Workshop Summary Reports

Beyond Containment
Assessing, Testing and Demonstrating Safety on Release of Synthetic Biology Devices and Chassis
An NSF SynBERC / MIT PoET / Smithsonian Woodrow Wilson Center Exercise

Woodrow Wilson International Center for Scholars
Ronald Reagan Building and International Trade Center
1300 Pennsylvania Ave., NW Washington, DC 20004-3027
January 13, 2010

CONFERENCE NOTES VERSION 2.0 PREPARED MARCH 11, 2011

TABLE OF CONTENTS

PARTIALLY ANONYMIZED NOTES

PREPARED PRESENTATIONS	NAMES ARE INCLUDED
FACILITATOR COMMENTS AND QUESTIONS	NAMES ARE INCLUDED
PARTIPANT COMMENTS AND QUESTIONS	NAMES ARE NOT INCLUDED

RULES FOR CIRCULATION

ALL CONFERENCE PARTICIPANTS WILL RECEIVE THIS VERSION OF THE NOTES (VERSION 2.0).

REFERENCES TO CONFERENCE PROCEEDINGS SHOULD BE BASED ON THESE NOTES.

Session I: Focus on Lumin biosensor as example of anticipated accidental release of device 9:00-11:00

Key Questions: What risks are associated with this early application of synthetic biology? How might risks be tested, reduced through redesign? How would the application fare under existing regulatory and insurance requirements?

- Briefing on Lumin device – pathway, chassis, physical device, methods of use (Mukunda)
- Briefing on conventional and life cycle methods of risk assessment (Dana)
- Critical presentation on MIT exercise on: (1) risks associated with Lumin device with JM 109 or rE.coli chassis (2) design of tests and demonstrations on areas of uncertainty 3) redesign of device to mitigate risks
- Assessment of existing regulatory standards and insurance standards using Lumin device as a test case / discussion of fit between these standards and benefits and risks of this application
- Assessment of public perceptions of risks and benefits associated with Lumin device

One of the goals of this conference was to bring together the technical, regulatory, and policy worlds at an early stage of product development, in this case the Lumin arsenic test kit, so that concerns might be addressed earlier and during the development stage.

In the first session, framed by Lumin's device to detect arsenic, Gautam Mukunda, CEO of Lumin, presented both the problem of arsenic contamination in Bangladesh and Lumin's efforts to create a low cost, low skill, accurate field test kit. The current design uses E.coli, lactose, and bromothymol blue where the presence of arsenic causes E.coli to break down the lactose and change the pH of the water, resulting in a change of the water color overnight from blue to yellow. The system uses parts that exist in nature; what is new is the combination of them together in same strain of e.coli. The test device uses JM109, a commercially available strain of e.coli, engineered so that it cannot survive outside the laboratory, and has been mutated to prevent the transfer of genes outside of itself. The device is still in the early stages of development and Lumin is a startup with many hurdles to overcome.

Genya Dana, from Dana & Sharp Risk Associates, spoke about approaches to ecological risk assessment for synbio devices and chassis. Currently, ecotoxicology methods have a limited ability to assess how living organisms will interact with complex ecosystems, as they are primarily based on dose response tests to determine toxicity of point-source chemicals and pollutants. She is proposing the application of new ecological risk assessment (ERA) methods to synthetically engineered organisms—methods which are ecological systems-based and participatory. These newer ERA methods have been tested in evaluations of other living organisms (e.g., non-native species and genetically modified organisms) introduced into complex socio-ecological systems, and they seem applicable for synthetic biology applications. A comprehensive risk assessment of synthetic biology application could begin by examining the most likely ways in which the stressor (the engineered organism) will enter the ecosystem and different exposure pathways. One such pathway for the Lumin biosensor could be at the disposal stage. In performing the problem formulation for an ERA, one should start by identifying stakeholders, defining the boundaries and scope of the analysis (e.g., the geographic region where the analysis will focus), developing conceptual models of the socio-ecological system, defining assessment endpoints, identifying human practices that may influence how the stressor moves through the system, and identification of hazards (i.e., interactions between the stressor and system components where there is the potential for harm). The next step of an ERA would be to conduct a risk assessment (determination of likelihood and consequences) of priority hazards. The risk assessment step will be difficult at this stage in synbio's development because there are not a lot of data on how frequently synbio applications would interact with different ecological entities and what the consequences of those interactions would be. However, *hazard identification* is possible, especially in the case of the Lumin biosensor where the developers have a specific receiving environment in mind.

One focus of discussion was on the disposal of the device. What will humans do? Will they leave the tubes on the ground? To answers these questions, you have to get out in the community. But that costs money and time. A human practice requirement would likely be prohibitive in the development of Lumin's biosensor. There are

tensions between “design is not final, so we don’t know how it’ll be disposed of” and request for information on disposal process.

What will happen if the organism enters the environment? Yes, every component already exists in nature, but what happens when they are combined. Will the new organism act the same? We don’t know. How does the e.coli strain change in the field? Is there a risk of it getting stronger? What happens when it dies?

Risk assessment imposes costs in both money and time that are likely to slow innovation/penetration. Poor risk assessment, however, may result in harm/lawsuits. The field also has larger concerns about its reputation since it is so nascent.

People agreed that this type of meeting, with a product in the early design phase, was useful to development. Some questions left to answer: When to involve stakeholders? For what role? Are there faster avenues to provide assurances of environmental safety? What are the current regulatory standards and insurance standards?

General Themes

- Significant advantages to focusing on safety problems at early design stages
- Significant advantages to companies able to do so
- Regulatory costs are likely to be difficult for small startups
- Lack of familiarity with synbio is likely to increase the unknowns and difficulty of assessments
- Setting boundaries and scope of risk assessment will be important and difficult
- Understanding human behavior and usage of the product will be key to any analysis – likely to require extensive field testing and experience
- A simplified regulatory process will be crucial if the field is to develop in a truly innovative fashion – pharma-style regulatory barriers will stifle innovation
    - Are there design approaches that would create regulatory shortcuts?
    - Does the government have the capacity to understand/create such pathways?
- Successful development of the field will require public support – this incentivizes the entire field to support first-movers in their attempts to do responsible risk analysis
- Development of the products is inevitable – where they will be developed is not.

General Open Discussion Notes

Looking at disposal of device - will it burst, drain? Who knows where the tubes will go after they are used for arsenic.

What are the dollar and time costs? In a similar situation in South Africa, the facilitator was the largest cost. There, a 4-day, 22 person participatory ecological risk assessment test case for genetically modified crops cost about $30,000 USD (facilitation was included in this price).

How do you do human practice identification? You have to get the data from the community – talk to people or watch people.

Does EPA have contractors to use to collect this data? There are government restrictions. Often do have to turn over to contractors, can be difficult – things get lost in the translation. EPA makes extensive use of contractors for many things.

You can't understand, you can only appreciate the assessment of the system. Any thoughts about the future of acceptance of ERA in regulatory, legislative, or legal sense? Scientific community is slow to move. Trying to get information into hands of decision makers.

What stage do these things get integrated into project design, when to start thinking about them? That's a question we are all here to think about. In Lumin's case, this is a good time to have that conversation. Get feedback from developers if they can address these things.

Human Behavior? It's still early for Lumin to look at this. First get a system that works at price point. Not finalized. Then launch into human behavior. Experience at McKinsey – even on large projects, very little focus on human behavior. When device leaves factory, no idea what happens to it. If it's a requirement to do the human behavior, it won't happen.

Will people even adopt this device? In this case, we are creating a cheaper and more accurate application. So people are already using something.

Gov regulator side. In US, regulatory agencies don't generally have discussions before the products are done. Some way to bridge that gap.

Looking at issue of fracking shale. Listening sessions set up with public. Integrated, hopefully, with environmental RA. Novel way that things are being done.

Open question on the issue of when an extensive RA would be required for Lumin. Tension between specific application, context, human practice and between market practice.

Are there shortcuts that can be taken to provide assurances of environmental safety?

Idea of everything exists in nature already. Still needs tests. Actual DNA sequences – raised more questions. In a few cases in the past few years, when an ERA has not been done, there have been lawsuits.

Idea that the natural pieces of the e.coli are the same as the naturalness of the combined. Don't know if new synthetic organism is going to act the same. If someone builds on Lumin's strain, does that change it? How do we classify the strands? Will we require a new assessment each time it is changed?

Have you tested the strain in the field? When you change the combinations, bring it out in the field, it's interacting with many different systems. Thinking about other products. Need to avoid the thinking of "well it's a naturally occurring thing, so it's okay"

Natural e.coli is more robust than the Lumin strain. This strain is designed to make it less dangerous. If design changes are made to make it safer, then should we have a lower level of scrutiny?

We don't know how these new things interact. Don't want to have the Jurassic Park situation. Releasing these things into the natural environment, and you can't pull them back.

Misleading term of synthetic biology. Pieces of natural system, engineered to do something I want them to do. RA in laboratory. Who is doing the experiments? Government funders and regulators should be concerned with this. Scientific piece is important. Whose role is it to support that?

Terminology. Important that lab exam is done. Translating that lab work to field is hard to do.

Natural can be harmful, ex. Change soil/bug environment and wreak havoc.

Who cares about the naturalness? Can you take a shortcut – since the strain won't survive?

Do organisms tend to evolve back to original (vs. engineered organism)?

Barriers to entry slows innovation/penetration. Hidden costs to regulation. How can we look at things at the beginning of development and head off any negative effects?

Once development space has so many RA requirements, less innovation – usually copy/slightly alter previous work. Initial science is critical. Science should be done with public support. One way or another, the products will be developed. Maybe not in US.

RA changes to commercial risk assessment. Need to initially ask permission from gov. Ex. Of pesticide application adopted because of fuel cost savings. Did not know this until it was developed.

What tests would you like to see done to be comfortable?

May be costly to follow the recommendations. Failure can lead to mistake. Initially may cost more, but long run is better. Really complex systems. Not just what happens to Lumin, but whole field.

2 questions: What's the organism? What's the receiving environment? Are you increasing the fitness of the organism? If it picks up another microbe's genes, can that increase fitness of the organism?

Way technology is deployed is important. Regional value? Socioeconomic. People have to understand how to use the device. Have to have a plan on how to educate the people. Popularize the device. Should be a standard disposal mechanism – should be transmitted to the people. Many people live in the rural areas. May not know about arsenic problem.

Things to look at: survival of e.coli. What happens when it dies? What happens to the DNA? How does it die?

This gets into values of stakeholders. Is there any harm? Issues of values become important.

Session II: Focus on rE.coli as example of chassis re-designed to mitigate safety risks on release11:15-12:15/13:15-14:15
Key Questions: To what extent can risks be assessed and safety improved through design of chassis without specification of pathways? With what implications for regulatory and insurance standards and for costs of compliance?

- Briefing on chassis functions, principles of safe chassis design, current status of rE.coli (Church/Carr)
- What are ideal properties of a chassis for safety? For other chassis functions? How can such properties be designed into chassis? How can properties be tested with technical fidelity and demonstrated with credibility?
- Assessment of risks associated with chassis designs as distinct from device designs / design of tests and demonstrations to address areas of uncertainty / redesign of chassis to mitigate risks
- Assessment of existing regulatory standards and insurance standards using rE.coli chassis / discussion of fit between standards and this application. Does the novelty of chassis design that enhances safety also increase the regulatory hurdles that must be cleared?
- Assessment of public perceptions of risks associated with chassis, with reference to E.coli as object of discourse.

Chassis Construction

There are various containment methods for keeping a GMO from surviving outside of the lab and transferring genetic material to and from other organisms. In particular, six safety components for synthetic biology were presented at the meeting. They are:

1. Physical containment (e.g. BSL1-4)
2. Genetics containment
3. Nutritional containment
4. Orthogonal antibiotics
5. Training & scenario brainstorming
6. Communication & transaction surveillance

In genetic containment, the chassis's genetic code is modified to prevent horizontal transfer with other viruses and microbes. By having a unique genetic code that differs wild type organisms, DNA that escapes the chassis could be stable, but would not be able to code for anything in the wild. Likewise, foreign DNA taken up by the chassis would not code for something functional. Genetic containment therefore prevents the chassis and wild organisms from correctly interpreting each other's DNA. It is easier and more robust to create organisms whose code is nonsensical to wild organisms rather that to create organisms that are chemically different.

In nutritional containment, the chassis is engineered to depend on novel amino acids rare in the environment. Eventually, genetic and nutritional containment will be robust enough such that physical containment would not be necessary.

In addition, surveillance should be broader then just commercial. Government agencies have experience with commercial surveillance, but academic surveillance is also needed.

As of right now, these modified chassis will probability interact with their environment and other organisms in a similar manner when compared to wild organisms. This is because the modified chassis differ only in their genetic code. But there are potentially additional ways to further modify chassis to create behaviors that differ more from wild organisms.

More individuals should be included to the conversation, especially those who have participated in biological safety meetings and those who focus on safety in all types of technologies.

rE.coli as the Solution

rE.coli seeks to build the containment into the organism or bug itself.

Genetic transfer relies on a comment language, as the code is living organisms is nearly universal. If organisms don't have that comment language, the concept of transferring DNA and gaining new functions is essentially lost. In transcription and translation, the DNA language is converted to protein language, which species lots of cellular functions.

If a different language used the DNA/RNA, then the tRNA will substitute the wrong amino acids into the protein, resulting in a protein with the wrong or no function.

In rE.coli one of the synonymous stop codons (UAG) is removed, which is the first step to creating genetic isolation. rE. coli will no longer need the UAG codon.

Potential trouble spot: The computational aspect of on genome design is relatively simple, the bigger the limitation is how well annotated the genome. It is difficult knowing for sure where all the stop codons are. The genome is currently heavily documented, but there are still unknowns.

If a wild gene from outside gets into rE.coli and is incorporated into the genome, what is the fate of that gene? If the gene has a UAG codon, it will likely code for a non-functional protein. If the gene gets integrated into the bacteria, over time, it is possible that single change can mutate from a TAG to some other form of a stop codon, and the gene would then be produced correct, or into the TAG would mutate into coding codon, allowing translation to continue until it meets another stop codon. Re.coli 1.0 is a stepping stone to more complex genome.

The rE.coli designs welcome scrutiny. It can be when the scrutiny doesn't come to a conclusion, which is often the case, it causes people to have an adverse reaction to scrutiny and being scrutinized. But if you have a high level of scrutiny on a novel mechanism and it swiftly comes to a conclusion where it works, that may in turn invite more people to be open to the idea of scrutiny.

The hurdles (relating to the Lumin device) are much more regulatory then biological in terms of what they can predict. You can't predict everything, and that is especially true for new and novel things.

Risk Evaluation

Having a safer chassis can give some reassurance on some issues. It can reduce some risks and reduce their associated costs. For the remaining risks, a risk analysis is still need.

There is some uncertainty in determining the current regulatory standards if the chassis is used. The novel approach to create increasingly exotic bugs to be housing stuff that can improve safety, but at the same time the artificial characteristic of the organism is increased as the organism is further engineered.

How will the public perceive the chassis? The public has a negative impression of GMOs and E. coli. The rE.coli chassis is based off of e.Coli. How can you distinguish the difference between the two? Is it a question of missed perception?

The goal strain has not yet been created thus the strain doesn't exist in nature as of right now, the conversation is hypothetical. The designers have a strong reason to believe the strain will increase safety, but has not yet been proven by experiments.

How would rE.coli be tested? First test would be to see the ability of individual bacteria to infect the new strain and the old strain. Need to account for growth, if there is any difference in the growth speed between the two strains.

More sophisticated tests would be to see if you can get to conjugation between bacteria, and whether you could have a very complex ecosystem physically contained in the lab to see if you could get selection for survival under circumstances that you wouldn't accept the survival of the strain.

RE.coli has 2 types of categories grouping

1. Genetic code minimization, where you are reducing the genetic code, where you are eliminating things
2. Reorganizing the genetic code.

That creates two different broad classes of properties related to rE.coli but distinguishes between them. They are not trying to put in any new gene functionally into rE.coli, they are trying to reorganize. They are not changing the proteins themselves. The amino acids are the same and the overall gene functions are intended to remain the same.

One opinion: Will never get a chassis or strain to the level where a case by case review is eliminated. Some understanding of the base organism will be gained from using it in different applications. But adding different things to the same chassis can lead to different consequences. Thus, you would still mostly likely have to do a case by case review. Can chassis share the DNA with each other?

