Increasing Microbial Biofuel Production by *In-silico* Comparative Genomic Studies

Gautham Subramaniam Ramakrishnan, Manali Mukund Kamath, and Vidya Niranjan

***Abstract*—Algal biofuels may be a kind of viable alternative to fossil fuels; however, this technology must overcome a number of hurdles before it can be considered to use in the market and be broadly deployed. One of the major hurdles are the low fuel yields per unit of biomass. In this study, we aim to overcome this challenge by identifying several genes which are responsible for increased lipid production, from numerous sources that can potentially increase the lipid synthesis in the autotrophic alga, *Chlamydomonas reinhardtii*. Using *in-silico* comparative genomics, we have made a shortlist of a total of 17 genes which, if incorporated into the genome of *Chlamydomonas reinhardtii* and overexpressed, could increase lipid production.**

***Index Terms*—Algal biofuels, *Chlamydomonas reinhardtii*, comparative genomics, lipid production.**

I. Introduction

The continued use of petroleum products is now widely considered unsustainable owing to the depletion of fuel reserves and the contribution of these fuels to global warming [1]. Renewable, efficient and eco-friendly fuels are necessary for environmental and economic sustainability. Interest in a variety of such renewable biofuels has been rejuvenated due to the instability of petroleum fuel costs, a reliance on unstable petroleum resources, and the dangers of increasing atmospheric CO_2 levels.

Biofuels can be solids, liquids or gasses so long as they are derived directly from biological sources. The most common solid biofuel is lignified cellulose (wood) that can be burned for energy. Liquid and gaseous biofuels generally require more refining and include bioethanol, biodiesel and engine-combustible hydrocarbons as well as methane from anaerobic digestion. The aforementioned liquid biofuels offer significant potential to augment or replace petroleum gasoline for transportation purposes. Currently ethanol dominates the biofuel market and may be produced by a variety of methods, primarily heterotrophic fermentation of sugars purified from biomass feedstocks [2]. Biodiesel and other hydrotreated biofuels are derived mainly from vegetable oil feedstocks (lipids) [3].

Ethanol and biodiesel are primarily derived from plant sources, often food crops, because the established scale of food crops made them a convenient source of biomass necessary to produce biofuel on a commercial scale. However, an increasing demand for biofuel feedstocks has negatively impacted food markets, and raised a global "food vs. fuel" controversy. Furthermore, the land and fresh water requirements for growing crops and the long growth-to-harvest periods limit the expansion of plant based biofuel industries to the amount of arable land. In contrast, unicellular algae requiring smaller amounts of land that does not need to be arable, have faster growing cycles, contain a higher percentage of oil, and have been proposed to be a better solution to the food vs. fuel debate. Therefore, significant attention has been focused on algae as a next generation feedstock for biofuel production [4].

Photosynthetic algae, both microalgae and macroalgae (i.e., seaweeds), has been of considerable interest as a possible biofuel resource for decades. Researchers have been exploiting algae for biodiesel production due to their short life cycle, less labor required, less affection by venue, and easier scale up procedure [5]. Several species have biomass production rates that can surpass those of terrestrial plants, and many eukaryotic microalgae have the ability to store significant amounts of energy-rich compounds, such as triacylglycerol (TAG) and starch, which can be utilized for the production of several distinct biofuels, including biodiesel and ethanol [6]. Fatty acids are also important precursors to biofuel [7].

Microalgae are especially attractive as a source of fuel from an environmental standpoint because they consume carbon dioxide and can be grown on marginal land, using waste or salt water [6]. In addition, it may be possible to leverage the metabolic pathways of microalgae to produce a wide variety of biofuels (see Fig. 1).

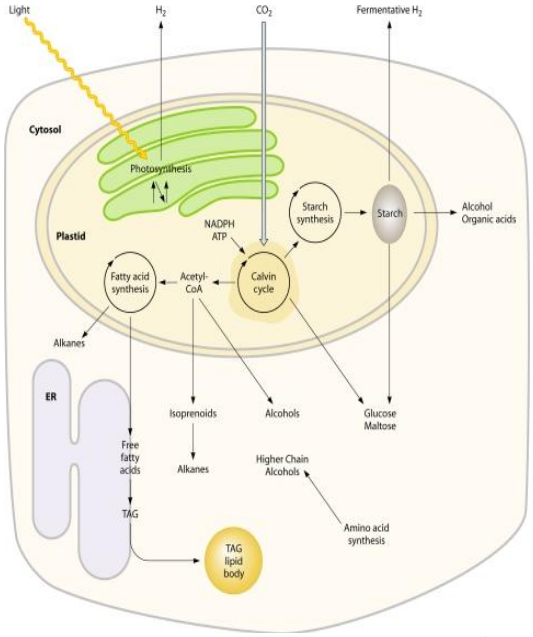

Fig. 1. Microalgal metabolic pathways that can be leveraged for biofuel production.

