate genome-wide investigations in plant biofuel feedstock species are the BFG (Biofuel Feedstock Genomic Resource) and pDAWG. BFG (http://bfgr.plantbiology.msu.edu) provides high-quality uniform and integrated functional annotation of genes as well as transcripts sequences from 54 lignocellulosic biofuel feedstock species (Childs et al., 2012). pDAWG (http://csbl1.bmb.uga.edu/pDAWG/) is a prototype of a one-stop shop database that offers complete information on plant cell wall genes/proteins, notably information related to phylogenomic, sequence and structure-function (Mao et al., 2009). Presently the database comprises complete genomes of 7 higher plants and 12 microalgal species, including 6 green algae (*O. lucimarinus*, *O. tauri*, *C. reinhardtii*, *V. carteri*, *M. pusilla* strain CCMP1545, and *M. pusilla* strain RCC299), 2 diatoms (*T. pseudonana* and *P. tricornutum*), 2 Stramenopiles (*Phytophthora ramorum* and *P. sojae*), brown tide algae (*A. anophagefferens*), and red algae (*C. merolae*). The above databases, although designed to expedite the process of functional annotation of genes putatively involved in plant lipid biosynthesis, additionally provide a platform for performing comparative genomic analyses in order to identify corresponding homologous genes of algal lipid metabolism. Further, they also serve as ideal templates to develop and enrich the relatively less explored algae lipid pathway databases and web resources.

Recent Application of “Omics” to Identify Genes and Metabolic Pathways Involved in Microalgal Biofuel Production

The global synthesis pathway of TAG begins with the basic fatty acid precursor, acetyl-CoA, and continues through fatty acid biosynthesis, complex lipid assembly, and saturated fatty acid modification processes until TAG bodies are finally

formed. A simplified overview of TAG biosynthetic pathway is shown as Figure 3. Acetyl-CoA carboxylase (ACC) is the initial rate-limiting enzyme that catalyzes carboxylation of acetyl-CoA to produce malonyl-CoA in both eukaryotes and prokaryotes (Sasaki and Nagano, 2004). Two physically distinct types of ACC enzymes have been reported to occur in nature. Plastidic heteromeric ACC composed of four subunits, including biotin carboxyl carrier protein (BCCP), biotin carboxylase (BC), and α , β monomers of carboxyltransferase (CT), is common among prokaryotes. In contrast, the eukaryotic cytosolic homomeric ACC is composed of a single polypeptide with each of these subunits functioning as distinct domains (Cronan and Waldrop, 2002; Huerlimann and Heimann, 2012). Once malonyl-CoA is synthesized, the malonyl group is transferred by malonyl-CoA:ACP transacylase to the acyl carrier protein (ACP) of the fatty acid synthase (FAS) multienzyme complex in order to form malonyl-ACP. Fatty acid biosynthesis is catalyzed by two types of FAS. The type I FAS as found in vertebrates, yeast, and some bacteria contain all the active sites on one or two multifunctional polypeptide, and in type II FAS of many bacteria and plants, the active site resides in discrete gene products. The β -ketoacyl-ACP synthase (KAS) components of the type II FAS enzyme complex carry out the condensation steps of the process. FabH (KAS III) catalyzes the first elongation step of fatty acid biosynthetic process using acetyl-CoA and malonyl-ACP as substrates to produce acetoacetyl-ACP. The cycle is completed by the sequential action of β -ketoacyl-ACP reductase (KAR), β -hydroxyacyl-ACP dehydratases (HAD), and enoyl-ACP reductase (ENR) to form acyl-ACP. The subsequent two carbon condensation steps are further carried over by FabB (KAS I) and FabF (KAS II), producing palmitic (C16:0) or a stearic acid (C18:0) as the final products (Gonzalez-Mellado et al., 2010; Li et al., 2009).

Completion of the *de novo* fatty acid synthesis is accomplished in one of three ways. Either the newly synthesized

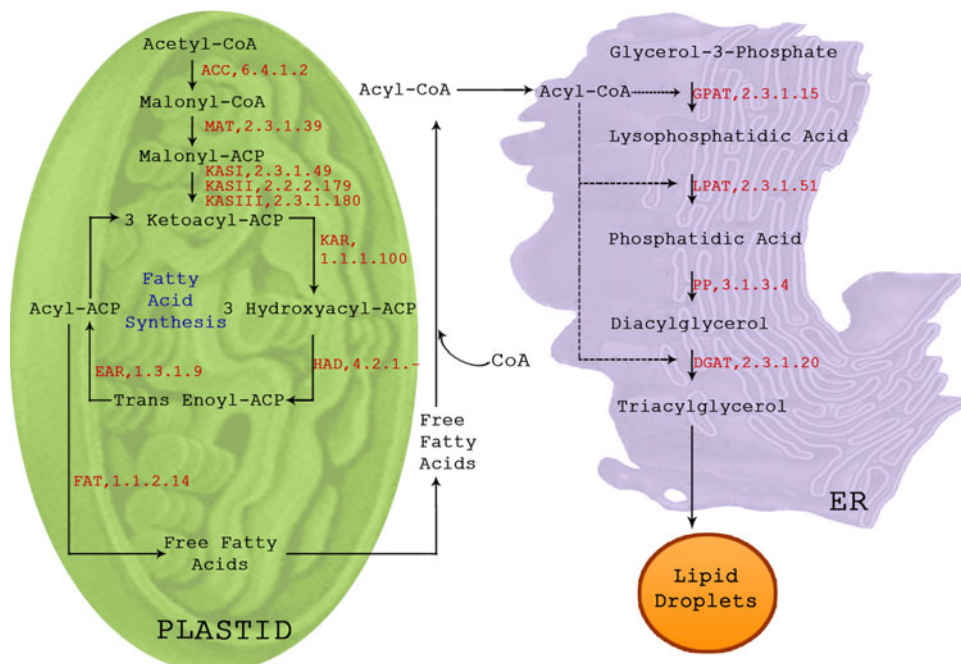


FIG. 3. Schematic overview of triacylglycerol biosynthetic pathway in microalgae.