One way of evaluating the effectiveness of rE. coli: It is very easy to monitor different types of gene function and whether the function is being demonstrated partially or completely. A simple example is color metrics of fluorescents, indicators of a single gene level. And then working your way up the hierarchy would be to then move on to more complicated populations and larger populations resulting in combinations of very small probabilities but with potentially very large population numbers. Looking at the Lumin example you're probably not going to see more then 10^9 cells, so if you have probabilities of failure modes that are mismatched from that by several orders of magnitude. But in regards to a large scale bio reactor you have several orders of magnitude more cells so you may have different criteria for that, relating to bio remediation.

Nature of Risk Assessment

These risk assessment are going to evolve incrementally/ change, the beginning ones are going to take bits and pieces of this depending on who the regulators are, and they ask for a wider beginning or may ask for a smaller beginning and they are going to work their way up from there. Several points: 1. The results of the risk assessment differ from the construction of the risk assessment. 2. There are only two reasons for doing risk assessment, 1 you're dealing with ignorance and 2 you're doing this to better understand those factors. You would have a reasonable deterministic decision possible and sufficient information would be available in order to make a decision. The thing about risk assessment is a series of things, trying to convince people it is necessary, then convincing people they are sufficient, and then having achieved the later trying to figure out what constitutes the degree of experiment that would actually lead to that degree of sufficiency. One's a scientific problem, the other is a policy problem that both cannot have an absolute answer.

The level of testing to support a risk assessment may be very different based on what is identified as what is valuable. Who are the stakeholders, public worry relating to device/ chassis and the ecological ramifications for the microbial system, all of their factors and ideas of what is valuable differ.

Evaluation: you have a feature you want to look at and you propagate the cell over many cell divisions and you look at when the feature performs the same. One of the concepts the many people are interested in is the robustness of the cell, what is the stability of the cell over time. The question is not if the cell is going to break, but more of when it will break. When it comes to controlled systems and fail safes, how long do you want the system to run for and then how long it can survive past that point of time. In the long term would like to see applications that are characteristics as sufficiently safe based on the wonderful benefits that they can provide. Starting something "a bit more realistic" like the Lumin device/ application can potentially pave the way for more futuristic applications, such as tumor eating bugs and intestinal watch dogs to fight off pathogen, as of right now that is still fairly academic.

The stakeholder should be the planet, break it down into humanity and the ecosystem in which we live. Starting to think about it that way for now you can set that as a threshold, which in some ways makes things a bit more absolute, bound and clear then identifying all the subsets that you think of as a stakeholder. When it comes to micro-organisms that are likely to be released, either in the extremes of being released into the environment or if it's released into the body, then pretty soon after it's going to follow the same path as if it was released into the environment.

The use of the chassis for contained uses doesn't feel it's going to raise as many difficult issues at it would in regards to bio remediation, dealing with issues of release to the environment and survival for an amount of time. A category of applications and you bring to a group of stakeholders from the science community, policy makers, NGOs who are concerned about these issues, not to do a risk assessment but to come together to try to define what are the key questions that are going to have to be answered. Later, when you start to get into specific applications you could do that again but could focus on the specific application itself, since the other broader questions have been defined. Creating a kind of tiered approach, because of the way it looks now at a very abstract level one can't really answer the questions fully. You have to look at what questions are answerable at each level.

Communicating the Results to the Public

Within the conversation several words that came up have meant different things to the different stakeholders especially to the public (example: chassis, chassis short cut, safety chassis, artificial, synthetic.) Coming up with a common langue, defining the words in relation to the conversation and sticking to using those words in the tense/ sense/ ideal you have preemptively defined them will help with educating your stakeholders and general understanding of the more technical stuff.

In regards to wording and public perception, the general public doesn't like to be confused especially when it comes to new technologies. There seems to be, within today's conversation a conflict in wording and the definition of those words. Example, the chassis/ engineered strain is wimpy (because it's a weak strain of e.Coli) but robust (because it has been altered to resist viruses). If one doesn't understand the science or technology of how theoretically it can be both, then you are sending out conflicting messages. It's hard to conceptualize the language/ rhetoric when you're not familiar with how it's all constructed.

There should be an emphasis on viruses and resistance, because viruses carry stuff. They are an organism that can transfer genetic material, so one would want to create a chassis that is virtually resistant, natural to the chassis you are putting together, but you also want one that is robust enough to serve as a "factory" for devices such as Lumin's. One that will be strong enough, that will prosper and survive to get through the pathways, do the job needed. The issue of survivability, as you move from industrial contained applications to the bio sensor on one hand to bioremediation and maybe even medicine and agriculture on the other, it has to survive. Moving the strain away from being "wimpy" type to a "stronger" type because you want it to survive but you want it still to be containable.

How do you actually bring stakeholders up to a level of understanding, where you don't have to be able to do the lab work but you do need to be able to ask the questions and not simply take things on faith?

Perhaps the key to that question is ignorance. Three categories of potential anti-people that have to be thought of, those who have moral abjections to this type of work, those who are "faith driven" (not in the religious sense) but who are in general fearful of this type of work. In theory education isn't going to work on them and those who are ignorant in the "they just don't know about this" sense. The latter of the three is probably the best group to shoot for in regards to educating. Idea of educating them when they are young, giving them the tools to have the conversation with the "older generation" of contributors who may be set in their ideas and ways when it comes to new sciences and technologies.

There are different reasons for resistance and concern that isn't simply because of ignorance for some. Environmental groups who work with groups around the world have concerns for very real potential risks to environmental safety, public health safety and social economic issues. It's not enough to say "well if you educate everyone then all those concerns are going to go away." Looking at what we've already learned from the conversation earlier in the day and what more we will learn as the conversation goes on, that there are some very serious and difficult issues relating to bioremediation and others. The more we learn about the science and the technologies, the more we learn more about the risks that one might be able to identify. It's not enough to just say it's a matter of "not knowing" or

"ignorance"; it's more about the knowledge that we've gained on the subject, the more concerns that can be/are raised. And it shouldn't just be dismissed by pure ignorance.

Comments on "ignorance" can be processed. First, looking at the role of a GMO engineer trying to educate people on the science/ technology and also looking at the role of a NGO, really trying to understand the points that the GMO engineers are trying to make in order to give a good critic from an informed perceptive. This question was brought up because the dynamic/ character of the conversation changed as we moved from the earlier session, discussing a topic that is a bit more translatable, to this session where some may be tentative to really poke and point out issues, chime in and ask questions about a topic/ science/ technology that isn't as understandable because one doesn't understand the science behind it.

Is there a slightly mistaken assumption that in order for one to engage in this conversation requires technical knowledge? To do the latter parts of the risk assessment, to answer some of the question in regards to risk and to understand those answers during the risk assessment may take technical knowledge but that in theory locks you into looking at only the technical question. If you look at the conversation of risk assessment as one part that sits within the bigger picture of risk management, which also can include the question of what are your concerns and what can be done to take can of those concerns. Not just saying "this technical reason takes care of those technical concerns," which can lead into big arguments about "how can I trust that technical answer if I don't understand the technology." Looking at risk management procedures, for people who don't have necessarily have the technical background but do have concerns. Their questions are "don't tell me all the reasons why I'm not going to get sick, tell me all the reasons, ways and what you are going to prevent that organism from getting out so I don't get sick." This directs the developer to thinking about ways to carefully run field tests and so on that don't necessarily require the technical knowledge of all the stakeholders. Creates an environment where the developer is taking these concerns into account and is addressing them and the other stakeholders can see how their concerns are being addressed. One should almost take a step back and look at the bigger picture, viewing the exercise from a risk management perspective when engaging in a multi stakeholder's assessment process

How do you know that you have all the right or correct answers when it comes to public policy? You don't want to come up with the wrong answers ending up with dead people. Look at all the worse case possibilities and systematically address them as best you can, finding out whether it's plausible to down size risk that has been missed by people doing best case judgment. And there are sometimes people who are promoters and some agencies best case judgment, that really want to believe that, that type of risk assessment will work without down sides.

Should the public need to know about the parameters that they are interested in? Technical details not everyone will be interested in, but would like to know some of the basic questions, "survivability, danger." If there is a way to somehow quantify that, if a hypothetically regulatory agency could come up with a measurement/diagram to show the level danger is if it gets out, to what degree does "it" have the ability to survive. In theory to someone who doesn't care about the technical answers, having those key parameters defined is more than enough information for them.

Also look at the ones who are fairly well clued in on the technology/ science that have their own concerns. They would like to aim for the technical knowledge not to be a high prerequisite for everybody. If the experiments are clear and list all the things that people are concerned about and test those concerns in a physically contained environment, that goes a long way. For scientists checking out every phage they know of that effects E. coli, is fairly reassuring that really the DNA is not functional because phage represents huge diversity. But for other people the experiments to put the e.Coli in to a complex environment including those phage and other things and determining at what time do all of the foreign DNA disappear is something that is fairly intuitively graspable and verifiable at other laboratories and so on. Process of being verifiable in multiple physically isolated laboratories will be a big step forward.

Further Tests and Regulation Framework

For all of the novel features built into an organism there will always be a few novel tests that will be required. Communicating to the public is to say "okay what are the normal tests to do to evaluate whether or not something is dangerous?" (Those with the engineered organisms) they would perform the same tests you would normally do if you were testing something else, ex. the release of a natural microbe into a different environment then the one it was found in. Communicate to the public that you've performed all the tests that you normally do to see if something is safe. If you've passed all those tests, then you have reached "this" one level (in safety) and then on top of those tests you run the special tests that you've designed just to test the engineered microbe. If you passed both parts then you can say that you know even more about the safety of this "new thing" then you do about any "normal" thing. Have to make it clear that you've done your due diligence – you have to have gone over and beyond the normal threshholds.

Taking the initiative: If you can determine which agency you think you need to go to in order to coordinate the frame work then you can begin the process of telling them what you are going to do (relatively small notification process.) Where you would describe where you got the gene from, where you're sticking it and what you think you are going to do with it. At that point determinations can begin to be made. If you walk into the agency saying "tomorrow, you would like to spread this "bug" over 50,000 acres" the door will be closed probably indefinitely. But if you're describing your intentions at this early stage, it's different. The concept behind an MCAN and a EPA it's defined as a heterologous, there is a technical definition what heterologous is and there are 5 page within the guidelines explaining what it really means to be heterologous. Within that text there are a ton of exemptions. E.Coli and some specific sub-classes of it do fall into those parameters, but the end point is you still have to ask to find out where you are within the guidelines. The general frame work is that if it's truly a heterologous gene, there will be someone with some sort of initial regulatory authority.

One interesting area is where even prior to synthetic biology the kind of "let's just go with the terrains of the world, the detergents of the world" extracting DNA from the soil, cloning it and sticking it back into something. The earlier stages of synthetic biology but they didn't know where the gene came from. Beginning to be able to answer the question "where did the gene come from" lets you in on where it may be going.

Looking at pesticides, basilious conjugate. If you moved a heterologous gene from Basilious 1 to Basilious 2 by recombinant DNA technology, you're not necessarily defined as recombinant because it could have happened by conjugation. "Could" being the key word.

Right now most regulatory standards that one could look at as an example of how to regulate the safety of SynBio is based off of chatter that was developed 30 years ago rooted in biology. SynBio has the opportunity to create some public/ reviewable/ sight able safety literature on safety on its own that then gives a regulatory body something to look at beside the confidential business plan and such.

Session III: Tagging key sources of uncertainty and tasking next stage research 14:30-15:30

Key questions: Of issues not resolved in morning sessions, which disagreements on assessments of risks, design of tests, and redesign of devices are rooted in questions of values? Which disagreements are grounded in uncertainty over specific empirical issues? How might these issues be addressed?

- Questions to / from environmental microbiologists and risk analysts?
- Questions to / from synthetic biologists?
- Questions to / from regulators and insurers?
- Questions to / from civil society?

Todd and Ken posted sheets with unresolved points collected from Sessions 1 and 2 on the walls of the room.

Facilitator: The purpose of this session is to review unresolved issues flagged in our earlier sessions and discuss how they might be addressed. Some are included in the posters on the walls. Take a moment and look them over. What points are most important? What important unresolved issues are NOT captured on the posters? Then we can talk about which sources of uncertainty we should focus on.

Added Unresolved Technical Issues including fitness and gene flow of JM109 and rE.coli with pathways.

Does the combination of chassis and pathway increase or decrease fitness? We don't know but need to know before we can judge the risks.

Interaction between genetic material. Gene transfer issues need to be understood better. This issue is being partially addressed with rE.coli, but what about other chassis? We need to know more about this before we can feel confident.

We also need much more clarity on what we mean by harm, with emphasis on how these issues affect human and environmental integrity.

What about the dead? The emphasis on survivability and gene flow in living organisms has deflected attention from another set of issues. If the microbe dies, what happens? What happens to genetic material in dead organisms?

Added Unresolved Regulatory and Behavioral Issues

We need public funding to attack policy relevant sources of uncertainty. Research on policy relevant sources of uncertainty should be in public domain. When uncertainty is identified, agencies should be involved. But we don't have many examples of funding.

Research should focus on benefits as well as risks. For example, neither benefits nor risks of the Lumin device have been well established.

The absence of clear regulatory structures in South Asia and the US for dealing with these problems contributes to uncertainty. We lack credible criteria for defining danger and safety, and that may inhibit effective identification and management of risks and inhibit technology development and diffusion.

We need to understand human behavior in practice in order to evaluate and manage risks. For example, as Genya noted in her presentation, how users will dispose of the biosensor will have significant effects on risks. How human behavior is incorporated into risk models is a key piece of the discussion that is not well understood.

Affordability is critical with respect to the Lumin device. This device has to be cheap. But we do not know how cheap. There is considerable uncertainty in the economics of the business plan.

Facilitator: Let's review the points in Todd's chart from the chassis session. What kinds of tests do we need?

With reference to risk assessment, we need to get policy makers comfortable with the approaches taken to testing. We lack real life examples where policy makers and stakeholders are involved in process of designing tests of safety.

How do we design test processes for a specific product such as the Lumin device? How for a generic chassis?

Facilitator: What of these questions are both key and answerable?

Note the specificity and clarity of the risk discussion in the first session. And note that we had trouble grappling with risk issues with reference to generic chassis. The reasons why we have difficulty engaging on the risks associated with chassis are worth probing.

What are the differences/similarities of GMOs and synbio? What can we learn from mistakes with reference to GMOs? How can we avoid repeating those mistakes?

Facilitator: If you were to pick one or two of these topics as most significant, where would investment of effort lead to better understanding?

Survivability/fitness of bug/what happens after death.

Still struggling with the notion of what stage of the process should risk assessment and stakeholder engagement occur? Should this take place with very early applications of technologies?

Go for stakeholder involvement and risk assessment at an early stage. Lumin is specific. The chassis poses challenges. Is it too early for rE.coli?

What of potential risks raised by stakeholders? What would you do with feedback? Can we try to specify what kind of research would help advance our understanding of potential risks? Don't just make a case for posing a question. Discussion should focus on how to answer questions.

Design of processes for stakeholder and regulatory engagement in design of testing methods. This is a sweet spot.

Facilitator: Let's not just talk about what needs to be done. Who should work on models of stakeholder engagement? Who should work on developing methods of developing testing methods? Who is responsible for moving on this? Can we give that job to someone in this room?

The chassis re-design group is starting to think of these details upfront. Perhaps these issues should be built into RFPs for grants. This may be an NSF function. It should be embedded in technology development, with funders and scientists involved up front. How do I identify these issues as the technologies are developed? What are the sources of uncertainty over risks and benefits? How are you going to test for safety? How are you going to address risks identified?

This needs to be done with public funding. Or by companies. It usually falls on the public to do the testing. When you have public money going to research, the public needs to get involved with evaluation and testing of risks associated with research. You might not be able to test for the things that you want to look for, so that testing technology has to be developed at same time. No lag.

Facilitator: Why would companies want to hold a stakeholder exercise? How can it be incentivized?

Some incentives exist for companies on the liability side. Companies are afraid of the consequences of something bad happening.

We need insurers in the room for this exercise. Their perceptions on risk and liability would be good. Insurers are concerned about this, especially because of the potential for serious and even irreversible harm. Such concerns translate into premiums. In another exercise, we must have the insurers involved. The insurers are intensely concerned because of their exposure. Industrial applications are contained. The uncontained applications raise more complex issues.

How do insurers evaluate risks? In areas of limited experience, insurers have to guess, they want input from researchers. This is the basis of mutual interest. This is positive. It is figuring out what areas should be excluded from coverage to limit risks, what areas should be included with what premiums to make money and what conditions on coverage should be imposed. Insurers that can make these decisions will have a competitive advantage. In doing that job well, they may also serve public interests in management of potential risks and in

incentivizing research on assessment and testing of risks. If a company wants coverage in an area with potential risks, then who should bear the burden of demonstrating safety?