Despite the immense potential, several technical barriers need to be overcome before microalgal biofuel can be

Manuscript received March 12, 2014; revised May 12, 2014.

The authors are with Dayananda Sagar College of Engineering, India (e-mail: gautham.sr@gmail.com).

successfully brought into the market [8]. These include developing low-energy methods to harvest microalgal cells, difficulties in consistently producing biomass at a large scale in highly variable outdoor conditions, the presence of invasive species in large-scale ponds, low light penetrance in dense microalgal cultures and the lack of cost-effective bioenergy carrier extraction techniques of most microalga-derived biodiesel. The major drawback of microbial biofuel production, however, in is the astonishingly low yield of fuel per unit of biomass [9]. In light of these problems, it seems genetic modification is likely to provide the key to unlock the feasibility of algal production strains.

Chlamydomonas reinhardtii is a green microalga (Fig. 2) whose lineage diverged from land plants over 1 billion years ago [10]. It has recently emerged as a model to test genetic engineering or cultivation strategies aiming at increasing lipid yields for biodiesel production [11].

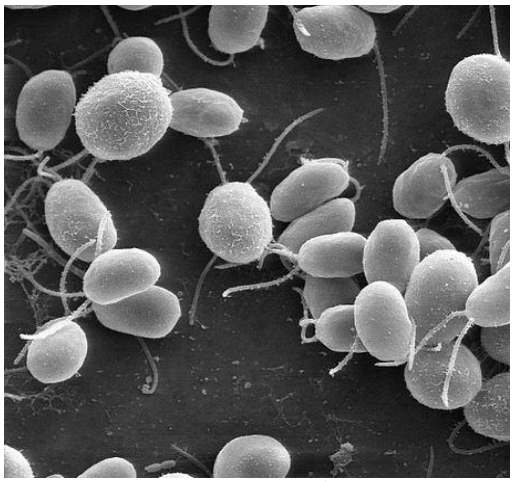

Fig. 2. Scanning electron microscope image of *Chlamydomonas reinhardtii.*

Our work aims to provide some insight on the possible genetic manipulations of *Chlamydomonas reinhardtii* to obtain higher yields of biofuel using methods of *in silico* comparative genomics.

II. Materials and Methods

A. Shortlist of Lipid Producing Genes

A selection of proteins involved in the regulation of lipid synthesis was shortlisted [12]. The shortlisted protein sequences were downloaded from the UniProt database [13]. The genes that were responsible for coding for these proteins were also deduced by study of available literature.

These genes, when expressed, were responsible either for the upregulation or downregulation of lipid synthesis. Hence, we conceived that manipulation of these genes could lead to an increased production of lipids. Only the genes that lead to an upregulation of lipid synthesis were selected for further study.

Based on homology, sources of the genes were narrowed down to 5 organisms, namely *Arabidopsis thaliana*, *Chlamydomonas reinhardtii*, *Ostreococcus lucimarinus*, *Ostreococcus tauri* and *Volvox carteri.*

B. Sequence Retrieval

The sequences of the genes encoding all the selected proteins were retrieved in FASTA format.

Once these sequences were procured, the *Chlamydomonas reinhardtii* genome sequence was obtained from the JGI Genome Portal [14]. The 121-megabase (Mb) draft sequence of the *Chlamydomonas* nuclear genome has the following characteristics [10]:

- It was generated at 13× coverage by whole-genome, shotgun end-sequencing of plasmid and fosmid libraries, followed by assembly into ~1500 scaffolds.
- Half of the assembled genome is contained in 25 scaffolds, each longer than 1.63 Mb.
- The genome is unusually GC-rich (64%).
- The *Chlamydomonas* nuclear genome comprises of 17 linkages presumably corresponding to 17 chromosomes, consistent with electron microscopy of meiotic synaptonemal complexes [15].
- Seventy-four scaffolds, representing 78% of the draft genome, have been aligned with linkage groups.