TABLE 3. RECENT APPLICATION OF INTEGRATED “OMICS” APPROACHES TO IDENTIFY GENES INVOLVED IN LIPID ACCUMULATION FOR MICROALGAL BIOFUEL PRODUCTION

Organism	Objective	Key results	Reference
<i>Nannochloropsis gaditana</i>	To present a draft genome sequence and a genetic transformation method for the marine microalga <i>N. gaditana</i>	Genes required for glycerolipid biogenesis were identified in addition to genes specific to <i>N. gaditana</i> genome and those conserved across stramenopile lineage.	Radakovits et al., 2012
<i>Chlamydomonas reinhardtii</i>	To annotate glycerolipid biosynthesis genes based on similarity to homologs from <i>Arabidopsis thaliana</i>	Reconstruction of the major glycerolipid biosynthetic pathways. A novel BTA1c gene encoding a <i>Chlamydomonas</i> Betaine lipid synthase was identified and isolated.	Riekhof et al., 2005
<i>C. reinhardtii</i>	To analyze purified oil bodies by using proteomics approach	LC-MS/MS analysis identified 248 proteins of which 33 were putatively involved in lipid metabolism, particularly acyl- lipids and sterol. Besides the gene already known to regulate TAG biosynthesis in <i>C. reinhardtii</i> , 19 novel genes were predicted including acyltransferases, hydrolase/lipases, lipid trafficking proteins and proteins of sterol/ergosterol metabolism.	Nguyen et al., 2011
<i>Cyanidioschyzon merolae</i>	To study lipid biosynthetic pathway by genomic and biochemical analyses	Genes involved in the synthesis of fatty acids and lipids were identified. Genomic analysis indicated that <i>C. merolae</i> possess a standard pathway of biosynthesis of glycerolipids except for the absence of plant-type galactosyltransferase enzymes, which underlines the marked difference in lipid biosynthesis pathway between red algae and the green lineage.	Sato and Moriyama, 2007
<i>Haematococcus pluvialis</i>	To study the correlation between fatty acid synthesis and gene expression patterns under different stress conditions	Good correlation between ACP, KAS, FATA gene expression and fatty acid synthesis was observed. In response to stress treatment, the fatty acid synthesis was upregulated while there was no change in the quality of the synthesized fatty acids.	Lei et al., 2012
Diatom species including <i>Thalassiosira pseudonana</i> , <i>Phaeodactylum tricornutum</i> and <i>Fragilariopsis cylindrus</i>	To perform comprehensive <i>in silico</i> evaluation of the genes underlying the carbon metabolism pathways of diatoms	The study identified a total of 164 genes including 51 genes from <i>T. pseudonana</i> , 55 from <i>P. tricornutum</i> and 58 from <i>F. cylindrus</i> . Among them, the enzyme PPK was indicated to play a pivotal role in carbon metabolism and thus a potential target for genetic engineering to enhance biofuel production in microalgae. Further, subcellular localization of the putative enzymes demonstrated the lower half and upper half of glycolysis to be conserved in mitochondria and diatoms, respectively. The results highlight the significant role of intracellular trafficking in the regulation of carbon flux. Genes/proteins up- and downregulated were identified. The potential role of carbon fixation pathways in lipid accumulation is also described.	Smith et al., 2012
<i>P. tricornutum</i>	To study gene expression profiles during nutrient-deprivation and lipid-accumulation conditions	The <i>Chlorella</i> genome was observed to encode all crucial ATG genes, which was further corroborated by expression profiles of the putative genes as determined by RT-PCR analysis. However, the receptor protein in cytoplasm-to-vacuole targeting and mitophagy were found to be absent in microalgae.	Valenzuela et al., 2012
<i>Chlorella</i>	To examine desaturases and elongases by phylogenetics and motif structure prediction approaches	The predicted 275 desaturases and 265 elongases homologs were characterized into 4 and 2 distinct subfamilies, respectively. The resulting Hidden Markov Model profile can be applied to other comparative genomic studies for functional annotation of novel genes.	Jiang et al., 2012
Eukaryotic genomes (fifty-six species)	To characterize the fatty acid desaturases on the basis of conserved histidine-rich motifs and phylogenetic profiles	67 desaturase genes were predicted and annotated along with reconstruction of the pathways involved in unsaturated fatty acid biosynthesis. A novel microsomal $\Delta 12$ desaturase was identified.	Hashimoto et al., 2008
Seven microalgal species including <i>C. reinhardtii</i> , <i>T. pseudonana</i> , <i>P. tricornutum</i> , <i>C. merolae</i> , <i>Ostreococcus lucimarinus</i> and <i>O. tauri</i>	To compare transcriptome and proteome of <i>D. tertiolecta</i> grown under N-replete and N-deplete conditions and high salt concentrations.	Several overexpressed genes and enzymes possibly involved in biosynthesis and catabolism of fatty acids, TAG and starch were discovered and the relevant pathways were reconstructed. Additionally, the pathways associated with biofuel precursor production were found to be interlinked to each other.	Chi et al., 2008
<i>Dunaliella tertiolecta</i>	To demonstrate the utility of a transcriptome as a search model for proteomic analysis in an unsequenced microalga	Various genes and gene products involved in TAG biosynthetic pathway were identified.	Rismani-Yazdi et al., 2011
<i>Chlorella vulgaris</i> UTEX 395	To perform comparative genomic analysis including subcellular localization, physicochemical characterization, exon-intron pattern, motif/ domain predictions of 398 putative lipid genes involved in lipid biosynthesis	The results indicated that although each of the algal species maintains the basic genomic repertoire required for lipid biosynthesis, they possess additional lineage-specific gene groups and the core lipid metabolic pathways in all the studied species are carried out by a comparable number of orthologous proteins.	Guarnieri et al., 2011
Five microalgal species including <i>C. reinhardtii</i> , <i>Volvox carteri</i> , <i>O.lucimarinus</i> , <i>O. tauri</i> and <i>C. merolae</i>	To perform sequence-structure analyses of plastid-located glycerol-3-phosphate acyltransferase (GPAT), a vital enzyme of TAG biosynthetic pathway	The phylogenetic tree along with motif and gene structure analyses indicated a close evolutionary relationship between red algae/diatom and green algae/plant lineages. Further subtle variation in amino acids of fatty acyl binding site were identified that might influence the substrate selectivity of GPAT.	Misra et al., 2012
Seven microalgal species including <i>C. reinhardtii</i> , <i>V. carteri</i> , <i>O. lucimarinus</i> , <i>O. tauri</i> , <i>C. merolae</i> , <i>P. tricornutum</i> , <i>T. pseudonana</i>			Misra et al., 2013

