
Transforming Modern Agriculture Through Synthetic Genomics

JIM FLATT

*Synthetic Genomics, Inc.
La Jolla, California*

jflatt@syntheticgenomics.com

In this talk, I will cover a few topics. First, I will describe some of the cutting-edge science that's underway and how we are applying it within Synthetic Genomics, not specifically on specialty crops as defined, but on low-acreage or potential crops of the future. Many of these concepts will apply to specialty crops.

We can think about progress in synthetic biology, much in the way that children progress through school (Figure 1). First, in 1995, whole-genome sequences were published for two simple bacterial species (*Haemophilus influenzae* and *Mycoplasma genitalium*), which, in the late 1990s, were followed by publication of several plant genomes, culminating in 2001 with the publication of the draft sequence of the human genome. That was equivalent to learning to read, as in early grade school, begging the question: "If we can read the genetic code, can we begin to write it?" In 2003, the synthesis of a small viral genome was achieved, at a little over 5,000 base pairs. In 2006, effort was initiated to chemically synthesize a bacterial cell. The result was published by J. Craig Venter Institute researchers in 2010: fundamentally, this organism came from nature and we recapitulated it with a few additional sequences.

WRITING STORIES

What's important now is to take the tools that we have developed in reading and writing to create our own stories. This is where the design aspect gets interesting in terms of developing crops that not only are more robust and higher yielding, but can be more beneficial from a health standpoint. Focusing on some of the tools, Figure 2 shows a simplified diagram of the process used to produce the first chemically synthesized cell, which has a genome of about 1.1 megabases, *i.e.* over a million base pairs. When this

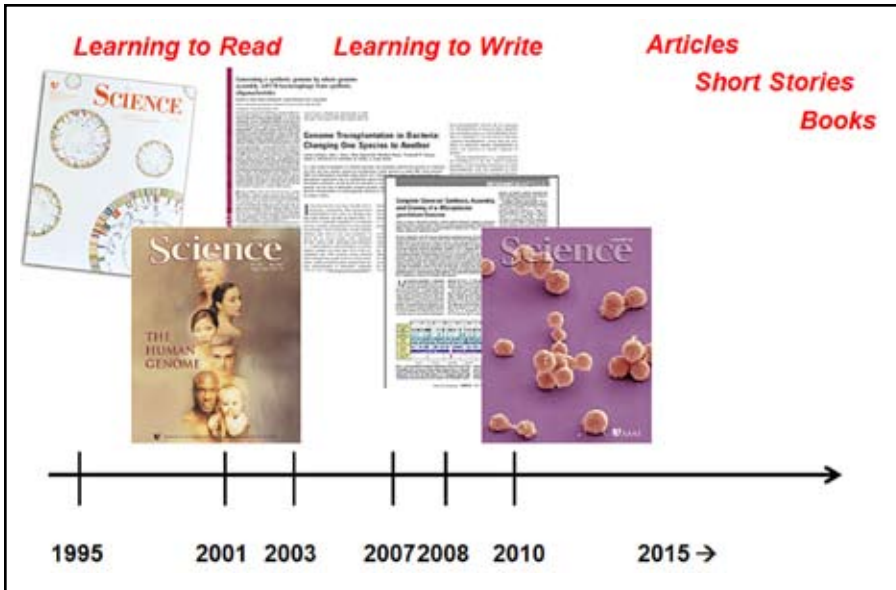


Figure 1. Progress in synthetic biology.