Work on risk assessment methodology would be useful. We need a factual basis for this undertaking. But one size may not fit all. What methods and rules for what circumstances? How to design so that there are rules that allow you to decide what portions of risk assessment would apply?

Facilitator: Do we need a menu or a flowchart. Who would do this?

Like a Pre-flight checklist? We need to focus on higher risk areas. In the Lumin case, disposal methods and risks. This might be helpful for small companies vs. large companies with independent analytic capacity.

If company doesn't do risk assessment because it doesn't have money, this could be bad. The external costs of not doing due diligence are substantial. Third parties may have interests.

What happens when goats eat the biosensor? This level of risk assessment should also be done. Investors should recognize Lumin has limited resources for risk assessment.

A lack of proof of harm does not equal safety. Where does burden of proof lie? We need to ask the right questions on risks and benefits. Consider the example of a Synbio company, created as private company in US then moved to Brazil. What of the regulatory challenges of US companies working in another country not covered by US regulations? There is the potential for export of harms. And we need to treat the socioeconomic. And issues of liability. We should not look to the precedent of liability with GMO's and crops as a guide. There are issues with that. It is too hard for farmers to secure compensation for damages from contamination of their non-GMO crops.

No absolute answers. Science deals in probabilities. That has to apply for this too. There is no guarantee of safety. No endpoint.

Looking at antecedents is actually good. We can see what really happens. Who actually gets involved? Is it useful to have forms of structured early engagement?

Facilitator: Please send us topics you would like to see tackled.

Please talk to EPA now to see what their questions and concerns are, and make them aware of the work going on to address safety issues proactively. Please talk to other people/groups to identify their concerns. We have not covered everything, but unless you talk to people, you don't know what they are worried about.

Researchers have talked to EPA on photosynthetic organisms being used but not specifically about rE.coli.

USDA has 18 million a year for risk assessment research grants. This is small. But there is usually a 50% success rate in getting funded – there are not many applications. This is through NIFA. There is not enough money to do all of what needs to be done, but still this is a potential source of funding.

Please continue to integrate the social sciences with the technical work. We need to know better how to address societal concerns, regulatory and insurance rules, and technology development

Ask research funders to donate money to address these issues?

Session IV: Scoping exercise on bioremediation as example of deliberate release of biological devices 15:45-17:00
Key Questions: How do risks associated with larger scale deliberate release with bioremediation differ from the accidental release issues found with the biosensor? What is the status of current natural and GM bioremediation? How are risks currently assessed and regulated? What sort of an exercise using what biological devices might be useful?

- Memos on bioremediation of atrazine (Lin) and petroleum (Mohr)
- Natural and artificial objects of remediation / wild type and GM organisms for remediation
- Scoping exercise on current GM bugs and status of current risk assessment and testing methods
- Scoping exercise on current regulatory and insurance standards for GM and non-GM bioremediation

Facilitator: This is a preliminary session focusing on prospective synbio applications beyond containment, to talk over whether an exercise on more challenging applications makes sense. We have identified two contrasting potential objects of bioremediation, atrazine and petroleum. Atrazine is artificial, petroleum occurs in nature. Atrazine is one compound, petroleum is a complex and varying combination of compounds. Both present environmental and health hazards. We will turn to short remarks by Allen and Scott on the objects of remediation, the prospects for remediation, and potential risks. Again, the purpose of this session is to talk about whether we want to have an exercise on bioremediation and how such an exercise might be structured.

Allen Lin atrazine presentation – referred group to the clear summary provided in advance of meeting. The specificity of the target molecule simplifies the task of design. Cutting to the core of the summary, the redesigned bug became more mobile with atrazine and has the ability to degrade atrazine into non-toxic substances subject to further degradation through natural processes.

Scott Mohr petroleum presentation – Petroleum is complex. Can't clean it up with one enzyme. There is natural bioremediation in oil spills. But there is not sufficient nitrogen in oil for natural processes to work quickly. Need to figure out a way for the bugs to deal better with specific molecules and to have bugs suited to specific environmental conditions. Surfactants are important – could be added to bugs. Issues include human effects, Impact on ecosystems, destruction of hydrocarbons, and evolution of bugs after release.

Facilitator: What are key differences between the two bioremediation examples and the biosensor Lumin case treated in our first session? Let's start with a discussion of the risks and how to evaluate them. Then talk about the need for a possible exercise.

The bioremediation cases involve deliberate release of a bug that has to survive and multiply on release to do its job. Gautam's device is released by accident and need not survive. This open and uncontained element of bioremediation is the key. We need to think about these bugs in lakes and rivers and soil. Or underground.

Governments would not want a petroleum eating microbe near their reserves. Dark oils have been undergoing natural bioremediation for a long time. In fact, there is a recently identified naturally occurring bioremediation bug in China. What type of bacteria? Potential human health risks are more than e.coli. (See Scott's map of the organisms, red and yellow).

We need to think about policy and philosophical questions to set up our questions on specific risks. Atrazine has been banned in many countries. Bioremediation might legitimize continued use, albeit with harmful effects of farmers. Petroleum bioremediation underground may translate into a threat of bioterrorism in Kuwait.

The Kuwait issue is complicated. Like the BP release in Gulf of Mexico. There are microbes that are eating oil right now. Is there anything different about synthetic microbes that would generate more risks?

With respect to atrazine, a conventional risk assessment conducted by today's standards would not look at possible increased use of atrazine if bioremediation were ready to roll.

We know little about environmental processes. Shoreline processes, if these organisms make their way into natural environments and even machines like autos. We need to highlight work on oxygen demands of organisms.

This is an enormous analytical problem, with unclear boundaries on analysis. There is a broad range of social issues. Bioremediation may be good for one part of population, and not good for others.

We need to create remediation bugs with suicidal properties – that's more appealing. Health and safety can be addressed easier.

If terrorists decide to use a bioweapon (anthrax), more work would be done on a biological approach (rather than the current approach of physical decontamination. That work would be nice to see.

If Synbio organisms could do something really neat like helping to recover leaked oil instead of destroying it, and then shut themselves off, that would be really nice. Can make money with such bugs.

There are some grounds for optimism. There are analogies in nature, in our own bodies. Can we put them together in a way that's controlled? If one cell doesn't shut down at the right time, that can be cancer.

Facilitator: How much reason do we have to believe that we can do better than nature? To what extent are we exaggerating what the technology can do in the near term? Maybe we should say "Gee this is interesting. Maybe 10 years from now we should revisit it."

Not necessarily blue sky. This could be near term and useful. If we pick specific examples of the most artificial things, those are the objects of remediation where the need for synthetic organisms would be greatest. And such bugs could be designed to be good digesters of only one thing.

First things first. If atrazine is toxic, let's stop using atrazine. We will not stop using oil. And natural and artificial oil eating organisms are going to find a way to survive. I can guarantee that it's impossible to calculate all the risks. This is going to concern the public.

Because these are microbes? What if remediation was based on plants?

Public reaction is unpredictable. The videos from the bottom of the ocean with the oil spill was what captured attention.

Some biologically derived materials are now pumped into wells.

Suicide genes have some origins in 1990s to develop releasable bacteria. There is a moratorium on suicide genes.

Do we have examples of recombinant based remediation being used anywhere in the world?

Facilitator: As you look around at artificial stuff that is bad, that would be the logical target for bioremediation. Is there something that you'd like to see gone, that is produced by humans, and not in nature?

PCBs. Are their examples of synthetic organisms doing bioremediation of PCBS?

EPA has funded studies using recombinant technologies to do this, but the results haven't been that positive. The idea is there. There is an EPA citizen's guide to bioremediation.

You can use iron to go after PCBs. Do we need synthetic biology for these issues? With atrazine, what else are these bugs going after or will the bugs release toxins, creating secondary problems?

Problem formulation is important. What's the problem? What are the options? Let's look at pros and cons and evaluate the options and technology. This provides a structure and a basis for comparison. The problem is that we have to do something about the arsenic, not just that we need to know if arsenic is present.

There is an arsenic eating bug under development at iGEM, but we didn't use it for this exercise. No commercial company wants to invest because they think it won't pass regulations.

Nothing would stop the Federal government from creating synthetic bioremediations. It could get to a pilot level demonstration. EPA does some bioremediation studies, but we do not know of their precise status. Nobody knows specific examples?

If remediation strategies are site specific, you can tweak the bugs to limit adverse effects.

Imagine a scenario of a site you want to remediate, and the public gets alarmed "what are these bacteria? Are they coming into my property?"

Facilitator: What questions would you want to have answered before having an exercise on bioremediation? What info do you need?

We should consider shifting to biofuels. There is extraordinary effort right now, and it is close to reality. There is rollout coming. The exercise should focus on stuff close to rollout.

Facilitator: Which synbio biofuels applications are technologies whose time has come?

Think of the arc of commercialization. We won't see anything for a few years if oil prices drop. If oil prices rise, then technologies will develop and diffuse more rapidly. Changes to level of common ingredient in household cleaner. If biofuels can be made cheaply, then they will be. China, India. China is onto this.

Facilitator: Do we need risk assessment for real time design of microbes for specific combinations of compounds and environmental conditions, as Scott's paper suggested?

It is not going to happen.

Our first steps –we need to be able to describe the state of the art. (a) contact the research group at EPA that works on bioremediation. (b) contact DoD – munitions cleanup cases. (c) contact oil companies … they are sensitive to issue.

Facilitator: Are there other places we could go for info?

Talk to Security folks. Look at bioterrorism. If this is leaked into oil reserves? Bioremediation issues are very interesting. Talk to Joule with their advanced work on biofuels. These are next steps.

END OF DISCUSSION

Comprehensive Environmental Assessment and Synthetic Biology Applications Workshop
The Woodrow Wilson International Center for Scholars, Science, Technology & Innovation Program
Ronald Reagan Building and International Trade Center
1300 Pennsylvania Ave., NW, Washington, DC 20004-3027
Thursday, July 28th, 2011, 8:30am to 4:30pm

CONFERENCE NOTES VERSION 1.0 PREPARED AUGUST 2, 2011

TABLE OF CONTENTS

PARTIALLY ANONYMIZED NOTES

PREPARED PRESENTATIONS	NAMES ARE INCLUDED
FACILITATOR COMMENTS AND QUESTIONS	NAMES ARE NOT INCLUDED
PARTICIPANT COMMENTS AND QUESTIONS	NAMES ARE NOT INCLUDED

RULES FOR CIRCULATION

ALL CONFERENCE PARTICIPANTS WILL RECEIVE THIS VERSION OF THE NOTES (VERSION 1.0).
REFERENCES TO CONFERENCE PROCEEDINGS SHOULD BE BASED ON THESE NOTES

About the CEA workshop (Opening Remarks)
The goal of this workshop was largely to try and generate a set of questions and research areas that would enable a more comprehensive risk assessment of synthetic biology applications. To do this, participants are asked to consider a scenario where cyanobacteria (*Synechococcus elongatus*) engineered to produce sugars at an industrial scale have escaped into the surrounding environment. Participants were asked to follow a Comprehensive Environmental Assessment (CEA) framework currently used by the EPA's National Center for Assessment (NCEA) to evaluate risks associated with nanomaterials, and consider each component of the CEA and discuss potential knowledge gaps and research areas for synthetic biology applications.

It is important to note that this was NOT a risk assessment of *S. elongatus*, but rather a test drive of the CEA framework to gauge its usefulness AND to generate research priorities to enable comprehensive environmental risk assessments of synthetic biology applications.

Major questions that this workshop aimed to answer were:
- Is the CEA framework useful for understanding risk and supporting risk assessments?
- Can we create a frugal tool or decision tree that can guide researchers in dealing with the ecological issues of synthetic biology applications?
- Will current risk assessment methodologies work for synthetic biology applications?
- What information/data would we need to be able to accurately assess the environmental risks of synthetic biology applications?
- Is there a set of research areas or questions that would be important to support future environmental risk assessments?
- What may be early stage risk management opportunities for synthetic biology researchers?

Session I: Introductory Presentations 9:15 – 10:30

Presentations:
- *S. elongatus* and sucrose production – *Daniel C. Ducat & Patrick Boyle, Harvard*
- Important gene flow concepts – *Allison Snow, Ohio State (via phone conference)*
- CEA framework and approach – *Genya Dana*
- Case study scenario of *S. elongatus* in a photobioreactor system – *Genya Dana*

A. Introduction to *S. elongatus* (P. Boyle)

S. elongatus is a good candidate for genetic engineering. It is naturally transformable (taking up DNA from the surrounding environment) and uses homologous recombination. Additionally *S. elongatus* colonies grow at good rate, doubling once every 24 hours. It also has relatively few nutrient requirements, largely needing CO_2 and sunlight along with small amounts of trace metals and nitrogen.

Because of *S. elongatus* requirements for sunlight and CO_2, it must interface with the environment (unlike *E. coli*) and therefore has a high likelihood of environmental release.

These requirements also create issues from an economic point of view. The economics of *S. elongatus* production get better as you scale up (more volume for less cost), but containment costs get worse (more surface area and greater opportunities for escape).

Questions posed:
Can this geometric/economic issue be addressed early in the genetic engineering phase?

How do you regulate an organism that uses sunshine and CO_2 and produces other things?

B. On Sucrose Production (D. Ducat)

Sugars are a biotech energy "currency." For cells to make desirable end products, such as biofuels, they need sugar inputs. Cyanobacteria could potentially be grown in nutrient poor environments, creating a source of sucrose that would not compete with food production or interfere with environmentally rich areas.

Sucrose is a compatible solute whose concentration in the cell can be greatly increased without damaging or poisoning the cell. Sucrose can even protect cells in some stressful environments. Sucrose is a solute that *S. elongatus* builds up in the cell in response to osmotic stress to maintain an appropriate osmotic balance with the external environment, for instance, when a freshwater cyanobacterium is put into salt water.

To make the sucrose produced by *S. elongatus* accessible, we need to get the product out of the cell efficiently and cheaply. By engineering the cell to produce the protein CscB and placing the cell in a salt solution, CscB will export sucrose out of the cell. However, such salt water use and sucrose extraction will decrease the growth of the microbial communities.

Rough early estimates suggest that sucrose production via *S. elongatus* could potentially be competitive with sugar cane sucrose production.

Production Method	% Sucrose	%Supportive Biomass
Sugarcane	20%	80%
S. elongatus in 100mM of salt	~30%	~70%
S. elongatus in 150mM of salt	~50%	~50%
S. elongatus in 200mM of salt	~70%	~30%

This sucrose production would be used for downstream biotech products, not for sugar. However, the levels of sucrose produced by this method would be very dilute for these biotech activities. Additionally, other organisms may be needed to remove salt which would reduce economic efficiency of this system.

In summary, this production is not perfect, but has potential. Sucrose production increases with culture density. Production by *S. elongatus* is continuous and doesn't need complex removal schemes. Additionally, it is non-competitive with food production.

Audience questions raised by this product:
How confident can you be that you are not turning on a silent operator in the genome that makes toxins or something else? *S. elongatus* is not known to produce toxins or have appropriate functional pathways for needed genes such as polyketide synthases, but if function is limited by regulatory regions, not synthetic genes, then could one turn on a silent regulator?

There is a risk with monocultures- how susceptible is this process to contamination (natural or anthropogenic) and how do you deal with it?

Do the cultures act as single cells or communities and how will this change how they survive/thrive in the environment?

If *S. elongatus* is naturally transformable, how stable are they genetically? Will they stay the same or change overtime? What is their rate of evolution?

Introduction to Gene flow concepts (Allison Snow)

To understand the movement and persistence of genetically modified organisms (GMOs) and/or their DNA, we need to look at horizontal gene transfer.
There are 3 mechanisms of horizontal gene transfer:

1. Transformation of free DNA: Free DNA from other, possibly dead, organisms is taken up by the host organism
2. Conjugation: The transfer of DNA from one organism to another through a plasmid or bridge-like connection
3. Transduction: The transfer of DNA from one organism to another by a virus

The persistence of this engineered DNA will largely be determined by the fitness cost, neutrality or benefit it provides.

To talk about risk assessment in general, and the *S. elongatus* case in particular, we will make the following general assumptions:

- Physical containment is not practical at a large scale production system. We should assume the GMO will enter local environment and disperse widely.
- The GMO can survive in the environment into which is released.
- The GMO has reduced fitness compared to its wild-type. We assume there is some fitness cost from introducing the engineered DNA.