fatty acid is hydrolyzed by fatty acyl-ACP thioesterase (FAT) enzymes, further modified by desaturases, or directly transferred to complex lipid formation using plastid acyltransferases (prokaryotic lipid pathway). Free fatty acids thus produced from plastid eventually enters smooth endoplasmic reticulum (ER) for further elongation, modification, or participation in the synthesis of membrane lipids or storage TAGs (eukaryotic lipid pathway). TAG biosynthetic pathway involves sequential esterification of glycerol chain by a class of acyltransferases. The initial esterification at the sn-1 position of glycerol-3-phosphate is catalyzed by glycerol-3-phosphate acyltransferases (GPAT), followed by esterification at the sn-2 position by lysophosphatidic acid acyltransferase (LPAT) to form phosphatidic acid, subsequently dephosphorylated by phosphatidic acid phosphatase (PAP), and finally esterification at the sn-3 position by diacylglycerol acyltransferase (DGAT) to form TAG (for an extensive review of plant and microalgal lipid biosynthesis, refer to Courchesne et al., 2009; Hu et al., 2008; Ohlogge and Browse, 1995).

The use of high-throughput methods to evaluate gene expression and availability of whole genome sequences in public domain has immensely facilitated characterization of genes and enzymes underpinning various crucial metabolic pathways of microalgae. However, the current understanding of the underlying molecular mechanisms responsible for lipid accumulation in microalgae is yet inadequate as compared to higher plants (Hu et al., 2008).

Systems biology represents the ultimate approach for a thorough understanding of metabolic pathways by harmoniously integrating the various “omic” platforms. Instead of focusing on individual genes, proteins, or metabolites one at a time, it takes into account the interconnections between these elements during regulation of a biological activity (Rodriguez-Moya and Gonzalez, 2010). Recently, several studies have successfully implemented “omic” approaches in parallel to identify the differently expressed genes and enzymes underpinning the metabolic pathways that are likely to be involved in algal lipid accumulation (Table 3). These genes and gene products represent the most potential targets for metabolic pathway reconstruction in order to develop engineered microorganisms with desired fuel-grade properties (Guarnieri et al., 2011; Lei et al., 2012; Misra et al., 2012; Nguyen et al., 2011; Radakovits et al., 2012; Riekhof et al., 2005; Rismani-Yazdi et al., 2011; Smith et al., 2012; Valenzuela et al., 2012).

Despite these commendable efforts, it is important to note that many of these studies were focused on the model organism *C. reinhardtii* or on single algal strains that are not potential biofuel feedstock (Fan et al., 2011). Thus, species representing different clades will probably have to be further explored in order to gain a comprehensive understanding of microalgal lipid metabolism. In addition, comparative genome-wide analyses of algal species showing variations in their fatty acid composition and accumulation will unveil the target genes and underpinning evolutionary processes responsible for the different phenotypes. Furthermore, a comparative analysis of lipid biosynthetic pathways in microalgae and higher plants is another promising area that is relatively underexploited. Although the core fundamental pathways are found to be largely conserved, recent molecular and biochemical studies have reported striking dissimilarities regarding the biosynthetic origin of TAG between microalgae and plants (Fan et al., 2011; Liu and Benning, 2012). In view of

these observations, genome data of microalgae is beginning to be mined using phylogenomics approach to determine the variation in gene contents and differential subcellular localization of TAG synthesis between algal species and higher plants such as *Arabidopsis*, thereby resolving several fundamental questions on algal evolution (Misra et al., 2012; 2013; Sato and Moriyama, 2007).