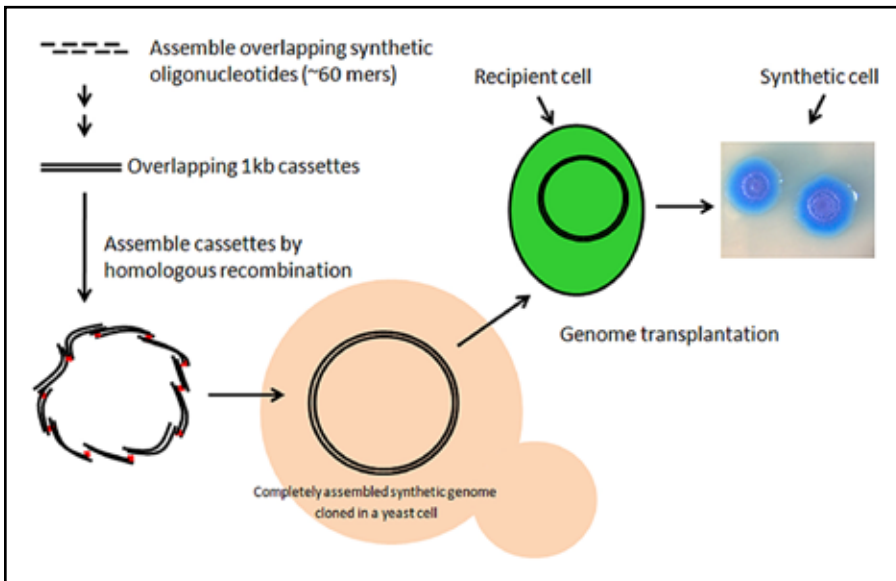


Figure 2. Approach used to synthesize a *Mycoplasma mycoides* cell.

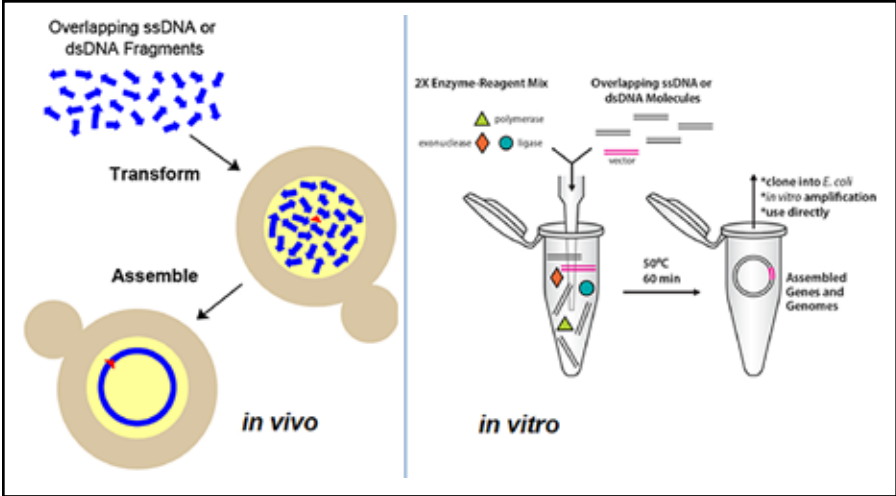
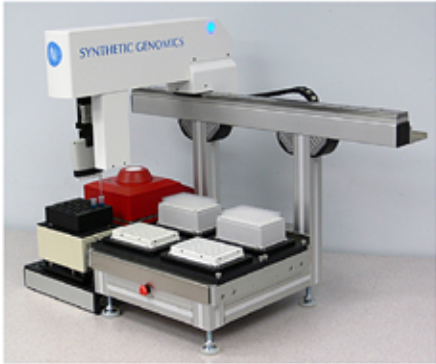


Figure 3. DNA synthesis/assembly methods: alternatives enabling combinatorial assembly.

Oligo Pools → Biological entities (DNA, proteins, viruses, and microbial cells)

produces 1 – 32 genes every 8 hours



- * SGI-Third Party relationship
- * Includes SGI proprietary DNA synthesis and error correction methods
- * Progress is monitored through email
- * Received at SGI in December, 2012
- * dependable: synthesis of 24 1.5kb constructs in a single run and >3kb fragments per run

MAX CAPACITY (bp/instrument)

Daily: ~190 kb	Monthly: 5.8 Mb
Weekly: 1.3 Mb	Annual: 70 Mb

Figure 4. Automation of DNA assembly.

project was undertaken, the largest DNA assemblies that had been produced were of about 30 kilobases—about a thirty-fifth the size. One of the challenges was stitching DNA together accurately in these large assemblies. Several methods were developed, but constructing these assemblies was only part of the problem; the next question was how to activate it and give it life. Methods were developed for handling whole genomes and transplanting them into related recipient cells that essentially served as the birthplace for the new genome. And again, in this case, much like a virus takes over the host cell, the transplanted genome was replicated, transcribed and translated, producing marked synthetic cells as in Figure 2.