The *Chlamydomonas reinhardtii* average gene properties were as follows:

- Gene density: 149 genes/ Mbp assembly
- Gene length: 3895 nt
- Transcript length: 1768 nt
- Exon length: 240 nt
- Intron length: 336 nt
- Exon frequency: 7.4 exons/gene

The protein and transcriptome data of this organism were also downloaded and stored [14]. The protein coding genes had the following features [14]:

- A reference set of 15,143 protein-coding gene predictions has been created.
- More than 300,000 ESTs have been generated.
- 8631 gene models (56%) are supported by mRNA or EST evidence.
- 35% have been edited for gene structure and/or annotated by manual curation (as of June 2007).
- Protein-coding genes have, on average, 8.3 exons per gene and are intron-rich relative to other unicellular eukaryotes.
- The average *Chlamydomonas* intron is longer (373 bp) than that of many eukaryotes and only 8% lack introns.

C. BLAST Using Bioedit

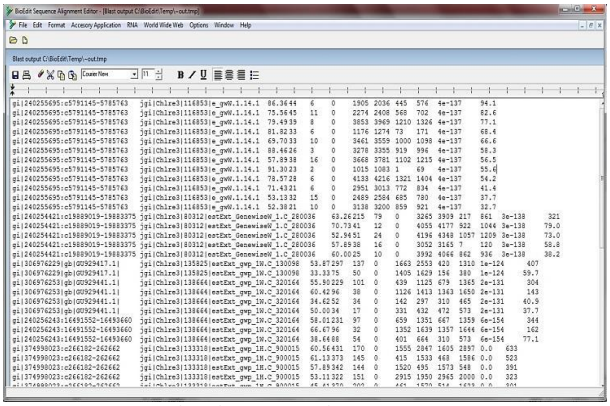

Fig. 3. The first few results of BLAST using BioEdit.

The next step involved the alignment of sequences using a standalone BLAST application called BioEdit [16]. As the genome is *de novo* and annotation of the protein is not complete, a BLAST [17] research of the listed proteins against the protein and transcriptome databases of *C.*

reinhardtii was performed.

Our objective was to map the homologous proteins responsible for lipid synthesis as well as to identify novel genes which are responsible for the lipid production. The result obtained gave us the percentage identity, alignment length, number of mismatches and the HSP (high-scoring segment pair) score of the said sequences.

The results were filtered to include only the proteins with HSP score greater than 500 and tabulated them. Only these proteins were chosen for the further study.

D. Screening the Candidate Genes

The resultant proteins and expression studies relevant to them were further studied. The results that were obtained from the sequence homology search against the protein and transcriptome databases of *C. reinhardtii* verified that the genes were transcribed and translated to proteins. This also ensured that regions with false positives, owing to the presence of pseudo genes, are not shortlisted. The number of favorable proteins was thus narrowed down to 34. Further detailed study was carried out and the function of each protein was found out along with the various genes involved in the upregulation of lipid synthesis. A total of 17 genes were screened and shortlisted. We proposed that incorporation of these genes into the genome of *C. reinhardtii* would help to achieve an increased production of lipid for biofuel, thus maximizing its marketability.

The final screening of the genes was conducted solely based on their functionality. The gene functions and their source organisms were identified as follows. The FadR (Fatty acid metabolism regulator protein), accA (acetyl-CoA carboxyl - transferase subunit alpha), accB (acetyl-CoA carboxylase, BCCP subunit), accD (acetyl-CoA carboxylase, beta carboxyl-transferase subunit), fabH (3-oxoacyl-[acyl-carrier-protein] synthase III) and fabI (enoyl-[acyl-carrier-protein] reductase, NADH-dependent) genes were shortlisted from *Escherichia coli*, str. K-12, substr. MG1655. This particular strain, a model organism, has a genome of length 4,641,652 base pairs. The genes accC (acetyl-CoA carboxylase, biotin carboxylase subunit) and fabD (malonyl-CoA-[acyl-carrier-protein] transacylase) were catalogued from *Escherichia coli str.* K-12 substr. W3110, having a genome of length of 4,646,332 base pairs. fabG (3-oxoacyl-[acyl-carrier-protein] reductase) gene was taken from *Francisella tularensis subsp. mediasiatica* FSC147 chromosome which consists of 1,893,886 base pairs. The gene fabZ ((3R) hydroxymyristol acyl carrier protein dehydratase) was shortlisted from *Lactococcus lactis*, strain IO-1. The gene Gut1 (glycerol kinase) was listed from *Saccharomyces cerevisiae* (budding yeast) S288c, chromosome VIII. Similarly, Ptb (phosphate butyryltransferase) and buk (butyrate kinase) were catalogued from *Bacillus megaterium* WSH-002 chromosome and *Bacillus megaterium QM* B1551 chromosome respectively. adhE2 (Aldehyde-alcohol dehydrogenase) gene was taken from *Azospirillum lipoferum* str.4B plasmid AZO_p1. DGAT gene was listed from *Mus musculus*, chromosome 15. The genes ME (Malic Enzyme) and PEPC (phosphoenolpyruvate carboxylase kinase) were taken from *Arabidopsis thaliana* chromosome 5 and chromosome 1respectively.