The ever increasing accumulation of genomic data has changed our perspective of the complexity of pathways involved in microalgal biofuel precursor production. Although the TAG biosynthetic pathway still appears to be the principal lipid metabolic process, it is now clear that other carbon fixation pathways such as autophagy comes into play and eventually signifies alternative pathways for lipid accumulation. Whereas considerable information is available on processes involving metabolic fluxes towards lipid accumulation, less is known about the autophagy genes and corresponding enzymes responsible for carbon fixation in microalgae. Only one such analysis has been recently performed for a green microalga, *Chlorella* (Jiang et al., 2012). Hence, a thorough knowledge of these additional routes and their precise role in biofuel precursor production will be useful.

The majority of the omic-based studies undertaken so far have primarily addressed identification of gene targets for improving lipid production in microalgae. Now it is apparent that modification of the fatty acid profile to include more stearic acid (C18:0) and oleic acid (C18:1) is also indispensable for improving the algal-derived biofuel properties (Knothe, 2009). As a fundamental step in their direction, *in silico* studies aimed at prediction of candidate genes whose combination determines the inherent fatty ester composition of microalgae has been reported (Chi et al., 2008; Hashimoto et al., 2008). These studies conclusively highlight the potential utility of the emerging “omic” approaches in providing greater insights into genome structure and lipid metabolism of various microalgal species.

Conclusion

A comprehensive understanding of metabolic pathways that direct carbon flux towards lipid accumulation is central in establishing genetic engineering strategies for optimizing biofuel production in microalgae. Several recent studies have employed omic-based approaches to unravel the fundamental genetic and cellular processes involved in the synthesis of biofuel precursors from diverse species of algae. Despite these efforts, several challenges still remain to be addressed to make algal-derived fuels cost-competitive with petroleum. For instance, a majority of the genomic studies has been reported for microalgal species of low oil content, particularly for the model microalga *C. reinhardtii*. This is perhaps due to the paucity of biofuel-relevant genomic data for other oleaginous microalgae. Therefore, access to more algal genomes is required to facilitate identification of novel genes and gene products in the metabolic pathways that could be appropriately harnessed for optimum biofuel production. Undoubtedly, with the progress in genome sequencing and “omics” technologies, appropriate bioinformatics resources including databases must also be made available to organize, visualize, and logically interpret the large datasets. Integration of complementary “omic” approaches is envisaged to provide the much needed insights

into the algal lipid metabolism for making biofuel production commercially scalable.

Acknowledgments

The authors gratefully acknowledge the valuable suggestions of the reviewers for improvement of the manuscript. This work was partially funded by the Department of Biotechnology, Government of India. N.M. acknowledges the support of Council for Scientific and Industrial Research, India for granting Senior Research Fellowship. The authors also thank Director, CSIR-IMMT for his support and providing laboratory facilities.

Author Disclosure Statement

The authors declare that no conflicting financial interests exist.