Some initial questions raised by gene flow involving GMOs are:

- Under what conditions do you see this reduced fitness?
- How much fitness reduction is needed to cause rapid extinction in all cases?
- Could DNA from dead GMOs be taken up and utilized by other microbes?
- Which other microbes can take up free DNA/exchange DNA with living GMOs?

For example, consider GM crops. Some plants can survive and become volunteers. Canola can easily start new populations (they are volunteers) where corn and soy do not likely to volunteer. Survival of these GMO volunteers plus rapid evolution can result in several outcomes:

1. GMOs can die out due to fitness cost that reduces their competitiveness.
2. GMOs can volunteer and form new population. If there is sufficient genetic diversity, would this population be able to evolve and create a high-fitness feral strain? (endo-ferality)
3. GMOs can volunteer and interact/combine with wild populations and create hybrid strains with high fitness (exo-ferality)

To understand the likely outcomes from these volunteers, we need to answer some general questions about rapid evolution:

1. If transgenes from GMO interact with wild populations, in other words there is gene flow, how important is level of exposure (consider low vs high exposure)?
2. Horizontal gene flow is very common in nature. Can we assume that populations are perfectly adapted at all times?
3. Or are the engineered genes stronger, enhancing fitness and increasing population growth rates? (e.g. antibacterial resistance)
4. Could the spread of the GMO and engineered DNA lead to unwanted consequences?

In this workshop we will ask: Can we answer these questions? What other information or research would we need to answer these? Do we need more discussion among experts from different disciplines? Do we need new research to inform risk assessments? Will consensus and research be sufficient for a thorough understanding of risks?

Audience questions raised by considering gene flow:

What constitutes extinction in the environment- how do you evaluate what that really is?

> Ideally, extinction would mean the death of all the GMOs, but this is unlikely and the DNA may still persist despite the death of the organism.

Some genes used in GMOs, such as resistance genes, already exist in nature. Do we regulate naturally occurring genes? Do we need to remove them from GM organisms? What about other genes?

> From a regulatory perspective, we do care about using naturally occurring genes. We do not want to add to an already existing problem of resistance development. Even though resistance genes occur in nature, they do not normally occur in such high concentrations. Using these genes in GMOs increases the likelihood or other organisms taking up the genes. So we are not only concerned just with artificial genes, but artificial concentrations of natural genes as well.

It is hard to maintain these cultures in monoculture production systems, so at some point someone may introduce pesticide production or herbicide resistance genes. Such genes would make engineered microorganisms hardier. Will these engineered genes be taken up by other organisms in nature? How likely is it that this will occur?

There are other methods of containment for GMOs other than physical containment, such as suicide genes and nutritional constraints. How confident are we in other methods of containment?

> These methods would contain the organism, but not the DNA. No matter how you contain the GMO, the DNA will be released into the environment which could be of big concern.

CEA framework and approach (Genya Dana)
Comprehensive Environmental Assessment framework- used for nanomaterials- tries to think very holistically about how materials or organisms move through the environment and change throughout the whole life cycle of a product.

Overarching Questions: What kind of info do we need to know to evaluate the organisms' movement and impacts? Where are the information gaps? What are our research priorities?

To help answer these questions, this workshop considered a cyanobacteria scenario:
A plastic bag photobioreactor system grows *S. elongatus* in an industrial area, near marine waters, and is subjected to a major environmental event (e.g., hurricane) that results in large scale release of the organisms.

Initial thoughts/questions raised by participants:

- How fast do these organisms change with time? We may end up with endless numbers of similar but genetically different versions. Are we worried about rapid evolution?
- We don't have comprehensive knowledge of genetic sequences of different strains already existing in nature. Do we need to compare them to our modified versions? What baseline would we use to compare GMOs to unmodified versions when evaluating risk?
- There is a gap in knowledge of the genomes found in nature that would potentially interact with modified genomes. How would they interact? How likely is functional DNA transfer?
- What happens to the animals that eat these GMOs?

Session II: Lifecycle of *S. elongatus* 10:30 – 12:30

Key Questions:

- What may be potential escape routes from the production system?
- What may happen to the organism upon escape: e.g., does it survive?
- What might be important research questions re: escape routes and organism survivability?

Facilitator: Initially we assumed the product is in the Product Manufacture/storage stage of its life cycle when released- The first major question is does the product survive or not? What information do we need to answer this?

Thoughts from the participants:

- We need a lot of information about the surrounding environment and the organism's compatibility with native flora and at different salinity levels to determine if it would survive.
- We would need to know the "fitness cost" of the engineered gene and what fitness cost would encourage rapid fall off or "extinction".
    - Models of evolution can predict the pervasiveness of specimen-based on fitness cost, population size, and competition with wild populations.
- We would need to be able to track these organisms when they're released. Otherwise we would not be able to separate them from wild populations to learn valuable information about the impacts of release. We would also need the instrumentation and methods to track this data.
    - Metagenomic analyses in general are improving rapidly, improving our prospects for tracking populations of microbes.
    - Sandia National Labs, working on bioterrorism, is able to break up DNA into small chunks for analysis. They are trying to develop genetic forensics, which could be useful for synthetic biology applications.
    - We can engineer in an innocuous marker (a barcode or watermark) so we can track GMOs among natural populations, enabling us to collect the information necessary to create mathematical models, conduct future risk assessments and facilitate remediation.
- Some cells can persist in dormant or resting states. Our definition of viability needs to be expanded to include temporal refuge. How long can they survive in a dormant state? What does survival mean?
    - What kind of experiments or lab studies could predict survival through dormancy or resting? How long would they take?
    - FDA requires that the genetic construct is stable and stays within the organism, and that it can be tested.
    - On a moral plane, do we object to the idea that something we've modified or "created" may persist for a million years even if it is relatively innocuous?

Facilitator: Now assume that *S. elongatus* dies and the engineered DNA persists. What sort of information will we need to know to evaluate risks related to DNA persistence?

Thoughts from the participants:

- Will the wild population take up the DNA just because it's there?
    - Organisms in aquatic environments take up DNA fairly readily. DNA can persist in some environments for a long time.
- How do we handle disposal of the organism or its product?
- What biological adjustments may reduce the likelihood of these events?

- How do you evaluate the probability of these "rare events" (mass escape)?
- What sort of tests would regulators require?
- How do you allocate funding to research for developing these tools and test?
- Do we know the prevalence of the natural gene in nature?
    - We'd need to redo DNA viability studies with newer methods.
- Are there benchmark levels of natural genes? Can we identify genes that are so prevalent in nature that we would not be concerned about re-introducing them? Is there a threshold where we no longer consider the DNA a great threat?
    - It depends on the product of the gene- does it makes sucrose or does it produce something more sinister or toxic
    - Scale matters
- We choose organisms because they are easy to modify (i.e., for their genetic handle). Doesn't that also make us more concerned about their ability to transfer DNA to nature? By choosing these organisms, how much are we increasing the probability that engineered genes will be picked up by other organisms?
    - Can we do a study to understand the genetic transfer in different environments? Can we measure gene transfer?
- Are there characteristics that can be checked in the host organisms to ensure that they do not have toxic or blooming abilities?
    - There are countless products/by-products that these cells produce or can produce that are harmful. We need to have a profile of how the genome and products of the cell are changed by the addition of engineered genes (through functional genomics, transciptomic and proteomic analysis).
- It is fairly cheap to run a sequence and gauge what the cells are doing on daily basis. Have they changed? Are they the same cells you started with? Are they behaving as desired?

Session III: Environmental Compartments and Organisms 1:00 – 2:30

Key Questions:

- Which environmental compartments may be exposed?
- What external factors might influence escape and movement into the compartments?
- What external factors might influence exposure/uptake of the organism or DNA?
- What kinds of research questions would be important to answering the previous questions?

Facilitator: Much of the discussion has already centered on water. Are there further things to consider in this environmental compartment?

Thoughts from the participants:

- Based on the cyanobacteria example, we are worried about creating (purposefully or inadvertently) organisms that can cross the salinity barrier from

freshwater to brackish water, resulting in toxin producing populations in new environments; this sort of problem does exist in nature.

- There is a limited amount of resources available for product research and risk assessment. How do we allocate these resources when considering the different environmental compartments? Are there priorities? Are some environmental compartments more important or in need of funding than others?
    - From a regulatory perspective, the EPA deals with companies and products on a case by case basis. There is no prescribed format for information required for approval. No checklist. Should there be? What would this include?

Facilitator: What about other environmental compartments? What about air?

Thoughts from the participants:

- Based on what naturally occurs we can determine if these organisms are transferable by air. Because information on air transport already exists, a literature search may be sufficient. We may not necessarily require lots of money and years of research for air transport. Air transport can also be evaluated by mathematical models- significant work already done has already been done in this area.
- On the other hand, air plays a huge role in transport. This could make site-specificity irrelevant. An escape in one place can easily be a global escape. So how does this impact risk assessment?
- Borrowing from the human health perspective, we should ask what informs the risk assessment to the greatest advantage. What are we most afraid of and how do we gauge this? In the human health sphere there is a list of seven deadly sins that raise red flags- changing the host environment, etc. Can we generate a similar list for synthetic biology? What are the major changes or possible things that we might change that should raise red flags? What should we be focused on?
- The similarities of the GMO to the wild type are both good and bad. It's reassuring because the GMO is not that different from what already exists, but it's bad because this enables easy gene transfer. If we make organisms increasingly artificial, on what criteria should we evaluate? In 10 years, we may not be talking about this wimpy sucrose producing organism sitting a bag, but a more robust organism in an open pond making something other than sucrose. What then should we be considering? On what criteria would we evaluate these?
- Expanding our view from sucrose producing cyanobacteria, let's consider the possibility of modularity. Joule may be looking at creating a cyanobacteria chassis with modularity and plug-in components. We are talking about large scale production of organisms that can produce different high value items with

just small tweaks in DNA. If we're not overly worried about sucrose, we are definitely worried about other materials that could be produced using these microbes.

- Need consider the industrial parks of the future with multiple GMO's producing multiple products in close proximity. Are there special risks in this case?
- If we engineer algae to produce compounds that are different from what's in nature, aside from gene transfer, we also a have the very real possibility of accidental and large production of such compounds.
- Is there a template from the pharmaceutical companies that we can consider? They largely regulate the product, but also regulate the process.
  - The microorganisms are highly contained by the use of HEPA (high efficiency particle air) filters and they incinerate waste and spills. However, scale and containment issues would prevent us from adopting their template. Production outdoors on an industrial scale is very different from limited indoor production by pharmaceuticals.
- The debate has largely centered on 2 different issues. First, what is the impact of genetically engineered material being present in the environment? Second, how do scale, production methods and end products influence the risk assessment needed?

Session IV: Research Priorities	2:45 – 4:00

Research Topics:

I. Stability of DNA- Investigating rates of evolution and changes of functionality
II. Persistency- Compatibility and survivability of organisms, dormancy and resting stages
III. Fate and transport of genetic material.- Does the target gene remain functional in other hosts
IV. The physiological differences and functionality between the wild and novel organism
V. Probabilistic modeling of gene transfer

This portion of the workshop aimed to generalize the questions generated by the cyanobacteria scenario.

Facilitator: Knowing that risk assessment can vary greatly depending on the product, can we generate any foundational research needed that would apply to issues across the board, so that when the organisms become more robust, and we are doing large scale, outdoor production we can do a thorough environmental risk assessment? Is there a set of investigations we can do in order to know more about how to assess the risks? Is there anything we can generalize?

Thoughts from the participants:

- Ultimately we need to investigate rate of evolution for changes in functionality. In terms of robustness, you want the rate of evolution to be as low as possible, and the organisms to only divide when necessary. Cell reproduction can be slowed, but evolution is definitely sped up by scale. We can study genetic diversity vs time. However, we may not be measuring what we want. We can easily measure changes in

base pairs, but it is harder to measure changes in functions. We can also do studies in comparative genomics, but the original culture changes over time as well.
    - We need to use functional genomics to look at the whole physiology of the new cell. (General omics)
- Investigate survivability and compatibility (genetic and organismal) with native flora and fauna under different environmental factors.
    - To test organism and DNA compatibility, we can "create a big soup, dump the stuff in and see what happens." We can also introduce DNA fragments and measure what is picked up. However, we do not have standards for how long to stir the soup, what type of soup. We would also need to study what is "natural", to put the results in context.
    - Regarding survival, we need to be concerned with competitiveness. How many survival competition tests do we need? It should be in a whole community analysis. We need to consider everyone (the grazers), not just the competitors.
    - Dormancy and resting stages are important. How long can an organism rest?

- Are we concerned about DNA transfer, the added risk of putting the DNA in a new location, or is scale the major issue?
- From the regulatory perspective, we need to remember that people can come in with an application for a desert location and once it's approved for commercial process, they can use that product anywhere unless the EPA has put restrictions on the product. In order to put these restrictions in place, the EPA would need to know beforehand what to consider and what they should restrict.

- How can we track GMOs in the environment? Genome sequencing is becoming cheap enough that we can barcoding/watermarking novel organisms. This helps manage risk, but also provides tools to trace and monitor GMOs in case something happens.
- Investigate the activity of assimilated DNA; does the target gene become functional in other hosts? This can be answered with "soup" experiments and functional genomics tests. Is this level of research on gene transfer practical for all genes or just for ones where the trait is particularly harmful?
- Can we develop better modeling? Can the modelers guide the parameters and data needed to predict gene uptake? The entire gene may not be adapted. What happens if new organisms pick up only parts of the new gene? It is doubtful that we will ever have the capability to model gene uptake and transfer. However, models can be used to predict the physical spread of the organism.
    - Would it be useful to have a model that separates naturally occurring genes into ones that are prevalent enough that we can assume they have been thoroughly sampled throughout evolution and ones that are rare enough that we cannot be confident of how they will react in nature? Can we create a threshold of exoticism for genes to guide us?
- Understand secondary metabolites. How many should we look at? At what concentrations are we concerned? Some secondary metabolites, though possibly harmful can also be useful.

- We could run toxicity tests as you would for other substances.
- From a business model point of view, this knowledge is likely to be generated in the process of product development/optimization and therefore easily available.

These generalized research questions and areas where then further condensed into 5 broader categories. Participants were asked to vote on which they felt were the most important and briefly say why. Below is a summary of the research categories, the smaller initial questions they encompassed, the number of votes they earned and why.

Research category	Specific questions	Vote tally	Reasons given by participants
Rates of evolution and changes in functionality	• Investigate the rate of evolution for changes in functionality.	0	NA (no votes)
Survival and persistence of the organism	• Is the organism compatible with the environment and other populations? • Can the organism survive in a dormant or resting state? • What is the "fitness cost" of the engineered gene and how much of a fitness cost would encourage rapid fall off or "extinction" of the organism in the wild? • How many survival competition tests do we need? Studies should include a whole community analysis, under a variety of env conditions. Consider everyone (the grazers), not just the competitors.	5	• Encapsulates the genetic history of organism and be useful in understanding its evolution. • Companies are not expected to do a lot of work in this area; this information is difficult to come by, but important.
Fate and transport of functional genetic material	• Ability of DNA to persist after death? • What may acquire the gene? • Does the target gene remain functional in other hosts? • In what ways can the target gene alter existing genomes? • Introduce fragments of the introduced cassette and measure what is picked up by other microorganisms.	11*	• A non-scientist would be very interested in this. A risk assessment would certainly need to cover this to satisfy the public. • Fills in gaps, leads to useful information for both regulation and development of organisms. • Least understood of what we talked about today and therefore most interesting. • Most relevant from the policy perspective. • A risk we don't understand • * limiting fate of genetic material

Physiological differences and functionality between the wild and novel organism	• What is the natural risk of these wild organisms? • How do we compare the additional risk due to novel genes? • Investigate secondary metabolites. How many should we look at and at what concentrations? • What are cells doing on a daily basis? Have they changed? Are they the same cells you started with? Are they behaving as desired? • Generate a profile of how the genome and the products of the cell are changed by the addition of engineered genes.	9	• Captures a broad understanding of the organism before it is modified and allows you to compare your modified organism to something. • By focusing on this category, we would be addressing issues contained in research categories 1 and 2 • Need to know before we can say if the new organism will change ecosystems • This category has the least amount of available data • This represents the hazard part of the risk assessment which is important • This will be the trigger of regulation • This information is important for the 1st step for the risk assessment and will temper what questions you ask in other areas
Probabilistic modeling of gene transfer	▪ Can modelers guide the parameters and data needed to predict gene uptake? ▪ Would a model separating naturally occurring genes prevalent enough to assume that they have been thoroughly sampled throughout evolution from ones that are rare be useful? Can we create a threshold of exoticism for genes to guide us?	0	NA (no votes)

Other issues raised by participants

- There needs to be an increase in the accessibility of data. Both research communities and the public need better access to research to make better decisions.
- On the CEA framework:
  - There is no component on economics, which is the controller, so a feasibility check should be early on.
  - It was useful to remind researchers of areas they may have forgotten or neglected.
  - EPA's Office of Pollution Prevention and Toxics looks at all these components during MCAN applications, though not necessarily in this organized fashion.
  - Provides an orderly way of thinking about, but the EPA sometimes like to combine compartments because they are not isolated in reality.