E. Phylogenetic Analysis

The literature survey showed that *Arabidopsis thaliana*, *Ostreococcus lucimarinus*, *Ostreococcus tauri* and *Volvox carteri* were homologous to *Chlamydomonas reinhardtii* [12]. To confirm this, we carried out a phylogenetic analysis. The sequence of the protein Heteromeric ACC biotin carboxylase subunit (BCC) was retrieved for all these 5 organisms from GenBank. These sequences were analyzed using Clustal Omega, a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments [18].

III. Result and Discussion

In the BLAST search against the *C. reinhardtii* transcriptome, we obtained a total of 197 hits. Upon filtering based on the HSP value, this number reduced to 110. Similarly, in the search against the *C. reinhardtii* protein database, we recorded a total of 198 hits, which reduced to 117 upon filtering. The number of proteins is common to both the transcriptome and the protein database boiled down to 34. Upon further literature survey, we found that only 13 of these proteins had been studied for expression and there were 35 genes which could affect the lipid synthesis. Finally, we came up with a total of 17 genes that satisfied our searching criteria. They have been shown in Table I.

TABLE I: Final List of Genes Enhancing Lipid Yield

GENE	SOURCE
FadR	*Escherichia coli*
accA	*Escherichia coli*
accB	*Escherichia coli*
accC	*Escherichia coli*
accD	*Escherichia coli*
fabD	*Escherichia coli*
fabH	*Escherichia coli*
fabG	*Francisella tularensis*
fabZ	*Lactococcus lactis*
fabI	*Escherichia coli*
GUT1	*Saccharomyces cerevisiae*
Ptb	*Bacillus megaterium*
Buk	*Bacillus megaterium*
adhE2	*Azospirillum lipoferum*
DGAT	*Mus musculus*
me	*Arabidopsis thaliana*
PEPC	*Arabidopsis thaliana*

Although we found 35 genes responsible for regulation of lipid synthesis, only 17 of these had been proven to increase lipid production if overexpressed. As a result, only these 17 genes satisfied our searching criteria.

During the literature survey, we came across a study of homology in a paper by Merchant SS, *et al.* stating that *O. lucimarinus*, *A. thaliana* and *O. tauri* were homologs of *C. reinhardtii*, thus proving the validity of our work. Later, we obtained evidence that *Volvox carteri*, which is a species of colonial green algae [19], has also been studied for lipid production and is a homolog of *C. reinhardtii* [12]. Verification of this observation by a phylogenetic analysis between the said organisms and *Chlamydomonas reinhardtii* (Fig. 4) showed that *Volvox carteri* is indeed highly similar to *C. reinhardtii*. The presence of lipid producing genes in all

the mentioned organisms and their evident homology suggests that transformation of *C. reinhardtii* with these genes would be a feasible endeavor.

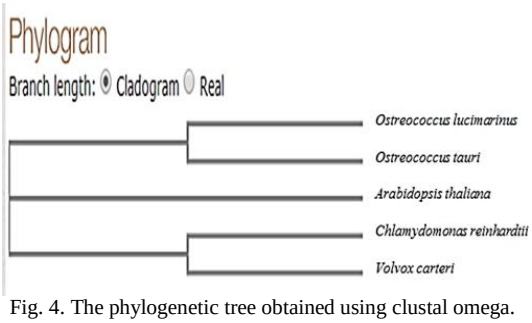

Fig. 4. The phylogenetic tree obtained using clustal omega.