References

- Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ. (1990). Basic local alignment search tool. *J Mol Biol* 215, 403–410.
- Archibald JM, Rogers MB, Toop M, Ishida K, and Keeling PJ. (2003). Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid containing alga *Bigelowiella natans*. *Proc Natl Acad Sci USA* 100, 7678–7683.
- Armbrust EV, Berges JA, Bowler C, et al. (2004). The genome of the diatom *Thalassiosira pseudonana*: Ecology, evolution and metabolism. *Science* 306, 79–86.
- Arvidsson S, Kwasniewski M, Riano-Pachon DM, and Mueller-Roeber B. (2008). QuantPrime—A flexible tool for reliable high-throughput primer design for quantitative PCR. *BMC Bioinformatics* 9, 465–480.
- Ashburner M, Ball CA, Blake JA, et al. (2000). Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25, 25–29.
- Bachvaroff TR, Concepcion GT, Rogers CR, Herman EM, and Delwiche CF. (2004). Dinoflagellate expressed sequence tag data indicate massive transfer of chloroplast genes to the nuclear genome. *Protist* 155, 65–78.
- Beer LL, Boyd ES, Peters JW, and Posewitz MC. (2009). Engineering algae for bio hydrogen and biofuel production. *Curr Opin Biotechnol* 20, 264–271.
- Beisson F, Koo AJK, Ruuska S, et al. (2003). Arabidopsis genes involved in acyl lipid metabolism. A 2003 census of the candidates, a study of the distribution of expressed sequence tags in organs, and a web-based database. *Plant Physiol* 132, 682–697.
- Blanc G, Agarkova I, Grimwood J, et al. (2012). The genome of the polar eukaryotic microalga *Coccomyxa subellipsoidea* reveals traits of cold adaptation. *Genome Biol* 13, R39–R51.
- Blanc G, Duncan G, Agarkova I, et al. (2010). The *Chlorella varii* NC64A reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. *Plant Cell* 22, 2943–2955.
- Blankenship RE. (2010). Early evolution of photosynthesis. *Plant Physiol* 154, 434–438.
- Bowler C, Allen AE, Badger JH, et al. (2008). The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature* 465, 239–244.
- Boyle NR, Page MD, Liu B, et al. (2012). Three acyltransferases and a nitrogen responsive regulator are implicated in nitrogen starvation-induced triacylglycerol accumulation in *Chlamydomonas*. *J Biol Chem* 287, 15811–15825.
- Caspi R, Foerster H, Fulcher CA, et al. (2008). The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res* 36, D623–D631.
- Castruita M, Casero D, Karpowicz SJ, et al. (2011). Systems biology approach in *Chlamydomonas* reveals connections between copper nutrition and multiple metabolic steps. *Plant Cell* 23, 1273–1292.
- Chen F, Mackey AJ, Stoeckert CJ Jr, and Roos DS. (2006). OrthoMCL-DB: Querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res* 34, D363–368.
- Chi X, Zhang X, Guan X, Ding L, Li Y, Wang M, et al. (2008). Fatty acid biosynthesis in eukaryotic photosynthetic microalgae: Identification of a microsomal delta 12 desaturase in *Chlamydomonas reinhardtii*. *J Microbiol* 46, 189–201.
- Childs KL, Konganti K, and Buell CR. (2012). The Biofuel Feedstock Genomics Resources: A web-based portal and database to enable functional genomics of plant biofuel feedstock species. Database Article ID bar061, 1–9. doi:10.1093/database/bar061.
- Chisti Y. (2007). Biodiesel from microalgae. *Biotechnol Adv* 25, 294–306.
- Chisti Y. (2008). Biodiesel from microalgae beats bioethanol. *Trends Biotechnol* 26, 126–131.
- Cock JM, Sterck L, Rouze P, et al. (2010). The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature* 465, 617–621.
- Courchesne NM, Parisien A, Wang B, and Lan CQ. (2009). Enhancement of lipid production using biochemical, genetic and transcription factor engineering approaches. *J Biotechnol* 141, 31–41.
- Crepineau F, Roscoe T, Kaas R, Kloareg B, and Boyen C. (2000). Characterisation of complementary DNAs from the expressed sequence tag analysis of life cycle stages of *Laminaria digitata* (Phaeophyceae). *Plant Mol Biol* 43, 503–513.
- Cronan JE, and Waldrop GL. (2002). Multi-subunit acetyl-CoA carboxylase. *Prog Lipid Res* 41, 407–435.
- Curtis BA, Tanifuji G, Burki F, et al. (2012). Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature* 492, 59–65.
- Derelle E, Ferraz C, Rombaut S, et al. (2006). Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proc Natl Acad Sci USA* 103, 11647–11652.
- Durrett TP, Benning C, and Ohlrogge J. (2008). Plant triacylglycerols as feedstocks for the production of biofuels. *Plant J* 54, 593–607.
- Duvick J, Fu A, Muppirala U, et al. (2007). PlantGDB: A resource for comparative plant genomics. *Nucleic Acids Res* 36, D959–D965.
- EASAC. (2012). The current status of biofuels in the European Union, their environmental impacts and future prospects. EASAC policy report 19.
- EIA. (2012). Biofuels Issues and Trends, Washington.
- Emanuelsson O, Nielsen H, Brunak S, and von Heijne G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequences. *J Mol Biol* 300, 1005–1016.
- Fan J, Andre C, and Xu C. (2011). A chloroplast pathway for the de novo biosynthesis of triacylglycerol in *Chlamydomonas reinhardtii*. *FEBS Lett* 585, 1985–1991.
- Fang W, Si Y, Douglass S, et al. (2012). Transcriptome-wide changes in *Chlamydomonas reinhardtii* gene expression regulated by carbon dioxide and the CO₂-concentrating mechanism regulator CIA5/CCM1. *Plant Cell* 24, 1876–1893.
- Finn RD, Mistry J, Tate J, et al. (2010). The Pfam protein families database. *Nucleic Acids Res* 38, D211–D222.
- Georgianna DR, and Mayfield SP. (2012). Exploiting diversity and synthetic biology for the production of algal biofuels. *Nature* 488, 329–335.