SIMPLE METHODS FOR DNA ASSEMBLY

Important for future progress was the development of some simple, but powerful, methods for assembling DNA. The right-hand side of Figure 3 illustrates the Gibson assembly method, now utilized in many laboratories; Dan Gibson's insight was to use DNA-repair mechanisms as the basis for developing an *in vitro* mix that is isothermal and reliable. This method is quite amenable to automation because of its relative simplicity, and Figure 4 shows a unit we received in December of 2012, the prototype of a benchtop instrument that will allow all laboratories to synthesize genes or even larger segments of DNA.

The instrument is loaded with the oligonucleotides encoding the gene of interest, a button is pushed and overnight the constructs are made. And this has actually been used for rapid production of influenza vaccines. SGI has a number of partnerships with leading companies, including Novartis Vaccines to develop a faster method to synthesize vaccine seeds so that Novartis can get a supply of influenza vaccine to the market more rapidly. The importance of this was illustrated in 2009, when the H1N1 epidemic was of great concern. As Figure 5 shows, the number of cases was growing exponentially before the vaccine became available. The reason it took so long to get the vaccine to market was both a function of a long process to develop the vaccine seeds as well as low yields with the H1N1 virus.

There needed to be a better solution, so the Centers For Disease Control Board/ Biomedical Advanced Research and Development Authority (CDC/BARDA) is funding Novartis and Synthetic Genomics to develop a better, faster and more reliable method for, not only producing vaccines for seasonal flu cases, but also to have a method that can be used for responses to pandemics. We have been successful in taking what is typically a six-week process using classical genetics to isolate virus-vaccine seeds to a process where, now, when the World Health Organization releases information about a strain, we take that sequence and, using our assembly methods, can synthesize the DNA constructs in about twelve hours, achieve infection of mammalian cells and recover active virus seeds in five to six days. This shortens the process by about six weeks. In the long-term, we will be able to survey all of the possible materials out there and bank the viral gene segments for assembly overnight when called for. This was recently reported with the H7N9 strain in China, which is currently not available in the United States. However, BARDA requested it and we were successful in assembling and producing a vaccine seed without having those viral strains available.