Closing Remarks (Ken Oye, David Rejeski, Genya Dana)

What are the next steps to be addressed in the future?

We need to talk about steps to mitigate risks in engineering such organisms. Another exercise could be to do a risk assessment on something that is closer to commercialization. How do we mitigate these risks, how do we test this risks?

We need to discuss bioremediation. Organisms for this purpose cannot be engineered to be weak, and they are not contained. They must be more artificial in order to eat more synthetic materials. What are the risks from such organisms, and how do we deal with them?

We need to discuss more novel, orthogonally contained organisms. Re-engineered *E. coli* from another Boston lab is an obvious case study organism that we should look at.

In general, when do we intervene and how in the engineering process? How do we balance IP concerns and product development?

What would be the "7 Deadly Sins" for creating genetically modified organisms (and what should we call them)?

End of discussion##

Managing Uncertainty: How to Assess, Test and Demonstrate Safety for Applications of Synthetic Biology

Woodrow Wilson Center / MIT Program on Emerging Technologies / NSF SynBERC Joint Exercise

Woodrow Wilson International Center for Scholars
Ronald Reagan Building and International Trade Center
1300 Pennsylvania Ave., NW Washington, DC 20004-3027
June 29, 2012 8:30 AM to 5:00 PM

Rules for Circulation:
All Conference participants will receive this version of the notes (Version 1.0). References to conference proceedings should be based on these notes.

These notes were prepared by Joshua Fass and Ralph Turlington.

PLEASE NOTE

This report was prepared by The Woodrow Wilson International Center for Scholars (Wilson Center) as a general record of discussion during the workshop Managing Uncertainty: How to Assess, Test and Demonstrate Safety for Applications of Synthetic Biology. This report captures the main points and highlights of the meeting. It is not a complete record of all details discussed, nor does it embellish, interpret, or enlarge upon matters that were incomplete or unclear. Statements represent the individual views of meeting participants not a consensus. Except as specifically noted, no statements in this report represent analyses by or positions of any of the participants in the workshop.

Goals of this Workshop:

In 2011, the Wilson Center and the Massachusetts Institute of Technology (MIT) conducted joint workshops on potential environmental effects of a rE.coli chassis engineered to limit lateral gene transfer and cyanobacteria modified to produce sugars.

The June 2012 workshop used these two genetically modified organisms as focal points to continue a discussion of possible environmental issues involving the development of genetically modified organisms. The goals of the discussion were to flag sources of uncertainty about the potential environmental effects of these organisms, and to evaluate methods of testing the organisms' evolutionary fitness, genetic stability, and propensity for lateral gene transfer. The topics discussed in this workshop are intended to provide a foundation for eventual development of protocols for assessing and testing for environmental risks associated with genetically modified organisms.

TABLE OF CONTENTS

Session I: Introductory Presentations:
Presentation A: Peter Carr and George Church on rE.Coli
Introduction:

- George Church, as a technologist has helped drive forward DNA sequencing and synthesis speed; since 2004, acceleration of capacity for reading/writing DNA has far surpassed Moore's law. We are beginning to bridge gap between capacity to read/write DNA sequences and capacity to build meaningful genomes.
- Longstanding history on technology and policy, campaigning in innovative ways for safety and security.

Safety components for synthetic biology:

1. Physical containment – some BSL4 (biosafety level 4, the highest level) areas
2. Genetics – resistance to all viruses and other horizontal transfers. Change the genetic code radically enough for a "win-win" combination of making the change economically useful and attractive at the same time as constructing an effective barrier to information exchange
3. Nutritional containment: E.g. mutant strains of E. coli dependent on diaminopimelic acid, other possibility is to insert a new amino acid that increases fitness but isn't found in the wild.
4. Orthogonal antibiotics – make engineered organism susceptible to a new class of antibiotics
5. More training and scenario brainstorming for safety is needed.
6. Communication and transaction surveillance is needed

rE.Coli and genetic containment:

Very simply, a lab strain E. Coli function can be transferred by bacteriophage, which carry their own genes and hitchhiker genes. Free DNA uptake capacity varies wildly from organism to organism. To review basics of translation: ribosomes physically translate RNA into amino acids, the building blocks of the organism.

The genetic code is nearly universal throughout biology. Can we rewire the genetic code of a simple organism? Here we fundamentally change genome throughout to produce a functionally similar cell except with orthogonality by producing all correct components, just rewiring the mapping between tRNA and messenger RNA.

Genetic code is almost fixed, tolerant to mutation, shared by nearly all life. From an engineering standpoint, we want to make it easier to engineer, more controlled, and unique to one organism. Starting with low hanging fruit, we remove stop codons from specific areas. This gives us new room for plug and play scenarios. It also opens the possibility for rearranging to cancel crosstalk. This rearrangement can interfere with protein translation.

rE.Coli projects:

Although synthesis is technically possible, we developed and used an even faster set of tools that have general utility:

- MAGE, multiplex automated genome engineering, which makes many small changes at once (Wang et al., 2009)
- CAGE, conjugative assembly genome engineering, which entails rearranging very large pieces of DNA.

There are a couple evolutionary precedents, but this is the first time we've been able to rationally engineer in this direction. Although we're still in the end phase of this first codon swap project, we have also made great progress on a 13-codon swap, which has already been successful in 47 of the most sensitive genes.

<u>Audience questions</u>

Can you explain the decision tree that led you to choose this method, i.e. why is this "low-hanging fruit?"

This started in 2004. Our two groups were contributing in different ways.

Smaller number of changes this way, increase in feasibility, easier troubleshooting hierarchy, some genes are overlapping. Making a change to one requires changing another.

First decision: organism. E. Coli is the easiest to manage, highly industrially relevant, impacted by phage infections. It's infeasible simply to remove surface proteins to avoid phage. Then thought this orthogonality would increase safety. This is a task worthy of whole-genome synthesis, added tools including editing tools.

Is current codon use merely historical? Or will you see reversion back to ground state? Although rare, there are natural organisms with alternative genetic codes.

Note that when the genetic code originally evolved, there weren't even proteins. We've engineered this code so that it cannot revert; full reversion can be lethal to the cell in two ways—essential genes made dependent on the new code, and lethal genes that can be toxic to the cell but are only translated under reverted code. This also provides a measure of safety and commercial value. Want to lock it down so it doesn't evolve anymore and it's not just about controlling input of natural functions, but also to prevent export of the engineered genes.

What other efforts have been made to re-engineer chassis to be safer?

1. Jeff Lucas is working on a project at Johns Hopkins University, working with yeast chromosomes. It is a similar project, not specific to safety. They are building a chromosome that recombines more often, with the primary objective of increasing ease of engineering.
2. Poliovirus. Live vaccine, 3 mutations have to be reverted, occasionally escape. A couple groups are reengineering multiple codons to make weaker. This is a "death by many cuts" scenario.
3. Richard Sayer is working on photosynthetic antenna in algae. The antenna is engineered to absorb more light of different wavelengths, leading to toxic damage. Maybe not a safety mechanism, first it's an efficiency measure, actually more efficient with lower up-front absorption.

Could you give us a sense of the functions to which this chassis might be applied?

First, it can be used in combating phage infections in industrially relevant *E. coli*. We could go to industry, ask people who have phage problems with their bacteria e.g. propane diol, diacids, dairy bacteria, etc. Refine those organisms, and then move onto other related organisms.

Next, this chassis could enhance the robustness of other control mechanisms. Knock some genes out that would make them more likely to evolve.

Is growth rate inhibited?

We check for growth rate inhibitions in addition to anaerobic, storage, electroporation efficiency, etc. If we see effects due to hitchhiker mutations, we go back to clean them up. Before we publish, we'll make sure it's reproducible and well-characterized.

The point has been raised that they'll be in a community. The microbiome may be doing something; the organism will force the microbiome to evolve. A given microbiome may have the same enzymes and the same products, but balance of organisms doing it may change drastically.

We will come back to this in the next session.

This work sounds very impressive with multiple safety mechanisms addressing predictable things that could go wrong. What else could go wrong that's not as addressed?

We can begin to address unknown unknowns. For instance, in BSL3 (biosafety level 3 containment), we can create as natural environments as possible while tightly physically contained, then very rigorously sample DNA from this synthetic environment even if there's no apparent value, and see what kinds of natural environments we can bring in to the laboratory. This is easier to do with bacteria as opposed to, say, elk. We need to become proficient at this while it's still easy.

What about evolvability of the organism itself? You've mentioned hitchhiker mutations, etc.

This project is not trying to make a particular genome, and a lot of auxiliary mutations that could help are still up for grabs. Rather, we want to lock down certain features of the organism, not the whole organism, although we can reduce mutation rate and so on. The question becomes: what kind of evolvability would be bad, reverting the key safety mechanisms? We do lab evolution; create organisms that have different characteristics.

Is E. coli fully annotated now?

Largely, we could also consider the synthetic biology approach to annotation: delete everything we don't understand, and what's left is fully annotated.

Presentation B: Patrick Boyle and Dan Ducat: Sugar-producing cyanobacteria: an alternative biotech feedstock source

We're interested in this because biotechnology relies on refined sugars as a kind of energy currency. Any biotech product comes from feeding a purified sugar to heterotrophic bacteria that turn it into a product—correspondingly need huge amounts of sugar for mass production, and most of that sugar in turn comes from agricultural sources grown on arable land. This results in an inherent conflict between food and fuel and raises many related concerns, driving interest in alternative feedstocks.

Cyanobacteria (specifically working with *Synechococcus elongatus*): a photosynthetic prokaryote 1-2 microns long. The bacteria also have granules for storage. Carboxysomes house machinery for fixing carbon dioxide into products.

- Driven by photosynthetic reactions in thylakoid membranes

- Some cyanobacteria are in arctic lakes, some in other freshwater areas, etc.—very broadly varied and distributed: this is not an exotic organism. Additionally, it has no known excreted toxins.

Sucrose production design: particularly interested in the simple disaccharides of fructose and glucose. The disaccharides do not seem to have deleterious effects on the host even at relatively high concentrations. Used as a compatible solute with protective effects under a number of stressful environments that balances osmotic pressure to keep the cell at homeostasis instead of desiccating.

Want to export these sugars so we can access them as a possible fuel source.

To help with this we used cscB, which is a sucrose/proton symporter in E. coli. It has a variety of homologs, and uses a proton gradient. The organisms can live in acidic environments relative to their cytoplasm.

If we took this and expressed it in *Synechococcus elongatus*, it would work the other way. It would basify their media (a pH of 10-11) yielding a higher proton concentration inside than out.

Insertion of cscB into genome:
- cscB under the control of LacI, which is a commonly used regulatable promoter. IPTG inhibits LacI which inhibits cscB. This allows switching on and off on demand by adding IPTG. In deployment, we will use a constitutive promoter.
- This construct is inserted directly into the genome using homologous recombination. There are a number of different neutral sites that used to be genes and are now non-functional.
- When we express this construct in salty media, we see the continuous export of sucrose. This was monitored for several months and continues as long as light is shining on the culture.

There is an inherent tradeoff between sucrose production and cellular biomass; reducing cell growth for more products. Normally the sugars are redirected into biomass, and siphoning off fixed carbon into sucrose means the carbon is unusable for cellular metabolism.

We have been able to tune the amount of export sucrose vs. cell biomass using induction regulated by salt concentration. So far, we have up to 85% fixed carbon redirection at high salt concentrations.

To put cyanobacterial sucrose production in context, we will compare it to existing sugarcane sucrose production. Sugarcane is the most efficient source currently, and sugarcane achieves only about 15% sucrose by mass, whereas cyanobacteria can reach up to 30-80% sucrose by mass. Assuming similar yields per hectare, we can get 5-6 times more yield, rendering it competitive with sugarcane production per land area. Range of cyanobacteria productivity depends on deployment details: for example, open raceway ponds achieve 25-30%, whereas contained facilities reach up to 100 times that, but at dramatically higher cost.

Potential benefits/constraints:

- Could feasibly be competitive with sugarcane (reducing need for more arable land)

- Inputs required for cyanobacterial growth are minimal (nitrogen, phosphate, sulfur, trace metals)
- Even minimally contained designs (e.g. lined ponds) would reduce total use of fertilizers and other supplements
- Sucrose production is continuous and on-demand. There are no longer constraints from seasonal growth and harvest cycles.
- Additional strain selection/modification could increase yields substantially. This first step, for instance, yielded 30% increases in the past year.

Two main hurdles remain:

- The purification of dilute sucrose could be economically difficult.
- Reactor design and contamination remains a problem. Sugary solution aggravates this even further, providing a highly attractive growth medium for opportunistic contaminants.

We're looking to see if we can create synthetic co-cultures of cyanobacteria with other heterotrophs to directly convert that sugar source into downstream products, which may make the overall culture more robust and resistant to incoming strains.

There has been renewed interest in micro-algal production of fuels/compounds; many companies are interested.

We were interested in presenting this project because it represents a minimal test bed in the cultivation of algae or cyanobacteria:

- genetic modification is minimal
- product is non-toxic
- sucrose production is effective and robust

Perceptions, concerns, and suggestions related to this strain may therefore apply generally to most photosynthetic production strains.

Physical containment is still a fundamental challenge. Because photosynthesis depends on actual capture of sunlight, production must be expansive on a horizontal plane, and ability to contain this system completely is limited by the economies of scale of what you're trying to produce. How fast does the per unit cost decrease as the size of the facility increases?

Plastic bag set-up and scale-up is a huge driver of whether this will actually work.

Discussions of containment need to recognize that total physical containment is probably not an option. Nutritional containment might be very difficult to reconcile with economic feasibility.

To address considerations from last meeting:

- Transmissibility - transfer of novel genes/pathways to other organisms
    - No easily transmissible vectors (integration)
    - Use multiple integration sites spaced throughout the genome for other modifications (reduces likelihood that a fully functional circuit can be transferred)
- Gene flow - poor understanding of how readily genetic material will transfer
    - Metagenomic study of water/soil microbes neighboring a test plant facility would be feasible and informative. Possible funding?
    - Risk consideration weighted by genes providing selective advantage?
- Effect of gene insertion on organism's physiology
    - We have done some characterization of this in past year – see publication (D Ducat, J Way and P Silver, "Engineering Cyanobacteria to Generate High Value Products," *Trends In Biotechnology*, February 2011.)
    - Will continue work to characterize the other natural products secreted by both wild-type and modified strains to see if there are unexpected products
- Consideration of natural exposure to homologous genes
    - Natural exposure of organisms to inserted genes—number of homologs with similar transporters.

<u>Audience questions</u>

Are there efficient ways to survey ability to survive and reproduce?
Moderator: Now open to points of clarification, also want to talk about parallel or more advanced projects in the field now.

Is your use of multiple integration sites an attempt to reduce horizontal gene transfer?

Spacing out individual elements of a circuit into multiple parts of genome will reduce probability of entire functional circuit being transferred.

Is there some conflict between sucrose/biomass?

We can tune when they produce sucrose. Can grow to whatever density you want, and then when you hit them with salt it produces sucrose on demand. Not dying off, just reducing the amount they're being produced.

Is this salt level similar to seawater?

0-200 mM salt is a brackish system. Chesapeake Bay water would include concentrations sufficient to activate this system.

Is it highly productive now?

The genes responsible for sucrose breakdown have already been knocked out. And there is no change at low salinity.

Session II
A. Insurance Issues

What would be most important to an insurer?

A model risk landscape is vital. We need to look for values that can be affected by this new technology:

- Leading values: humans, animals, plants, environment, and so on.
- Financial values: property loss. Quantify to dollars at stake at the end of the day.

Can you imagine a scenario when you might exclude other scenarios?

- All products are all risk products. We must estimate frequency of problems. Is this product economically viable?
- We must have family of peer risks, comparisons to similar products, and basic principles of insurability. Who else in insurance business is on this?

There is also variation across liability frames. Legal action may not be covered. There would be talk about legal practices and what kind of losses could occur.