- Glockner G, Rosenthal A, and Valentine K. (2000). The structure and gene repertoire of an ancient red algal plastid genome. *J Mol Evol* 51, 382–390.
- Gobler CJ, Berry DL, Dyhrman ST, et al. (2011). Niche of harmful alga *Aureococcus anophagefferens* revealed through ecogenomics. *Proc Natl Acad Sci USA* 108, 4352–4357.
- Gonzalez-Ballester D, Casero D, Cokus S, Pellegrini M, Merchant SS, and Grossmann AR. (2010). RNA-seq analysis of sulfur-deprived *Chlamydomonas* cells reveals aspects of acclimation critical for cell survival. *Plant Cell* 22, 2058–2084.
- Gonzalez-Mellado D, von Wettstein-Knowles P, Garces R, and Martinez-Force E. (2010). The role of β -ketoacyl-Acyl carrier protein synthase III in the condensation steps of fatty acid biosynthesis in sunflower. *Planta* 231, 1277–1289.
- Grossman AR. (2005). Paths towards algal genomics. *Plant Physiol* 137, 410–427.
- Guarnieri MT, Nag A, Smolinski SL, Darzins A, Seibert M, and Pienkos PT. (2011). Examination of triacylglycerol biosynthetic pathways via de novo transcriptomic and proteomic analyses in an unsequenced microalgae. *PLoS ONE* 6, 1–13.
- Hackett JD, Scheetz TE, Yoon HS, et al. (2005). Insights into a dinoflagellate genome through expressed sequence tag analysis. *BMC Genomics* 6, 80–93.
- Hagopian JC, Reis M, Kitajima JP, Bhattacharya D, and de Oliveira MC. (2004). Comparative analysis of the complete plastid genome sequence of the red alga *Gracilaria tenuistipitata* var. *liui* provides insights into the evolution of rhodoplasts and their relationship to other plastids. *J Mol Evol* 59, 464–477.
- Hannon M, Gimpel J, Tran M, Rasala B, and Mayfield S. (2010). Biofuels from algae: Challenges and potential. *Biofuels* 1, 763–784.
- Hashimoto K, Yoshizawa AC, Okuda S, Kuma K, Goto S, and Kanehisa M. (2008). The repertoire of desaturases and elongases reveals fatty acid variations in 56 eukaryotic genomes. *J Lipid Res* 49, 183–191.
- Henry IM, Wilkinson MD, Hernandez JM, Schwarz-Sommer Z, Grotewold E, and Mandoli DF. (2004). Comparison of ESTs from juvenile and adult phases of the giant unicellular green alga *Acetabularia acetabulum*. *BMC Plant Biol* 4, 3–18.
- Hu Q, Sommerfeld M, Jarvis E, et al. (2008). Microalgal triacylglycerols as feedstocks for biofuel production: Perspectives and advances. *Plant J* 54, 621–639.
- Hu ZZ, Huang H, Wu CH, et al. (2011). Omics-based molecular target and biomarker identification. *Methods Mol Biol* 719, 547–571.
- Huerlimann R, and Heimann K. (2012). Comprehensive guide to acetyl-carboxylases in algae. *Crit Rev Biotechnol*, 1–17.
- Hummel J, Niemann M, Wienkoop S, et al. (2007). ProMEX: A mass spectral reference database for proteins and protein phosphorylation sites. *BMC Bioinformatics* 8, 216–224.
- Hunter S, Apweiler R, Attwood TK, et al. (2009). InterPro: The integrative protein signature database. *Nucleic Acids Res* 37, D211–D215.
- IEA. (2012). Tracking Clean Energy Progress: Energy Technology Perspectives 2012 excerpt as IEA input to the Clean Energy Ministerial, Paris, France.
- Jiang Q, Zhao L, Dai J, and Wu Q. (2012). Analysis of autophagy genes in microalgae: *Chlorella* as a potential model to study mechanism of autophagy. *PLoS ONE* 7, 1–16.
- Kanehisa M, Goto S, Furumichi M, Tanabe M, and Hirakawa M. (2010). KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* 38, D355–D360.
- Khozin-Goldberg I, and Cohen, Z. (2011). Unraveling algal lipid metabolism: Recent advances in gene identification. *Biochimie* 93, 91–100.
- Knothe G. (2009). Improving biodiesel fuel properties by modifying fatty ester composition. *Energ Environ Sci* 2, 759–766.
- Kowallik KV, Stoebe B, Schaffran I, Kroth-Pancic P, and Freier U. (1995). The chloroplast genome of a chlorophyll a+c containing alga, *Odontella sinensis*. *Plant Mol Biol Rep* 13, 336–342.
- Lam MK, and Lee KT. (2012). Microalgae biofuels: A critical review of issues, problems and the way forward. *Biotechnol Adv* 30, 673–690.
- Lei A, Chen H, Shen G, Hu Z, Chen L, and Wang J. (2012). Expression of fatty acid synthesis genes and fatty acid accumulation in *Haematococcus pluvialis* under different stressors. *Biotechnol Biofuels* 5, 18–29.
- Lemieux C, Otis C, and Turmel M. (2000). Ancestral chloroplast genome in *Mesostigma viride* reveals an early branch of green plant evolution. *Nature* 403, 649–652.
- Li L, Li H, Li J, et al. (2010). A genome-wide survey of maize lipid-related genes: Candidate genes mining, digital gene expression profiling and co-location with QTL for maize kernel oil. *Sci China Life Sci* 53, 690–700.
- Li MJ, Li AQ, Xia H, et al. (2009). Cloning and sequence analysis of putative type II fatty acid synthase genes from *Arachis hypogaea*. *J Biosci* 34, 227–238.
- Liu, B., and Benning, C. (2012). Lipid metabolism in microalgae distinguishes itself. *Curr Opin Biotechnol* 24, 1–10.
- Lluisma AO, and Ragan MA. (1997). Expressed sequence tags (ESTs) from the marine red alga *Gracilaria gracilis*. *J Appl Phycol* 9, 287–293.
- Lopez D, Casero D, Cokus SJ, Merchant SS, and Pellegrini M. (2011). Algal Functional Annotation Tool: A web-based analysis suite to functionally interpret large gene lists using integrated annotation and expression data. *BMC Bioinformatics* 12, 282–292.
- Maheswari U, Mock T, Armbrust EV, and Bowler C. (2008). Update of the Diatom EST Database: A new tool for digital transcriptomics. *Nucleic Acids Res* 37, D1001–D1005.
- Maheswari U, Montsant A, Goll J, et al. (2005). *Nucleic Acids Res* 33, D344–D347.
- Malcata FX. (2011). Microalgae and biofuels: A promising partnership? *Trends Biotechnol* 29, 542–549.
- Mao F, Yin Y, Zhou F, et al. (2009). pDAWG: An integrated database for plant cell wall genes. *Bioenerg Res* 2, 209–216.
- Mata TM, Martins AA, and Caetano NS. (2010). Microalgae for biodiesel production and other applications: A review. *Renew Sust Energ Rev* 14, 217–232.
- Matsuzaki M, Misumi O, Shin-i T, et al. (2004). Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* 428, 653–657.
- Matthews L, Gopinath G, Gillespie M, et al. (2009). Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res* 37, D619–D622.
- May P, Christian JO, Kempa S, and Walther D. (2009). ChlamyCyc: An integrative systems biology database and web-portal for *Chlamydomonas reinhardtii*. *BMC Genomics* 10, 209–220.
- Mekhedov S, de Ilarduya OM, and Ohlrogge J. (2000). Toward a functional catalog of the plant genome. A survey of genes for lipid biosynthesis. *Plant Physiol* 122, 389–402.
- Merchant SS, Prochnik SE, Vallon O, et al. (2007). The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* 318, 245–250.