IV. Conclusion

Microalgae has emerged as a popular feedstock for the production of biofuels and other products [20]. They are an extremely diverse group of organisms, many of which possess unique metabolic features that can be altered for the production of renewable biofuels [6]. These include:

- High efficiency of conversion during photosynthesis.
- Rapid rates of biomass production.
- The capability to produce a wide variety of biofuel feedstocks.
- The ability to thrive in diverse ecosystems

Although microalgae have long been considered a promising feedstock for the production of biofuels, studies have concluded that the economic viability of microalgal biofuel production is in need of significant improvement. However, in contrast to previous efforts, we are now equipped with an arsenal of new genetic tools, genome sequences, and high-throughput analytical techniques that will equip and empower scientists to analyze and manipulate metabolic pathways with unrivaled precision [6].

To date, molecular genetic tools have been developed for only a few algae, including the green alga *Chlamydomonas reinhardtii*, and the diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*. Among these, only *C. reinhardtii* has well established systems for recombinant protein expression from both the nuclear and chloroplast genomes, and even in this species only a handful of recombinant proteins have been expressed to any appreciable level [21].

Though expression studies have also been done for bacteria such as *E. coli* and there are molecular genetic tools available for the same, the prokaryotic genome of bacterial cells lacks certain advanced post translational modification systems that are seen in the higher eukaryotes. Additionally, microalgae such as *C. reinhardtii* are photosynthetic and are easier to cultivate. They can even be cultivated in natural, for instance, open pond systems with waste water as a source of nutrients [9]. It has recently been demonstrated that *C. reinhardtii* can be engineered to produce novel hydrocarbon molecules that are superior biofuels, demonstrating the potential of microalgae as a biofuel source.

In silico screening of genes for genetic engineering is a versatile methodology to aid us in creating mutants with high biofuel productivity. These methods at organism and functional level can reduce years of wet lab experimentation by narrowing down the huge number of genes and even pinpoint the most desirable ones. Optimizing microalgal biofuel production using metabolic engineering tools requires an in-depth understanding of the structure-function relationship of genes involved in lipid biosynthetic pathway [12].

Our bioinformatics approach considers a wide range of data with various parameters and criteria. This enhances the quality and scope of the research we have conducted.

Acknowledgment

We would like to thank the professors of R.V. College of Engineering, Bangalore, India for supporting us in our work and for their valuable inputs to help us write the paper. We also extend our gratitude to the staff of Dayananda Sagar College of Engineering, Bangalore, India, for providing their opinions and suggestions on the original idea.