Insurers are worried about a lot of things. We have to assess things in which the problem defines the outcome. Check what companies, scientists, etc. are doing in terms of anticipating failure and its different modes. This is more promising than just trying to come up with a lot of scenarios that are more or less mindful. Insurers try to come up with benchmarking exercises to compare safety and design performance. They try hard to prefer the one that's doing the best job. The idea of relative risk management is invoked. Similar to what happened with nanotech companies, some who developed nanoparticles less responsibly, and others who were a lot more precautionary in their work.

Overall, premium alone cannot compensate for bad risk. Insurance is about atomizing risk into small bits and pieces to make it bearable.

How do you set premiums to reflect your understandings? What conditions on coverage do you impose? How do you quantify these risks etc.?

The companies take examples from life sciences and pharmaceuticals. They look at the return period from expected loss. The loss is a dollar figure heavily driven by the legal system. They are partially compensation for physical, punitive, and other damages. This reflects an expected feeling of risk from the technology. This system creates a risk landscape of everything that could go on and is covered by insurance.

I understand underwriters and actuaries could do this, map coverage to behavior on the basis of empirics, but how do you do it in synthetic biology?

It's just gambling. Insurance companies make bets using game theory. They win if they're slightly better than average.

Leave aside agriculture and bioremediation for right now. Let's think bioplastics, biofuels, etc. In Europe the company has to buy insurance. "Hey, I'm just an innocent guy developing my rE. Coli" what questions would you pose to these technologists?

Very few questions would be asked; overall we just want to prevent the destruction of capital. Instead, the company would rather pose the question of whether your activity poses a reputational risk. Here we cannot afford to delve too much into the details, just need to condense it into a very limited number of risk buckets.

- What are your anticipated failures?
- Worst-case scenarios?
- Plans/intentions to prevent these scenarios from happening?

You want to quantify a whole universe of values and effects. When assessing a drug, you look at the number of people taking the drug, the adverse drug outcomes, the probability of suit, the success in a suit, and the average of indemnification, among other things. An insurer would apply a similar method to look at environmental damage in terms of ecosystem services in dollar loss figure.

Bottom Line:

- Insurance will not care about the transfer of genes; will care about effects of these transfers.
- We don't yet know enough about the natural ecology to make good predictions.

B. Environmental Issues

We've mentioned that evaluations would need to be done. The DoE called for evaluation of algal test beds. Need to think about where you're doing things and what you're doing?

Need to consider potential of bacterial industrial parks and industrial symbiosis. One organism's product becomes another organism's input. How do you regulate and assess risks of an integrated system? Such integration might dramatically increase efficiency.

Have you (Dan and Patrick) talked to anyone if your sucrose output would be someone else's input? We should also hold onto the concept of thinking in terms of bacteria in consortia, assemblages, communities, etc. rather than in monoculture isolation.

This is actually one of the central appeals of cyanobacterial systems.

How about parking algae near power plants to fix CO_2?

This is a type of industrial park; one output is another input.

We've heard one application of one minimally modified algae for sugar production, are there others under consideration?

- Synechococcus has been modified to produce fatty acids, isobutanol, and ethanol.
- Methods for engineering will also have implications on risks.
- These areas might be proprietary in their nature; someone may have already patented that strain of algae.

Could you lay out some of the design principles being used in your research?

The company Joule Unlimited is making hydrocarbons and various alcohols. These are immediately compatible with current engines. Hydrocarbons are not easily consumed by heterotrophs, reducing the contamination issue.

It may be worth moving the discussion in the direction of systems that might have environmental risks, not ones that are just benign.

(To Dan and Patrick) Getting around the Rubisco speed bump through hydroxyproponate bacteria: is this a way to increase fitness?

- Rubisco is ancient, and kind of slow, it also has problems with Oxygen. Can we get new carbon fixation pathways as an alternative to Rubisco?
- Possibly a problem that these organisms will become a weed, faster and better than existing cyanobacteria.
- Thinking only to do so in auxotrophic chassis so it's not spreadable from the lab.

Confusion over the definition of safety, could we discuss what safety is?

- Safety is preventing harmful outcomes. There are concerns about unknown outcomes, not just harmful. Preventing gene flow is a safety component even though it's not a known risk. Safety to the environment as an engineered organism could take over. There are health risks to humans or agricultural species. Atmospheric environmental risks (global warming, etc.).
- Extent to which many are against the word "safe" is quite extreme. Raises the societal expectation that is indefinable and unachievable. Really try to keep in risk vocabulary and away from safety vocabulary. In practice, ability to condition expectations isn't really good to take account of uncertainty.
- Lessons must be taken from the FDA, because initially the big deal is to get approved, and then to market. Only recently realized that ongoing monitoring is important. We are looking at something that is fraught with uncertainty. Can't just say "safe, it's fine" anymore.
- The complexity of these interactions is unmanageable even *before* we start introducing the synthetic aspects.

Three main pathways of interaction discussed so far (fitness, stability, and gene transfer). Are there other pathways?

What are the concerns of the environmental microbiologists at this point?

- Overall, we want a solid definition of fitness.
    - Ability to leave offspring relative to other phenotypes.
    - Are we looking at competition?
    - Reproduction?
    - Both?
    - What specifics would you then look at?
    - Would you look at cell size, cell metabolism?
    - What are we actually talking about?
- The word "fitness" is a very big word, and *very context dependent.* When someone sees the cyanobacteria example a tradeoff between sucrose and biomass, we need to remember that there are a lot of co-associations, symbiotic interactions, etc. that may benefit other co-evolving organisms, possibly creating a net-positive fitness from the associations. We need to consider the positive, not just the negative effects that will happen in nature.
- A simple way of defining fitness is the number of offspring of a particular "genotype" how many copies of the genes are being promulgated. This is what allows a new genotype to spread.
- There may be consortium changes. Not only the organism changes in fitness, the microbial community changes as well.

C. FITNESS
How do we test for fitness?

- We can use a physically contained bioreactor with a diverse set of natural environments (sewage, agricultural runoff, stream water, etc.). We may want to make a large parallel set of environmental simulations.
- Create a "tournament" for a single organism against another. This is a more complex evaluation of a single organism.
    - The tournament idea is a great first start for evaluating a single strain or competitions between them. We need to consider that there are many of environmental factors and it's impossible to test all of them; we will never achieve what we'd see in the environment.
- Another definition of fitness is on a gene basis; this connects to the barcodes idea. Dead organisms may still be hazardous because their DNA is still there. We can use next-generation sequencing and PCR which are powerful techniques for detection.

Take extreme end of one bug vs. another strain. How do you pick your cocktail or mix of organisms?

In the 1980s and 1990s the EPA sponsored a lot of basic research in microcosms. You start with what's known and then move into particular organisms you're investigating. This means that you're not starting from scratch though. You can always test for a fitness cost, but if you don't see one or actually see a fitness increase, you're left with a lot of questions, for instance:

- What happens if you add virus resistance to an organism that is normally regulated by viruses, especially in the oceans?

Tests could be done with actual ocean water. There have been many works published on ocean viruses. We can make bigger and bigger experiments. These are not far-fetched ideas. Could be

used as a safety mechanism for cyanobacteria, have to make sure you go the whole distance to test if it's safe and won't dominate its ecological niche.

- In most microalgae, at a production scale the worst-case scenario is that the culture gets contaminated. Add one outside organism in and outcompetes your engineered organism. We want to engineer the other direction. We can envision taking genes from dinoflagellates so that they can outcompete a contaminant.
- This type of production increases the risks when the culture is released, which is also inevitable. Currently, these experiments are being run all the time in industrial scenarios.

Most of the work going on is people having a product that they think solves some human need. We have to put into context of what you're trying to accomplish and then say what are your unintended consequences near the river in whatever state?

Do the people who think about testing all these qualities think the same way as we think about the testing of chemicals?

We are trying to make things detectable, have a sensitive assay to make the deleterious effect manifest. There are drawbacks:

- There is heated rhetoric about the nature of trying to be conservative and precautionary.
- The controlled nature of the tests (pristine, etc.) raises its own questions.
    - Many diseases require environmental messiness (example of lymphomas which cannot be seen in the lab.)

The risk framework used for chemicals and materials in the environment could be used. Things will change and transform over time. The real difference is in exposure. Exposure might change over bioconcentration/biomagnification, but with fitness, what happens if exposure changes completely due to fitness interactions in the environment? Evolution of chemical testing tracked the changes in chemical industry with regulations. As time went on, the instruments became finer and finer (ppm to ppb), they co-evolved and changed the regulatory enterprises. Sequencing technology in particular will be extremely influential as well.

Can we consider evolution? Have these tests already been run. If there is a huge fitness advantage, what are the chances this has already been sampled evolutionarily?

We seem to be back in the 1990s with rDNA (recombinant DNA) technology. As far as fitness, we've already been testing fitness, survival, and competitiveness with parental organisms. With synthetic organisms we may lose the parent organism to test against. The EPA has been testing with microcosms. We can move to mesocosms and beyond to field trials, if the containment works. May see things arise, after years of monitoring, before ever moving onto actual commercialization.

Fitness is a highly context specific trait. Cyanobacteria creates a novel biological niche that's exploited by a contaminant. What is the niche within which our organism is the fittest, or set of niches where it is very fit? This offers a second way of mapping how this organism will behave in the natural environment.

What was the average cost to do these tests?

Industry has never had anything that's gone through the testing of mesocosms.

- We were going to release a gram-negative bacterium with a toxin, but due to the inability to make progress releasing a recombinant organism their final product was a dead recombinant organism. Transmit this over to what has been released so far in transgenic crops; we need to think about what makes these microbes different within a risk assessment frame.

That was back then when it was new, but now people know what to expect, could be easier.

- But the cost for the whole thing? Couple of million per year. The Environmental Use Permit and related paperwork, but not just for paperwork but for field trials with people of certain credentials, a lot of analytical stuff.
    - Repeat: we decided to use a dead rDNA organism. We grew it to high density. But how do you prove you killed everything in a hundred liter tank (without denaturing its DNA)? How do you set up the statistics, tests, and evidence ongoing that you killed something of this magnitude?
- With tests like this, we won't have to repeat the early stuff; we can start from a different stage. Also working with pesticides, regulations are more stringent for hazard assessment than usual.

Not an ecological concern, but there is an additional economic concern that hasn't been as present until recently in the Gurtz case in the Supreme Court. The issues arises if your product contaminates someone else's. If you work with cyanobacteria that can infect something used as food by other folks, or contaminates a natural or organic product, a lawsuit could happen. But the contamination of rice strains by pharmacological rice resulted in the company making the farmers somewhat whole. When the regulations fail, it moves to trial lawyers, which is a real cost.

- A lot of this testing is organism-specific. If you have something not designed to be pathogenic or toxic, the type of data you'd want to see is different than with a potentially toxic organism.
- Seasonal variations are also an issue; this is at least a one-year commitment.

How can we resolve the difference between natural experiments and lab experiments with different species? Might this change how we do containment testing vs. natural experiments? E.g. testing fitness or stability of genetic function over time? How might that be incorporated?

- Metabolic composition of the consortia, not just metagenomics. Evaluate if your metabolic concentrations are indicative of changes in consortia. RNA-sequencing of all RNA in environment, shifting levels indicates bigger shift in metabolic activity.

When setting up facilities how much are you benchmarking before you start building, or start introducing other stuff?

You have to think about the other stuff as well.

D. GENETIC STABILITY, GENE FLOW, AND ASSOCIATED RISKS

First, any points of clarification, second thoughts, etc. related to fitness and survival?

What is a consortium?

- A natural assemblage of microbes.
- Rule rather than assemblage. Current focus on monocultures is probably not efficient since nature involves consortia.

We need to consider the possibility of a hit-and-run-event, i.e. a microorganism changing your consortia and then leaving. For example, *Helicobacter pylori* can invade the stomach, cause cancer, and then leave. The challenge is that we need to have assays not just for DNA sequences we've made but for their functional consequences.

The point was brought up in the discussion when talking about fitness: if the organism persists in the environment is that in and of itself something of concern? Or, should we only deal in consequential endpoints (e.g. harmful algal blooms, extinction of species)?

Start to think about consequences, if this might help define some of the tests to inform risk assessment.

You can summarize risk analysis into these central questions:

- What could go wrong?
- How bad will it be?
- How likely is it?

Risk management can likewise be boiled down into:

- What can you do about it?
- What are the tradeoffs?

To take the example of nuclear risks, the process involves assembling a group of people who are very knowledgeable about a process to systematically address these questions. It takes a group of a dozen people familiar with many different aspects of the total system, systematically working through combinations for a week or two. If this hasn't already been done for synthetic biology, it might be useful way to approach these issues, determine which kinds of tests could or should be done.

- You've just described our normal process for evaluation of microorganisms subject to our rules. (Company rules)

In discussions of what happens if the transfer gene is integrated into another organism, we seem to assume that they must be non-functional or weaker. How do we know that won't change?

If you engineer intentionally in the direction of safety, not just changing nonstandard amino acids, but changing the organism to be dependent on a substance which cannot be found in the environment, you've added yet another barrier to help reduce likelihood of successful escape.

How do we know that making so many changes doesn't bring us into a completely new space?

The idea of creating more and more independent probabilities makes sense, but it often actually turns out they're correlated. "If you change the codon, won't it be nonfunctional?" Doesn't it move from something I know will fold to something that won't. How do we know that it doesn't change to be something totally different and unpredicted?

- Getting away from language of safety to language of risk is also applicable here. You have the question of how many. Probability that x substitutions will change protein folding landscape. Not an absolute all-or-nothing principle.

One solution is to take many groups and bring them in from the beginning.

We also skipped something: STABILITY. We've been discussing gene flow.

We don't have anyone from the FDA here, and actually the FSA and pharmaceutical industry have joined together to exclude some forms of synthetic biology from RAC review. The FDA in their genetically engineered animal review does require addressing stability.

- May be useful to consult that study.

In reality, economics will drive stability. That's the biggest pressure to use genome insertion vs. plasmids to convey genetic information—overall making sure your platform doesn't crash out of your strain. We are more interested in preventing phage from hurting your organism than other way around.

So are you suggesting that economic self-interest means we have less to worry about?

We still have to worry about consequences of stability. It's nice if your organism died if your gene wasn't stable, but there is the danger of large financial losses if the organism dies due to contamination.

In defining the function of the given circuit how much effort is given to determining how important stability is to the function of the circuit? It's unclear what the concern is from a risk standpoint if that circuit breaks down in metabolic engineering.

In the primer, the authors strongly suggested that relying on kill switches is risky because stability disables the protection mechanism, which certainly harms fitness. Thoughts?

We should define it as conditional stability because we want things to be different in the lab vs. outside. In industry it is necessary to describe how to engineer things to survive in the raceway, you want it wimpy outside and strong inside. But add to it, the stability and likelihood that mutations or change would affect weakness vs. survivability.

- In all these scenarios, the unnatural niche is required for the engineered system. For engineering it for a specific environmental niche, it's easy. I worry more about stability of gene flow into my engineered organisms.

Unfortunately, the two examples we have would favor stability. In standard biotech, you'll occasionally want high copy number plasmids, which can be maintained in their artificial systems. Then you worry how stable it is after it gets out of the fermenter. But that's not a question they need to ask in the two examples discussed today.

- It sounds like, with regards to stability etc. from a regulatory perspective, there's no specific set of questions. The onus will be on you to produce a quantitative risk assessment with data sources, reasoning calculations, etc. Sounds like you should plan to enter those risk assessments or work with consultants or regulators if they are not being

done in a standard efficient manner, so you don't have to do the whole model from scratch every time.

In industry, when filling out a regulation submission, there has never been an extension to section five, stability, where I wrote down that I needed it to be stable to make money. Is my 17kB what I put in the first place? Does it stay the same over many generations? It raises the question of: what is the nature of my strain repository? How do I know this entity is the right one when I pull it out of a repository. This all comes down to characterizations. Sometimes comes down that at pre-production scale you get process stability. Some of these questions while interesting are how the process exists now. There may be sequential changes in the way we submit data, this might change submission requirements or practices.

Thinking about this from an invasive species standpoint gets to evolutionary stability. What are the chances of this surviving continuously? I care whether this can survive in the Chesapeake Bay. Is there a mechanism to test the chassis from an environmental standpoint for survivability? How would you do that? Does this get back to the threshold issue?

Start small and controlled, and then move up in scale. Currently there are many tools for assessing evolvability. There are other tools for assessing adaptation of specific function. For the simple things, like mutations rates or function knockouts, there are easy ways to measure this.

We've been talking about the actual organism surviving outside of your controlled environment when you move this out, however it gets out (intentional bioremediation etc. or unintentional). What are the threshold numbers we want?