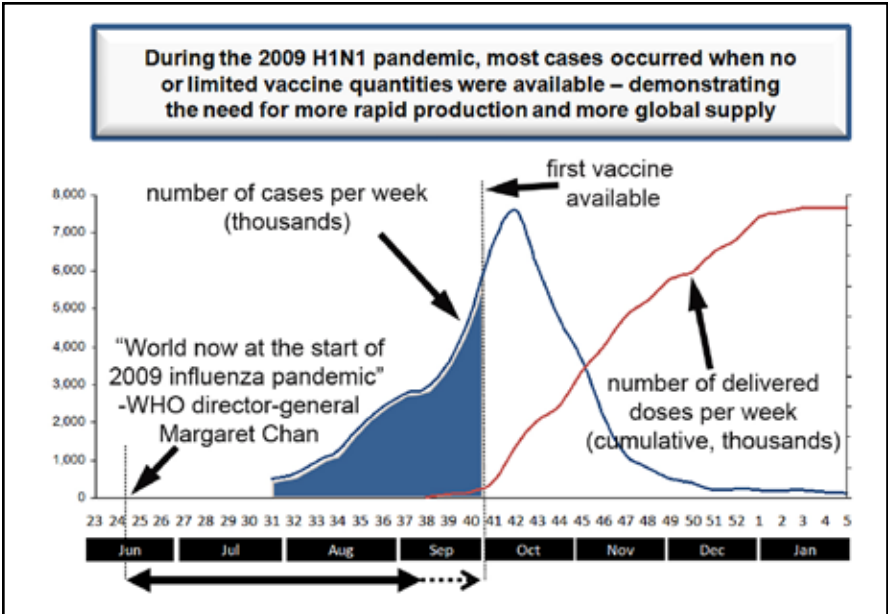


Figure 5. Synthetic Genomics case study: Influenza vaccine.
 (Data provided by Phil Dormitzer at Novartis Vaccines & Diagnostics.
http://www.cdc.gov/h1n1flu/estimates_2009_h1n1.htm)

These improvements in speed are important, but so too is accuracy. We have worked hard to develop methods to weed out errors that are inherent with oligonucleotides, the 60-base-pair segments that are the building blocks for genes and, ultimately, genomes. They can be only 60 percent to 70 percent pure, therefore, if large numbers are stitched together, assembly errors are almost guaranteed, which can be problematic. To address this, we have used bioinformatic capability to develop error-correction methods that we now incorporate routinely. This allows us to accurately assemble DNA segments of up to seven kilobases without intermediate sequencing to verify accuracy.

Another important factor is cost. At the time, building the genome of *Mycoplasma mycoides* cost about \$1 million in reagents alone. It was a fascinating project, but not something that would be taken on routinely because of the necessary expenditures. This field as a whole depends not only on improved accuracy, but on lowering the cost of DNA assembly. Figure 6 shows the exponential decrease that has occurred in sequencing costs since 1990, which has fostered applications of “omics” technologies.

We haven’t seen the same reduction in cost of gene synthesis (Figure 6). We and scientists in several other labs are working on methods that involve ultra-low-cost DNA sources from microchips as well as next-generation sequencers that allow retrieval of validated correct sequences to begin assemblies. If successful, these methods could lower the cost by at least an order of magnitude, bringing the assembly cost down to a penny or two per base, presenting the possibility of new ways of improving specialty and other crops.