- Miller R, Wu G, Deshpande RR, et al. (2010). Changes in transcript abundance in *Chlamydomonas reinhardtii* following nitrogen deprivation predict diversion of metabolism. *Plant Physiol* 154, 1737–1752.
- Misra N, and Panda PK. (2013). In search of actionable targets for agrigenomics and microalgal biofuel production: Sequence-structural diversity studies on algal and higher plants with a focus on GPAT protein. *OMICS J Integr Biol* 17, 173–186.
- Misra N, Panda PK, Parida BK, and Mishra BK. (2012). Phylogenomic study of lipid genes involved in microalgal biofuel production—Candidate gene mining and metabolic pathway analyses. *Evol Bioinform* 8, 545–564.
- Moreau H, Verhelst B, Couloux A, et al. (2012). Gene functionalities and genome structure in *Bathycoccus prasinus* reflect cellular specializations at the base of the green lineage. *Genome Biol* 13, R74–R90.
- Mukhopadhyay A, Redding AM, Rutherford BJ, and Keasling JD. (2008). Importance of systems biology in engineering microbes for biofuel production. *Curr Opin Biotechnol* 19, 228–234.
- Nagaraj SH, Gasser RB, and Ranganathan S. (2006). A hitchhiker's guide to expressed sequence tag (EST) analysis. *Brief Bioinform* 8, 6–21.
- Nguyen HM, Baudet M, Cuine S, et al. (2011). Proteomic profiling of oil bodies isolated from the unicellular green microalga *Chlamydomonas reinhardtii*: With focus on proteins involved in lipid metabolism. *Proteomics* 11, 4266–4273.
- Nikaido I, Asamizu E, Nakajima M, Nakamura Y, Saga N, and Tabata S. (2000). Generation of 10,154 expressed sequence tags from a leafy gametophyte of a marine red alga, *Porphyra yezoensis*. *DNA Res* 7, 223–227.
- O'Brien E, Koski LB, Zhang Y, et al. (2007). TBestDB: A taxonomically broad database of expressed sequence tags (ESTs). *Nucleic Acids Res* 35, D445–D451.
- OECD–FAO. (2012). OECD–FAO Agricultural Outlook 2012–2021, http://dx.doi.org/10.1787/agr_outlook-2012-en (last accessed on 5/21/13).
- Ohlrogge J, and Browse J. (1995). Lipid biosynthesis. *Plant Cell* 7, 957–970.
- Ozdemir V, Pang T, Knoppers BM, et al. (2011). Vaccines of the 21st century and vaccinomics: Data-enabled science meets global health to spark collective action for vaccine innovation. *OMICS J Integr Biol* 15, 523–527.
- Palenik B, Grimwood J, Aerts A, et al. (2007). The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proc Natl Acad Sci USA* 104, 7705–7710.
- Pienkos PT, and Darzins A. (2009). The promise and challenges of microalgal-derived biofuels. *Biofuel Bioprod Bior* 3, 431–440.
- Prochnik SE, Umen J, Nedelcu AM, et al. (2010). Genomic analysis of complexity in the multicellular green alga *Volvox carteri*. *Science* 329, 223–226.
- Radakovits R, Jinkerson RE, Darzins A, and Posewitz MC. (2010). Genetic engineering of algae for enhanced biofuel production. *Eukaryot Cell* 9, 486–501.
- Radakovits R, Jinkerson RE, Fuerstenberg SI, et al. (2012). Draft genome sequence and genetic transformation of the oleaginous alga: *Nannochloropsis gaditana*. *Nat Commun* 3, 686–711.
- Ravindranath NH, Lakshmi CS, Manuvie R, and Blachandra P. (2011). Biofuel production and implications for land use, food production and environment in India. *Energ Policy* 39, 5737–5745.
- Reith M, and Munholland, J. (1995). Complete nucleotide sequence of the *Porphyra purpurea* chloroplast genome. *Plant Mol Biol Rep* 13, 333–335.
- Remm M, Storm CE, and Sonnhammer EL. (2001). Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J Mol Biol* 314, 1041–1052.
- Riano-Pachon DM, Ruzicic S, Dreyer I, and Mueller-Roeber B. (2007). PlnTFDB: An integrative plant transcription factor database. *BMC Bioinformatics* 8, 42–52.
- Riekhof WR, Sears BB, and Benning C. (2005). Annotation of genes involved in glycerolipid biosynthesis in *Chlamydomonas reinhardtii*: Discovery of the betaine lipid synthase BTA1_C. *Eukaryot Cell* 4, 242–252.
- Rismani-Yazdi H, Haznedaroglu BZ, Bibby K, and Peccia J. (2011). Transcriptome sequencing and annotation of the microalgae *Dunaliella tertiolecta*: Pathway description and gene discovery for production of next-generation biofuels. *BMC Genomics* 12, 148–165.
- Rodriguez-Moya, M., and Gonzalez, R. (2010). Systems biology approaches for the microbial production of biofuels. *Biofuels* 1, 291–310.
- Sasaki Y, and Nagano Y. (2004). Plant acetyl-CoA carboxylase: Structure, biosynthesis, regulation and gene manipulation for plant breeding. *Biosci Biotechnol Biochem* 68, 1175–1184.
- Sato N, and Moriyama T. (2007). Genomic and biochemical analysis of lipid biosynthesis in the unicellular rhodophyte *Cyanidioschyzon merolae*: Lack of a plastidic desaturation pathway results in the coupled pathway of galactolipid synthesis. *Eukaryot Cell* 6, 1006–1017.
- Scala S, Carels N, Falcatore A, Chiusano ML, and Bowler C. (2002). Genome properties of the diatom *Phaeodactylum tri-cornutum*. *Plant Physiol* 129, 993–1002.
- Schenk PM, Thomas-Hall SR, Stephens E, et al. (2008). Second generation biofuels: High-efficiency microalgae for biodiesel production. *Bioenergy Res* 1, 20–43.
- Sharma A, and Chauhan R. (2012). *In Silico* identification and comparative genomics of candidate genes involved in biosynthesis and accumulation of seed oil in plants. *Comp Funct Genom Article ID* 914843, 1–14.
- Singh A, Nigam PS, and Murphy JD. (2011). Mechanism and challenges in commercialisation of algal biofuels. *Bioresour Technol* 102, 26–34.
- Smith DR, Lee RW, Cushman JC, Magnuson JK, Tran D, and Polle JEW. (2010). The *Dunaliella salina* organelle genomes: Large sequences, inflated with intronic and intergenic DNA. *BMC Plant Biol* 10, 83–97.
- Smith SR, Abbriano RM, and Hildebrand M. (2012). Comparative analysis of diatom genomes reveals substantial differences in the organization of carbon partitioning pathways. *Algal Res* 1, 2–16.
- Stirewalt VL, Michalowski CB, Loffelhardt W, Bohnert HJ, and Bryant DA. (1995). Nucleotide sequence of the cyanelle genome from *Cyanophora paradoxa*. *Plant Mol Biol Rep* 13, 327–332.
- Sucaet Y, and Deva T. (2011). Evolution and application of pathway resources and databases. *Brief Bioinform* 12, 530–544.
- Tatusov RL, Fedorova ND, Jackson JD, et al. (2003). The COG database: An updated version includes eukaryotes. *BMC Bioinformatics* 4, 41–55.
- Thimm O, Blasing O, Gibon Y, et al. (2004). MAPMAN: A user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J* 37, 914–939.
- Thomas PD, Campbell MJ, Kejariwal A, et al. (2003). PANTHER: A library of protein families and sub families indexed by function. *Genome Res* 13, 2129–2141.