Competition experiments would give us some useful data. Not only: "does it survive?" but "does it compete against other natural organisms?" Again, these tests could be scaled up.

How do you balance that with stability to work?

We need a point of reference: cyanobacteria in its original niche can be compared to sucrose production in other organism's niches (both the host organism and the strains you added). Is there a high probability of transferring to other organisms? If there were a waterway would it work? The point of reference has to be based on host organism and traits, and some reasonable approximations.

- One interesting observation involved sequencing a grave organism and found it was similar to present-day plague, and the sequence is conserved over time. There are evolutionary mechanisms that keep these organisms from changing too rapidly.

E. DEFINITION OF SYNTHETIC BIOLOGY

Synthetic biology as a term is disorienting. It seems to be a complete blurring of terms. Why is Dr. Church's research clearly synthetic biology?

The background in which this is happening is also important to the definition. All this new metabolic engineering might be happening on a new chassis that might reduce the risk.

We are engineering a chassis to improve functional properties, modularity, and also decrease the likelihood of gene transfer.

The extent to which one can evaluate gene transfer, etc. without reference to pathways is central to synthetic biology questions. Cyanobacteria is a baseline example, the major change is quite simple, could add many other layers of synthetic biological control. The difference between synthetic biology and metabolic engineering is pretty small and may be meaningless.

In this case, we want to deploy an organism. In this case, there is modular design. However, using more exotic techniques down the road might look very different. A big advance is that we can synthesize whatever we want, no longer limited by existing genes. Additional new capacities include codon optimization, no scar sites, etc. Another substantial distinction is between rationally writing genetic code and introducing random mutations to create desired phenotypes.

What methods of design and construction (scaffolding, etc.), have environmental implications? Is gene transfer/stability more likely using some methods? What is unique?
Overall the two important implications of synthesis:

- Codon optimization
- Expands how far across the tree of life you can move genes.

Interesting distinction, but since the 1990s everyone's done codon optimization, polyadenylation site usage.

How do new capacities in terms of tools allow monitoring, modularity, measurements allow a different kind of grasp on fitness or risk?

What is risk, what do you mean by risk in these contexts? What is the problem at the end of the day when the gene is transferred? The fact that it happens is not intrinsically a risk, just the consequences. Before we start a new conservation, what aspects of these conversations are unique to synthetic biology as opposed to, say, metabolic engineering etc.?

There's no doubt there's overlap between these disciplines. These are "big tent" disciplines. The semantics may have more to do with funding. Learning to apply the tools of other engineering disciplines to the design/building/testing of biological systems characterizes synthetic biology. The sorts of things we're trying to produce in terms of complexity and applications represent a huge leap from genetic engineering.

- The design-build-test cycle is moving a lot faster, which changes the risk profile. This acceleration is a stated goal, a fundamental part of the engineering, a hope and aspiration for these techniques. There is more control, but also new risks and new benefits.
- Likelihood of larger-scale use. Anything that affects scale-of-use affects effect.

So isn't synthetic biology quantitatively different, not qualitatively novel?

We can now do recoding. This is not something we could have just done inefficiently using old techniques. There are many other applications in synthetic biology that are clearly not just metabolic engineering. Biocomputing, for instance, can use DNA strands to perform complex functions. Synthetic biology completely diverges from metabolic engineering at this point.

Synthetic biology may be enabling changes that couldn't occur evolutionarily whereas metabolic engineering is accelerating what might have happened before or could happen again. Metabolic engineering is an outcome of what you want from engineering organisms whereas synthetic biology is an assortment of attempts to engineer organisms better and more rapidly. There's going to be more of the kind of recoding that George and Peter are doing. As this becomes more common as an aid to design, this will add another layer of "synthetic" to biology:

- New amino acids
- New nucleotides
- New proteins with novel functions due to design or directed evolution

Suggested analogy for biology people: comparing genetic engineering to synthetic biology is a lot like comparing prokaryotes to eukaryotes. No one would agree they're the same thing even though they use the same amino acids, etc.

F. TESTING AND TESTING TECHONOLOGY

Let's move from potential properties to empirical issues. In rE. Coli, designing to have certain properties, how have you tested for these properties? Obviously this is a fascinating design exercise.

A lot of our transcription profiling is trying to be open-ended in a way that isn't hypothesis-driven. We are testing a variety of phages. We want to create strains that are nutritionally dependent before making them multi-virus resistant. We also have experiments on synthetic ecosystems, tested in increasingly realistic ecosystems. Going to combine *rE. Coli* with this to see what happens to the DNA in these complex ecosystems. Does it proliferate, incorporate into the genome, disappear, etc.

Any tests you wish you could do but can't?

Short answer is "no." Long answer is "maybe" but scale is the big difference. Chances of escape mutation in 10^{12} vs. 10^{9} cells is very different, we want to know how different.

Is there work on standardized synthetic environments? It would be useful to have contained environments not just for rE. Coli but more broadly as well.

Back in the 1980s people set up their various microcosms to look at soil, etc. They had mechanisms to look at populations but didn't have ability to sample the metagenome. Now we can do that without having to look for the specific products. This all started out with soil in a flask. It needed to be more realistic than that. Intact soil cores with roots were brought in. Came up with these and were field tested with a number of different bacteria.

- Freshwater
- Marine water
- Brackish water, etc.

There have been studies that have literally dug up giant plots and brought them into a lab. This obviously cost a lot. We can now measure these fully functional ecosystems (minus land animals) brought into a controlled environment. Duluth lab in Minnesota concocted artificial

systems based on observations. Selected plants, and other organisms and had replicates that were similar.

We can do this at different scales. There are currently measurement mechanisms at the molecular level that we didn't have before. There's quite a bit of literature on the relevance of observations from micro and mesocosms to environmental impacts. Move up to mesocosms when ready to sacrifice the ability to replicate, and precisely control, but can get more realistic environments.

In what ways have measurement technologies changed?

This is all about the tools. What do you want to measure?

1. Is it a novel gene?
2. What is the function?
3. In what organism?
4. Any impacts on other organism's presence?
5. Any impact on organism function?
6. Any impact on function/structure/composition of community?
7. Any impact on non-human organisms we care about?
8. Any impact on humans?

When you ask those 8 questions, you need to know something before you get started. You need to know something basic about what's happening with your organism in the lab and in nature. Going to be asking about different components depending on what you know about that organism. You may be satisfied that it will never be a pathogen, narrowing your risk assessment. It's not going to be completely open-ended. You're going to be looking for specific things you'll hopefully have the ability to measure.

If I were a synthetic biologist and didn't know what I was going to plug into my chassis, how would I think about the important ecological questions? I don't think I'd want to do all my work on my organism and then dig soil up later and try to figure stuff out. Instead we should want to engineer from the ground up to prevent gene flow. What is the mechanism of gene flow? Is this something we can plot out in a series of events? This is a simple two or three step process before getting started.

So these are classical trade-offs. But how do we improve the terms of these trade-offs?

- Two approaches really in play:
  - Hypothesis-driven
  - Observation-driven
- We could start with a specific concern and then do as much to prevent that one event as possible.
- The macrocosm approach is not hypothesis-driven, start by creating an environment as realistically as possible. Then observe it extremely carefully at first since you don't know what's going on.
- Do all of the above. What if we're wrong about our hypotheses about what could go wrong?

- At some point, you can't do everything with everything. You have to have a way to decide what things go through what types of testing. Depends on what you know about the organisms, gets into uncertainty, where you know more or less. You need to be cognizant of that element.

How do you evaluate the extent to which the chassis provides protection independent of pathway, or with reference of pathway?

One of the purposes of rE. Coli is to increase protection in this regard. rE. Coli has a chassis that's been modified substantially, and it's going to be much more deeply modified still.

If the objective is protection, then protection from what?

An easy answer to the "what?" question is horizontal spread of DNA. Test with human, chicken, pig, bacteria, plant DNA, etc. If someone designs this for a crime, what would we do?

On the topic of modularity of risk assessment: not having to do a do a completely new risk assessment for every single new organism. We should have a database, pool of knowledge about all of these data. Get probability data about all of this, the various nodes on your fault tree. Someone might be able to compile all of this. An example may be nuclear process systems. Using things like rates of mutations under various conditions, etc. might be able to give us interesting data.

We want to demonstrate it won't survive and there'll be no gene flow. It is difficult to supply negative evidence. What kind of testing program do you actually want to demonstrate desirable properties of the chassis?

We can't frame it as all or nothing, but $1{:}10^{15}$ or whatever it later is improved to. We are trying to define a situation as a base-point, and then we are deciding whether it is a threat or not. A lot of this mesocosm research belies a belief that we understand real biology, whether we can predict ecology given synthetic predictions. No one agrees on these predictions in use. We must have careful understanding of limits of these conclusions. We can imagine situations that put up extremely obvious stop signs. We're characterizing part of this discussion in terms of ecosystem impacts, let alone exogenous organisms.

It's not easy to predict these things. Connecting that to functional outcomes is different. Aside from a few simple cases, it's difficult to pinpoint these effects. Trying to put numbers or probabilities on a transfer event is exceedingly difficult. Talk about when transfer takes place in a way that increases fitness, and then it becomes more complex.

G. BIOMES

What's a healthy microbiome?

Only answer is that if we can enhance fitness of E. Coli, if you enhance it such that it competes when it shouldn't, then it will disrupt existing internal balances of microbiota. As a regulator it

would be great if I didn't have to think about gene transfer. In gene transfer entered a realm of uncertainty and unpredictability. If the purpose of this chassis is to not have to worry about transfer, what frequency of gene transfer actually happens productively?

- Examples of gene transfer can only be studied on evolutionary time scales "Highways of gene sharing among prokaryotes." Inference is that there are things more likely to allow you to metabolize products.
- Take a gene-centric approach to assessing risk. As we get better at synthetic biology, focusing on genes and genetic circuits will be more relevant than bugs themselves.
- Lots of bacterial sequences are actually viral. Not always adverse, depends on what the sequence is and where it transfers to.

Imagine a genetic circuit that expresses a nuclease. Nuclease destroys the DNA before lysis. We cannot eliminate the selective advantage of a piece of DNA. Solutions to prevent horizontal gene transfer are almost impossible. Need to focus on making sure that the genes are okay to transfer.

After discussion of what we know about natural bugs, what happens when you return to the world of the artificial? What kind of tests do you wish you could use?

Three genes key to placental functioning comes from viruses. They have a huge environmental impact. The time lag is important to think about. Evolutionary terms may not be useful for us.

- Steve Benner has done good work on nucleotides. The other organism won't have system that would allow you to use a transferred piece of DNA. A type of alien biology, increasing the gulf between synthetic and natural biology.
    - Japanese researchers discover resistance transfer factor on plasmids, this goes way back. A thorough review of horizontal gene transfer documented that the gut is the most common place for gene transfer to occur.
- Gene transfer is going to happen or a eukaryote could gobble it up, cf. the origin of mitochondria or *Prochlorococcus* requirements for metal.
- Various populations in Atlantic controlled by cyanophages, raising issues of transduction.

Antiviral, antibiotic resistance traits are obvious examples of potential significance, what else can insurance care about risk wise?

What is the program of testing you would then lay out to ensure insurers that what you're doing is safe? As a designer, it may be more meaningful to hear what would be convincing evidence to regulators?

Gene transfer is going to happen. We're never going to be down to zero gene transfer. Are there examples from GMOs that illustrate how much gene transfer is a problem? The question becomes: what frequency is acceptable?

- Let's put it into the context of antibiotic resistance instead of being too abstract. Could an antibiotic resistance gene be put into rE. Coli and then test the relative frequency with which the bacteria uptakes antibiotic resistance relative to in the natural environment?
- Measurements of rE. Coli need to look at the transfer of function, not just the DNA.

For engineered organisms, specifically with a chassis or other features, what is the half-life of that in as many environments as possible? What is the likelihood of gene transfer? What is the

phylogenetic novelty of the function? That is to me the unknown risk. It may still be benign, but it's unknown. In contrast, if it's been out there in evolution it may be less of a risk.

Chemical process and nuclear industry have made an argument that x level of risk is acceptable due to comparison with highways, but history of selling that to the public is not very successful. Public doesn't want involuntary risk imposed on them. From a regulatory standpoint and risk assessment standpoint, the idea that you might be able to estimate a risk may have merit but may not be bought by the public.

It's not just the demands on developers, but vice versa. They have a right to place demands back on the systems. You can't measure everything.

Chicken and egg: what do we expect from them? It costs to measure. We need to determine tests and protocols that all find acceptable.

The EPA handles this by asking that the clients talk to them 1-2 years before they plan to deploy, explain how it works, where you expect to be, and ask them questions and they'll try to give insight about what you should be thinking about. A menu of a bunch of things people should be thinking about. Does this make sense? Is it appropriate for my organism?

H. REGULATION

From a regulatory standpoint we have different categories of questions:

- What organism are you dealing with?
    - Starting with a non-pathogenic host, make sure you can show us documentation that you know what you're working with.
- What does it do?
- What have you done to it?
- If you say that it exhibits a certain amount of stability, can you show us how you know this?
    - We're going to ask about inserted material, potential for gene transfer, and see if you've addressed these kinds of questions.
- Are there explicit hazards associated with your organism that you know about that we should know about?
- Is it going to be released?
- How much will be released and by what modes?
- If you want to limit release, what have you done to limit it?
    - And again, how do you know this is going to work, have you done any testing?
- We usually see people working fermenters. But in algae in open ponds it's a totally different scenario. Have you thought about where your ponds are (esp. the receiving environment around the ponds)?
- What's your weather going to be like?
    - Are natural disasters common?

Back to a case-by-case study, do we really care about the trait being transferred? It is hard to generalize for any possible scenario. Humans, animals, and plants all have gene flow and dissemination issues.

There is a question of how to increase efficiency of the conversation 2-3 years out. How are we scoping out claims that they should be making insurance- and regulation-wise. How can we make it more general to the process? What could streamline and make this more transparent? What might get in the way of this process, specifically IP?

What evidence do regulators seek, hypothetically?

Depends how the organism is constructed. If it's on a plasmid it might be more demanding than if the engineered functions are being incorporated into the genome.

Since this is case-by-case. Can the inflow of new organisms exceed EPA's capacity?

Probably by next week, it will. The sheer speed of innovation we can foresee may overwhelm the current system rather quickly.

Is there some point you can generalize?

It's one of the things we have to do. We're not ignorant of what's facing us. Industry often talks to us 2-3 years in advance, to try to anticipate. We use information tools that will allow us to make do with the information resources we have. There may also be familiarity if the same chassis was used over and over again. Even if the genetic modifications will be slightly different, we can extrapolate from something that's gone through the system more easily. If someone has come through a year or two in advance, we don't have to wait until it shows up on our doorstep, meaning less delay since we will have already started working on it.

I. NGOs

What do you guys want to know?

NGOs want to know that there are appropriate regulatory frameworks in place that can really assess what's happening in this field. We do have a lot more concern about long-term environmental effects of the technology. We also have heightened concerns about effects on humans if they are exposed to it. In this meeting we haven't talked about workers and worker exposure. There was a genetic engineer working for Pfizer who got a 1.4 million settlement for infection by lentivirus but didn't get any medical records. Overall:

- We have to make sure review frameworks fit what you're trying to do.
- Transparency – what makes environment sick shouldn't be a private business matter.
- What we think are difficult current problems to which we don't know the solution are the process issues, risk frameworks, etc. What we've talked about are generalizations and we have to realize that when we look at a specific organism it's very context-dependent. But although we may be getting better at getting a lot of parties in the room to talk about generalities there are not a lot of opportunities for other people to get a comment in, i.e.

Federal Comment might come much later on. It would be good if there were a mechanism by which other concerned parties could be involved more and earlier on specifics on instead of on generalities, which only gets us so far.

If you were to request tests, what do you want to see?

Using the example of the Department of Energy biofuels program, we would want a rigorous comparison of old vs. new ways of getting to these biofuels. The equivalent of natural and organic vs. genetically engineered in this field.

J. FLAGGING SOURCES OF UNCERTAINTY

ID areas that need more research. Highlight key points of uncertainty that need to be addressed.

- What's the problem with gene transfer?
- What are its effects?
- Which questions need the most attention?
- The regulation answer is 'sure, we care and it happens, but how much do we care?'
- What is that actual problem with gene transfer?
- What is the impact on communities? Not just the impact on the organisms it's designed to, but the ecosystem as a whole.
- Hit and run event, how do you figure that out? This is an interesting measurement question.
- Time lag question, 1 year, 2 years, 20 years. How do you even begin to develop a model to figure that out?