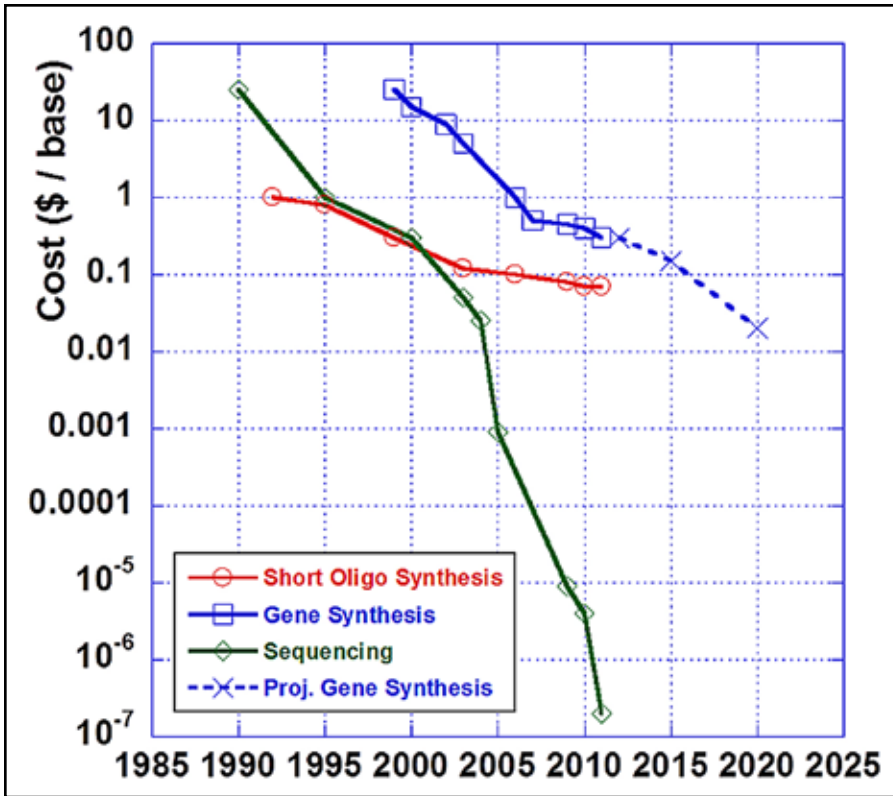


Figure 6. Cost trends: DNA synthesis and assembly.
 (Adapted from R. Carlson, www.synthesis.cc.
 Includes SGI Projection.)

TRAIT DISCOVERY AND PATHWAY ENGINEERING

Over the past six years, we have developed a web-based scalable comprehensive bioinformatics platform to allow not only computational biologists, but also novice users, to analyze genomic information and to use that information in design and construction of DNA assemblies for recombinant cells. And we have done an extensive amount of biodiscovery and characterization of fungal, plant and algal genomes, which increases the diversity of information available for our work.

Using our proprietary enrichment and isolation methods, we have acquired some 4,000 microbial isolates—associated with wild grasses—that have been screened for various properties. One of the interesting things that came out of this was the idea that, with these bioinformatics methods and knowledge of gene structure/function, we could then use bioinformatics as an assay in the discovery effort. With a view to discovering new *Bt*-type insecticidal toxins, we took our *Bacillus* isolates—a subset of about 200 in

- 188 mutants of enzyme A created (two plates)
 - alignment with enzyme B and crystal structure analysis
 - 1 - 43 aa were mutated
- 110+ mutants folded properly (absorbance assay)
- Screened reactivity against substrates 1 and 2
 - native enzyme A has no reactivity with either substrate
 - engineered enzyme A products obtained with both substrates
 - >10 mutants showed products from a target substrate

Figure 7. Engineered enzyme activity results: the substrate range of the target enzyme was modified creating the desired activity.

number—and isolated the plasmid DNA potentially encoding such toxins, pooled those plasmids, went through a next-generation sequencing effort and then were able to assemble and annotate that information. Within a six-week period, we discovered fifty full-length novel *Bt* genes. This pilot test demonstrated the potential utility of this extensive amount of information when analyzed with the new tools.

We now have the ability to efficiently and cost-effectively assemble DNA structures from starting-material oligonucleotides. We can now construct digitally designed protein variants rather than employ traditional methods that involve either random changes such as error-prone PCR or gene shuffling or site-directed mutagenesis, which allows us to make base-pair nucleotide changes, but only in specific regions. Being able to design this on the computer gives us unlimited flexibility.

We had an enzyme for which we were trying to modify the substrate specificity to induce a desired carbon-carbon bond-cleavage reaction. We had an enzyme that performed other chemistry on those identical substrates and then we had “enzyme A,” which in fact did not work on those substrates but catalyzed the reaction of interest. Through protein modeling, we identified a number of changes that would be beneficial within and outside the active site and made a set of 188 protein variants, which we screened for function (Figure 7). Within one round, we obtained ten mutants that worked on both of the substrates of interest. In essence, we had engineered the desired change in substrate specificity.

MOLECULAR BREEDING FOR SPECIALTY CROPS

One of the foci of our AgraCast subsidiary is the development of castor-oil plant (*Ricinus communis*) as a source of specialty chemicals. Castor is of interest due to its high content

of ricinoleic acid, a hydroxylated fatty acid that provides a platform for producing branch-chain chemicals. It is used by BASF to produce lubricants, for example. However, a barrier to broader scale adoption is lack of availability, lack of consistency and high cost. Although it grows wild in Texas and in Mexico, commercial production is mainly in India by small farmers with relatively low yields, one to two tons per hectare. Typically, it is harvested manually because of the plant's architecture. Several years ago, we began both classical and molecular breeding programs to identify plants with larger racemes and higher yields (Figure 8). We are up to about four tons per hectare.