- Tirichine L, and Bowler C. (2011). Decoding algal genomes: Tracing back the history of photosynthetic life on earth. *Plant J* 66, 45–57.
- Turmel M, Otis C, and Lemieux C. (1999). The complete chloroplast DNA sequence of the green alga *Nephroselmis olivaceae*: Insights into the architecture of ancestral chloroplast genomes. *Proc Natl Acad Sci USA* 96, 10248–10253.
- Turmel M, Otis C, and Lemieux C. (2002). The chloroplast and mitochondrial genome sequences of the chlorophyte *Chaetosphaeridium globosum*: Insights into the timing of the events that restructured organelle DNAs within the green algal lineage that led to land plants. *Proc Natl Acad Sci USA* 99, 11275–11280.
- Usadel B, Nagel A, Thimm O, et al. (2005). Extension of the visualization tool MapMan to allow statistical analysis of arrays, display of corresponding genes, and comparison with known responses. *Plant Physiol* 138, 1195–1204.
- 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, 40–57.
- Wahlund TM, Hadaegh AR, Clark R, Nguyen B, Fanelli M, and Read BA. (2004). Analysis of expressed sequence tags from calcifying cells of marine coccolithophorid (*Emiliania huxleyi*). *Mar Biotechnol* 6, 278–290.
- Weber A, Oesterhelt C, Gross W, et al. (2004). EST-analysis of the thermo-acidophilic red microalga *Galdieria sulphuraria* reveals potential for lipid A biosynthesis and unveils the pathway of carbon export from rhodoplasts. *Plant Mol Biol* 55, 17–32.
- Weiss TL, Johnston JS, Fujisawa K, Okada S, and Devarenne TP. (2011). Genome size and phylogenetic analysis of the A and L races of *Botryococcus braunii*. *J Appl Phycol* 23, 833–839.
- Weiss TL, Johnston JS, Fujisawa K, et al. (2010). Phylogenetic placement, genome size, and GC content of the liquid-hydrocarbon-producing green microalga *Botryococcus braunii* strain Berkeley (SHOWA) Chlorophyta. *J Phycol* 46, 534–540.
- Wijffels RH, and Barbosa M. (2010). An outlook on microalgal biofuels. *Science* 329, 796–799.
- Worden AZ, Lee JH, Mock T, et al. (2009). Evolution and dynamic adaptations revealed by genomes of the marine picocoeukaryotes micromonas. *Science* 324, 268–272.
- Yu WL, Ansari W, Schoepp NG, Hannon MJ, Mayfield SP, and Burkart MD. (2011). Modifications of the metabolic pathways of lipid and triacylglycerol production in microalgae. *Microb Cell Fact* 10, 91–102.
- Zhang W, Li F, and Nie L. (2010). Integrating multiple ‘omics’ analysis for microbial biology: Application and methodologies. *Microbiology* 156, 287–301.

Address correspondence to:

Dr. Prasanna Kumar Panda
 Bioresources Engineering Department
 Academy of Scientific and Innovative Research
 CSIR-Institute of Minerals and Materials Technology
 Bhubaneswar-751 013
 Odisha
 India

E-mail: pkpanda@immt.res.in