References

[1] Y. Chisti, "Biodiesel from microalgae," *Biotechnology Advances*, vol. 25, no. 3, pp. 294-306, 2007.
[2] D. Chiaramonti, "Improvement of crop plants for industrial end uses," *Bioethanol: Role and Production Technologies*, pp. 209-251, 2007.
[3] J. V. Gerpen, "Improvement of crop plants for industrial end uses," *Biodiesel Production*, pp. 281-289, 2007.
[4] W. L. Yu, W. Ansari, N. G. Schoepp, M. J. Hannon *et al.*, "Modifications of the metabolic pathways and triacylglycerol production in microalgae," *Microbial Cell Fact*, vol. 10, p. 91, 2011.
[5] M. H. Liang and J. G. Jiang, "Advancing oleaginous microorganisms to produce lipid via metabolic engineering technology," *Progress in Lipid Research*, vol. 52, no. 4, pp. 395-408, 2013.
[6] R. Radakovits, R. E. Jinkerson, A. Darzins, and M. C. Posewitz, "Genetic engineering of algae for enhanced biofuel production," *Eukaryotic Cell*, vol. 9, no. 4, pp. 486-501, 2010.
[7] F. Zhang, M. Ouellet, T. S. Batth, P. D. Adams *et al*., "Enhancing fatty acid production by the expression of the regulatory transcription factor FadR," *Metabolic Engineering*, vol. 14, no. 6, pp. 653-60, 2012.
[8] P. T. Pienkos and A. Darzins, "The promise and challenges of microalgal-derived biofuels," *Biofuels Bioprod. Bioref*, vol. 3, pp. 431–440, 2009.
[9] M. Hannon *et al.*, "Biofuels from algae: Challenges and potential," *Biofuels*, vol. 1, no. 5, pp. 763–784, 2010.
[10] S. S. Merchant *et al*., "The *Chlamydomonas* genome reveals the evolution of key animal and plant functions," *Science*; vol. 318, no. 5848, pp. 245-50, 2007.
[11] M. Siaut, S. Cuiné, C. Cagnon, B. Fessler *et al*., "Oil accumulation in the model green alga *chlamydomonas reinhardtii*: characterization, variability between common laboratory strains and relationship with starch reserves," *BMC Biotechnology*, vol. 11, p. 7, 2011.
[12] N. Misra, P. K. Panda, B. K. Parida, and B. K. Mishra, "Phylogenomic study of lipid genes involved in microalgal biofuel production - candidate gene mining and metabolic pathway analysis," *Evolutionary Bioinformatics*, vol. 8, pp. 545–564, 2012.
[13] The UniProt Consortium, "Activities at the Universal Protein Resource (UniProt)," *Nucleic Acids Res*., vol. 42, pp. D191-D198, 2014.
[14] Joint Genome Institute: *Chlamydomonas reinhardtii* v 3.0. DOE Joint Genome Institute. [Online]. Available: www.jgi.doe.gov/chlamy.
[15] R. Storms and P. J. Hastings, "A fine structure analysis of meiotic pairing in *Chlamydomonas reinhardtii*" *Exp Cell Res*., vol. 104, no. 1, pp. 39-46, 1977.
[16] T. A. Hall, "BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT," *Nucleic Acids Symp*., vol. 41, pp. 95-98, 1999.
[17] Altschul, F. Stephen, T. L. Madden, A. A. Schaffer, J. H. Zhang, Z. Zhang, W. Miller, and D. J. Lipman, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Res*., vol. 25, pp. 3389-3402, 1997.
[18] F. Sievers, A. Wilm, D. Dineen, T. J. Gibson *et al.*, "Fast, scalable generation of high‐quality protein multiple sequence alignments using Clustal Omega," *Molecular Systems Biology*, vol. 7, p. 539, 2011.
[19] M. D. Guiry and G. M.Guiry, "*Volvox carteri*," *Algae Base*, World-wide electronic publication, National University of Ireland, Galway, 2008.

[20] E. Stephens, I. L. Ross, J. H. Mussgnug, L. D. Wagner *et al*., "Future prospects of microalgal biofuel production systems," *Trends in Plant Science*, vol. 5, no. 10, pp. 554-64, 2010.
[21] S. P. Mayfield *et al*., "*Chlamydomonas reinhardtii* chloroplasts as protein factories," *Curr Opin Biotechnol*, vol. 18, no. 2, pp. 126-133, 2007.

Vidya Niranjan was born in February 1972. She has 19 years of core teaching and bioinformatics research experience along with 8 years of exclusive teaching experience. Currently she is heading the Department of Bioinformatics at R. V. College of Engineering, Bangalore, Karnataka, India, she has mentored more than 300 people studying M. Tech, B. Tech and M. Sc courses in bioinformatics research. She is also a consultant as bioinformatics domain expert in IT industries and has 1 patent in her name. She is a PLOS reviewer from 2011-till date and has excellent verbal and written communication skills. She has publications in international journals like Nature, Genetics, Genome Research, NAR, TIBS. She also has experience in Next generation sequencing data analysis and has competency in design, development and maintenance of research tools. She also has leadership qualities for inspiring and leading a team of professionals/students and ability to work independently along with strong code of ethics and dedication to uphold quality standards.

Gautham Subramaniam Ramakrishnan was born in December 1991 in Bangalore, India. He is currently pursuing his bachelor's degree in engineering at Dayananda Sagar College of Engineering, Bangalore, Karnataka, India in biotechnology. He has presented a poster on biofuels to a panel of 20 environmentalists at Indian Institute of Management, Bangalore [IIMB]. He has received accolades for various other poster presentations at AICTE recognized national level symposiums. He is also an active member of the Association of Microbiologists of India (AMI), Bangalore Chapter.

Manali Mukund Kamath was born in August 1992 in Pune, India. She is currently pursuing her bachelor's degree in engineering at Dayananda Sagar College of Engineering, Bangalore, Karnataka, India in biotechnology. She has undergone several practical training programs in areas like enzyme technology and immunology. She has participated and won first place for poster presentation at an AICTE recognized national level symposium. She is also an active member of The Association of Microbiologists of India (AMI), Bangalore Chapter.