ALGAE AS A MAJOR CROP OF THE FUTURE

By FAO estimates, we will need to increase the food supply significantly in the foreseeable future, and we will have to do it without accessing more arable land, usage of which has been stagnant for decades. That, coupled with issues associated with climate change, increasingly important issues around water availability and the fact that we are seeing increases in demand for animal protein that are primarily correlated with increased economic development, we see the need not only for increased productivity of our major crops, but also of new crops that can be cultivated on non-arable lands with minimal inputs of fresh water.

We have had a collaboration with Exxon-Mobil since 2009, researching algae for use as biofuels, but we may see this commercially utilizable in production of algal-based proteins. The potential is shown in Figure 9: algae—even the current forms—are much more efficient producers of protein than are terrestrial counterparts. Our characterizations show that algal protein provides complete sources of amino acids and are highly digestible. Major barriers associated with algae as commercial sources of protein include developing domesticated species that will grow robustly in the wild. We see a need to use synthetic genomics techniques to combine beneficial traits that have utility under specialized conditions and combine them with photosynthetic efficiency, to allow us to channel carbon to target molecules and show improved tolerance of environmental stresses. One of the most significant things that we'll be reporting on towards the end of 2013 is research to improve photosynthetic efficiency.

Figure 10 shows that when algae grow in dense culture, light becomes limiting and overall productivity decreases. In part, this is because, when light is low the algae acclimate and build larger antennae that shade their neighbors. We have engineered semisynthetic algae that are deregulated in their response to light, allowing significantly more light to penetrate the culture without compromising the photosynthetic processes and functionality of the cell. We've also taken similar steps to increase lipid productivity in just a matter of a few months, again based on bioinformatics and our ability to modify biosynthetic pathways. Our algal research is at an early stage, but our synthetic genomics techniques give the ability to accelerate the developmental process. Ultimately, I think that we will see large-scale algal-production facilities providing a growing share of protein requirement in the future.

<u>Why Molecular Breeding?</u>	<u>Molecular Breeding Targets</u>
<ul style="list-style-type: none"> • Faster crop improvement • Reduced phenotyping costs • More rapid hybrid development • Quality control • Protect IP (Elite Lines) 	<ul style="list-style-type: none"> • Plant size • Proportion of female flowers • Raceme size • Disease resistance • Toxin elimination
<p>Molecular breeding will sustain Agracast competitive advantage in castor</p>	

Figure 8. Molecular-breeding rationale and targets: the molecular-breeding approach developed for large-acreage crops will be increasingly applied to specialty crops like castor bean.

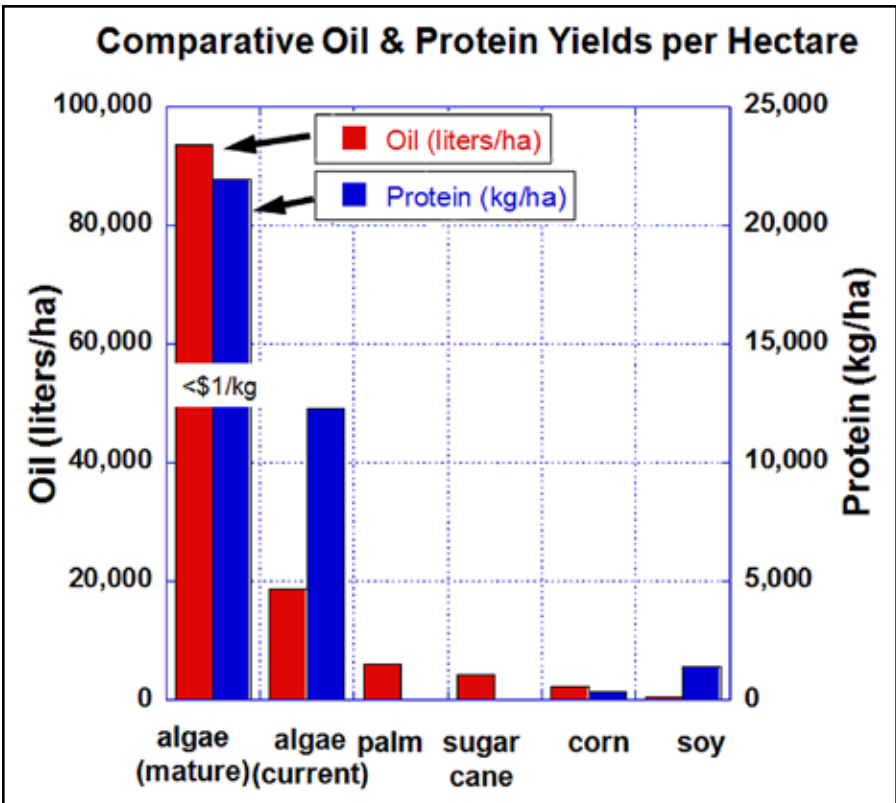


Figure 9. Algae is the best, scalable production system in a land-, water-, and carbon-constrained world (data based on various literature reports).

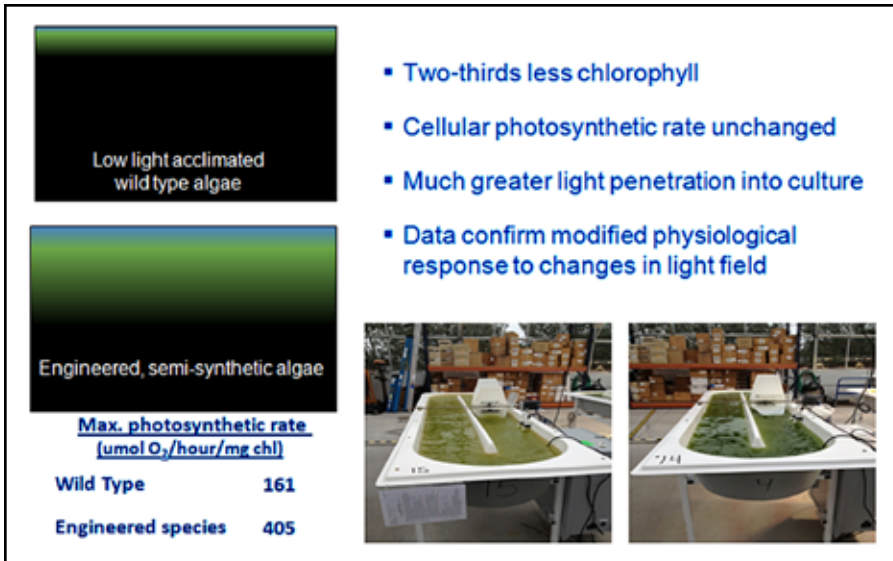


Figure 10. Engineering algae for improved photosynthetic efficiency: increased light penetration and improved photosynthetic efficiency.



JIM FLATT is the chief technology officer at Synthetic Genomics, Inc., a leader in the development and application of synthetic biology for sustainable production of fuels and chemicals and applications in agriculture. He has been involved in the industrial biotechnology field for over 20 years. Prior to SGI,

he was the executive vice president of research and development and operations at Mascoma Corporation, a leader in the development of cellulosic biofuels. Before joining Mascoma, he served as senior vice president of research for Martek Biosciences Corporation, leading the development of nutritional fatty acids from microalgae that are now included in many infant-formulas and other food products. And prior to Martek, he was involved in microbial biotechnology research at Merck and Monsanto.

Dr. Flatt received his undergraduate degree in chemical engineering from Massachusetts Institute of Technology and graduate degrees in chemical engineering from the University of California-Berkeley and University of Wisconsin-Madison. He has served as chair of the industrial advisory board for the National Science Foundation Engineering Research Center for Marine Biotechnology at the Universities of Hawaii and California-Berkeley.