What other topics? And we'll talk about what to discuss?

- Agent-based models need to be looked at. Questions of community invites talking to population or ecosystem dynamics experts ask them to what extent they can help with these problems.
- One can imagine something so fundamentally different from biology, but the reason we do biology is to do chemistry. So there is still metabolic interaction with the environment. Even if it doesn't have any way of transferring DNA to environment, still capable of forming metabolic communities, and metabolite networks.
    - Note that this question highlights tracking metabolism and metabolites: 22 billion potential organic molecules.

What's the problem with persistence in the environment?

- So far, persistence isn't a big issue in corn and soybeans.
- Canola oil, however, has plants growing well out of where it started.
- Similar problem will occur with microbes.
- Biofuels have a greater likelihood of persistence because they survive well; they are not domesticated like corn.
- Persistence of gene, not just the organism
    - So, an organism can act like a vector; can move a gene/function to other locales where it could cause a problem.

Synthetic biology wants a more formal function tree, how can the capacities like that change the discussion? Generalize risk of more complex function.

How can we attack these sources of uncertainty, what research is useful?

All of these issues seem to be related to the natural reproductive cycle. We need to have indexing with respect to these reproductive cycles.

What is the impact of genes on community? If you had infinite access to funds, and were trying to tackle this, what would you do? Can they survive? How fast will they reproduce? How many will there be?

We should start with microcosm/mesocosm research. It should be ongoing in a circular fashion. Also look at enhanced photosynthetic activity with new ways of fixing carbon. We should always assumption kill switches don't work.

- Any time you work with a kill switch you should assume the selective pressure will be strongly against it.
- Connections among three of these questions: effects on community, effects on metabolic functions of the community, effects on functions. Take a look at metabiome and see what's changing.

What is the community?

We have little understanding of how this community works, let alone with exogenous organisms or global climate change. We are going to have to be very careful in determining this in relation to the community beforehand to measure potential impact relative to an actual baseline.

If you revisit the old literature that conditions our knowledge of all these issues, what needs to be redone with new methods?

- There is an interesting gap: thinking about pollen flow and dispersal. The literature at the time was not robust enough because detection wasn't sensitive enough. Thinking about mechanisms of dispersal informs the idea that communities you may impact may be only a migratory path away.
- We often reference older information technology. We have obvious changes in measurement and observational methods—they have improved tremendously.
- The older studies asked 'who's there,' but determined it by culturing, we should do metagenomics not cultures.
- New techniques might look at other mechanisms of impacting the microbiomes we're interested in, in comparison with what we're proposing to do, perhaps old practices have a bigger footprint than what we want.

We need gene transfer experiments that we can't even culture, not just assuming we have one donor one receptor in soil in a beaker.

- We just need more data here.

Further studies of directed evolution are needed to see what they're capable of evolving into. These studies should be done with and without other species around.

- Derek Welby's lab, directed evolution of co-cultures is a very hot topic research. We should see if we can get something better. We see microstructures evolving in ways which are fantastic.

Do we need a tool to anticipate human disruptive behaviors? Do we need to integrate that at this level or does it come necessarily later? Is there an instrumentation problem with this? Saw with nanotechnology that ENGS research is lagging because of problems of metrology. This is a big funding issue too, making sure we have metrology research now.

We need more research into appropriate bioreactors that can mimic other environmental variations and patterns.

Scale needs further research and testing. Especially things like dose-response. If you're using your organism to make specialty chemicals in small quantities is that different from when you have huge tubs when a hurricane comes through? How important is scale to dispersibility?

What about Algal biofuels?

We've been growing algae in aquacultures for a long time. No one has looked at issues of putting algae into a non-indigenous environment. Is there data useful with synthetic biology/GM algae? Is there anything unique?

- We need to know more.
- There's confusion over what's indigenous or not.
- Certain strains of algae are grown in Israel in large open ponds. Contamination by other organisms is not an issue, not grazed on, large basis. This algae has some very unique stress-responsive genomic characteristics: it cannot be transformed, sequesters CO_2 20% by weight, has a tough cell wall, so it's difficult to remove. Also, it doesn't affect the Red Sea.
- *Synechococcus* in Hawaii e.g. has worked, and naturally.
- Gets back to understanding the organism you're working with.
    - Understand you can't know everything.

Call for upstream research:

- What sustains diversity?
- What is required for survival?
- Can we turn this into a predictive understanding?
- Currently constrained by evolutionary historical constraints, but we can use synthetic biology to understand this more theoretically.

This should all be seen as experimental biology, with the goal of understanding living systems better. We need better cultivation of the data on successes and failure.

We're all assuming the natural organism escapes unintentionally. What about security of organisms or robustness to change perhaps due to another human being changing them?

- The issue of malevolent reanalysis and modification being done more easily due to modularity. Gautam Mukunda, Ken Oye, and Scott Mohr did a piece on biosecurity and modularity (What Rough Beast: Synthetic Biology, Uncertainty, and the Future of Biosecurity).

- DARPA is funding solutions for identifying engineered organisms.
- Serious repercussions are possible as our economy becomes more dependent on this field.
- Malevolent misuse is a major concern.

Perhaps we need a symposium on what kinds of species have the highest benefits and highest risks for biofuels? Gene flow, harmful toxins, environmental effects, etc.

The algae working group's assignment has been to develop a list of potential commercial algae. The group has been told by Biomass R&D board that this should be expanded to an information resource with all algae identified. Next fiscal year, definitely want to do that.

- Working group has no budget, yet. The board may do it, assuming approval.
- DOE should be funding this.
    - This is a moving target, since most people are working on strain discovery.
    - This is why it should be a dynamic information resource rather than a static document.

It is well worth the investment to convene land-use/land-cover change community and talk about where the algal fuels would be planted. Because we don't know where there'll be, this causes a lot of generalized NIMBY backlash. This is a good, cheap, early investment.

USDA has a risk assessment research grant program. The program has $4 million per year, and is generally focused on algae-related issues, and is congressionally mandated. Those who are interested—there is conflict over whether this is product development disguised as a risk assessment project. This research goes through peer review every year. EPA and USDA sit on review panel, more for TSCA and EPA folks.

K. WHITE BOARDS:
Board 1
What is meant by safety?
Preventing harmful outcomes/unknown outcomes-gene flow
safety to environment-overtake natural species and change areas
so is safety undefinable?
Health:
human
agriculture
atmosphere
What do we include in the definition of natural world?

What does insurance want to know?
What is the risk landscape?
What are the different values (monetary) that can be affected?
human
animal
plant
financial
Are there peer risks?-ongoing monitoring

What are the microcosms that the organism could affect?
    millions $/yr for testing

Fitness
changing interactions with organisms
    symbiotic relationships
    can result in a net positive fitness in environment, even if individual tests show weakness
Testing
what do we want to accomplish with these tests?-assays? Differences between rE. Coli/cyanobacteria?
    what are the end products?
What are the functional elements?
Range:
    environmental conditions
organism consortium or land changes
    non-engineered organism vs. engineered organism
    virus resistance
What is fitness?
How will it compete?
    reproduce
different conditions
    number of offspring
        copies of gene
Survival
Evolution of measurement
sequence technology
gene flow/stability
PCR development
RNA sampling
Marine Eukaryotes
    organism gets outgrown
    engineered gets out competed
    contamination biggest issue for large scale cyanobacteria (according to industry)
Exposure issues
    leads to fitness?
Gene transfer itself is not a risk, what happens after is what matters (insurance)
what changes need to be made to our methods?
Design-build-test under synthetic biology is accelerating and becoming more predictable
    this changes risk footprint

Board 2
Hit and run events:
    •What about occasions when bacteria enter, cause damage, and then leave?
Persistence of an organism in the environment:
    •What are its immediate impacts?
    •What are the overall consequences of the organism persisting in the environment?

Persistence of transgenes:

- •Algae farms may be a good test case.
  - •What is currently known about the persistence of transgenes?
  - •How can we compare what is known to synthetic biology?

Stability:

- •Is there an economic drive for stability?
  - •How strong is this incentive?
- •What about conditional stability?
- •How should we define the function(s) of an organism?
- •How can we best match an organism to an ecosystem?
- •What about gene flow into an organism?

Wishful tests:

- •Scale is a limiting factor.
  - •Field tests come after several iterations of microcosm tests.
- •What do we measure?
  - •How the novel gene functions
  - •How the organism behaves in the environment
    - •What are the effects on other organisms?
  - •What are the impacts of the community?
    - •Does community structure change?
    - •Do organisms' functions change?
  - •What about metabolism?
    - •What are the recent updates on metabolic engineering?

Gene transfer:

- •What is the problem with gene transfer?
  - •What are the effects?
- •Knocking out protein function
- •How is gene moving?
- •What is this protecting from?
  - •Horizontal Gene transfer
- •How can we perform modularity testing?
- •Can we create a modular approach to risk assessment?
- • Should we be following a gene centric approach?
  - •What about episodes of individual transfer?
- •Each case will depend on gene traits, as each is separate and distinct

What are we designing vs. what are we preventing from transferring?

- •What is the frequency of transfer?
- •What is the transfer function and mode of transfer?
- •What is the phylogenetic function of the transferred material?
- •Do these different methods have different environmental effects?
  - •Can we compare to genetic engineering?

Random insertion vs. surgical improvements and the ability to synthesize DNA:

Lab level:

- •What are the mechanisms that currently exist?
- •Is this a chicken and egg issue? Does it travel from regulators, to developers, and back to regulators?

Scale:

- •How important is scale?
- •How does scale affect persistence?
- •How do different scales change local communities?

Time lag:

- •This can use this to further understand evolutionary biology
- •Time lag can be used for different tool applications
- •As the technology advances the design-build-test system will move faster and faster

Biosecurity of organisms:

- •In the near term we need an overall ranking of algae species

Synthetic biology capacity:

- •Going forward, how will risks change?
- •What will be the changes in functional transfer?

Research Questions going forward:

•What are the time scales we should be looking at?
•How does this new technology measure up as a new way to fix Carbon?
•How will synthetic organisms change natural metabolism?
•What will be the effects on population and ecosystems dynamics?
•How should we measure the impacts of synthetic organisms?
•What are the other ecological factors (runoff, etc.)?
•How is this different or the same as directed evolution?
•How can we account for disruptive human impacts?
•How can we model the natural environment in bioreactors?

Summary Report of the Meeting to Discuss Data Needs and Testing Methods for Assessing the Safety of Environmental Introduction of Synthetically Designed Algae for Biofuel Production

A Joint Workshop of the Woodrow Wilson Center, the MIT Program on Emerging Technologies, and the U.S. EPA

December 14, 2012

Report Date:
February 5, 2013

Report Authors:
Shlomiya Bar-Yam
Kelly Drinkwater
Todd Kuiken
Julie McNamara
Scott Mohr
Ralph Turlington
Kenneth Oye

NOTICE

This report was prepared by the Wilson Center and MIT PoET as a general record of discussion from the workshop *Data Needs and Testing Methods for Assessing the Safety of Environmental Introduction of Synthetically Designed Algae for Biofuel Production.* This report captures the main points and highlights of the meeting. It is not a complete record of all details discussed, and it does not interpret or enlarge upon statements made over the course of the meeting that were incomplete or unclear. All included points represent the individual views of meeting participants and should not be viewed as a consensus. Except where specifically noted, no statements in this report represent analyses by or positions of any of the meeting hosts or report authors.

MEETING SUMMARY

During 2011 and 2012, The Woodrow Wilson International Center for Scholars (Wilson Center) and the MIT Program on Emerging Technologies (PoET) co-hosted multiple workshops studying varying aspects of synthetic biology. Discussions repeatedly illuminated conspicuous data gaps throughout the field, with such uncertainties often significantly impeding forward progress on safety and security discussions. These findings, in combination with an interest by the U.S. Environmental Protection Agency (EPA) to more explicitly consider the ramifications of such gaps in the face of future Toxic Substances Control Act (TSCA) applications, led to the organization of this meeting, *Data Needs and Testing Methods for Assessing the Safety of Environmental Introduction of Synthetically Designed Algae for Biofuel Production*. Co-hosted by the Wilson Center with funding from the Alfred P. Sloan Foundation, MIT PoET, and EPA, the meeting was designed to bring together industry, academics, non-governmental organizations, and several agency stakeholders to discuss what the relevant data gaps are—and how they might be addressed—when considering the implications of environmental release of synthetically engineered organisms. For the purposes of considering a tangible concept, the meeting was conducted specifically through the lens of algae engineered to produce biofuels

Contents

Agenda

This workshop provides an opportunity for academia, industry, government, and non-governmental organizations to improve scientific understandings of ecological issues of relevance to evaluation of synthetically designed organisms, and to identify what methods exist or should be developed to assess the safety of a field release regardless of statutory or regulatory mandates. This workshop is one of a series dealing with scientific issues surrounding synthetic biology put on by the Wilson Center and the MIT Program on Emerging Technologies, with the support of the Sloan Foundation and the NSF Synthetic Biology Engineering Research Center. These workshops have assessed risks; identified scientific uncertainty associated with synthetic organisms and their interaction with the environment, and developed research agendas to address some sources of scientific uncertainty.

Session I: Introduction		8:00-8:30
Welcome	David Rejeski, Woodrow Wilson Center	
Overview of Schedule	Kenneth Oye, MIT and NSF SynBERC	
Self-Introductions	All Participants	

Session II: Overview of Current and Emerging Industrial Applications		8:30-10:15
Synthetic Genomics	David Hanselman	
Algenol	Pat Ahlm	
Sapphire	Yan Poon/Tim Zenk	
UCSD	Stephen Mayfield	
Agilent Technologies	Stephen Laderman	
Discussion:	What are the properties of algae optimized for biofuels production?	

Break 10:15 – 10:30

Session III: Overview of Current Review Process		10:30-11:00
EPA perspective	Mark Segal	
DOE perspective	Daniel Fishman, Kristen Johnson	

Session IV: Previous Workshops - Addressing Data Needs		11:00-11:45
EPA findings from 1990s workshops on biotechnology		Gwen McClung
Wilson Center and MIT findings from previous workshops		Todd Kuiken

Session V. Identification of Ecological Endpoints to be Assessed 11:45-12:00

Lunch 12:00-1:00

Session V. Identification of Ecological Endpoints to be Assessed 1:00 – 3:00
Defining potential receiving environments (terrestrial, freshwater, marine)
Defining endpoints within these environments
Defining immediate vs. long-term data needs to assess endpoints
Defining minimum data set needed prior to any environmental introduction vs. data set needed for large scale acreage
Panelists: Rex Lowe, Bowling Green University; Bruce Tonn, University of Tennessee; Kent Redford, Archipelago Consulting; Robert Stevenson, Michigan State University; and others.

Break 3:00-3:15

Session VI. Methodology & Protocols 3:15-4:15
Methods/Tools
Instrumentation

Session VII. Wrap-Up 4:15-5:00
Summary of data needs
Summary of instrumentation needs
Identification of areas of uncertainty
Identification of research paths to address uncertainty

List of Acronyms

ASU	Arizona State University
CBI	Confidential Business Information
CWA	Clean Water Act
DARPA	Defense Advanced Research Projects Agency
DNA	Deoxyribonucleic Acid
DOD	Department of Defense
DOE	Department of Energy
EISA	Energy Independence and Security Act
EPA	U.S. Environmental Protection Agency
FIFRA	Federal Insecticide, Fungicide, and Rodenticide Act
FISH	Fluorescence *In Situ* Hybridization
FOIA	Freedom of Information Act
GFP	Green Fluorescent Protein
GM	Genetically Modified
GMO	Genetically Modified Organism
HAB	Harmful Algal Bloom
HGT	Horizontal Gene Transfer
HPLC	High-Performance Liquid Chromatography
JBEI	Joint Bio-Energy Institute
MIT	Massachusetts Institute of Technology
NIH	National Institutes of Health
NREL	National Renewable Energy Laboratory
NSF	National Science Foundation
ORNL	Oak Ridge National Laboratory
OPPT	Office of Pollution Prevention and Toxics
PBR	Photo-bioreactor
PERF	Petroleum Environmental Research Forum
PNNL	Pacific Northwest National Laboratory
TERA	TSCA Experimental Release Application
TSCA	Toxic Substances Control Act
UCSD	University of California, San Diego
USDA	U.S. Department of Agriculture

Summary of Session II: Overview of Current and Emerging Industrial Applications

Representatives from Agilent, Algenol, Sapphire Energy, and Synthetic Genomics provided brief updates to inform workshop participants of the present state of the industry. Following, new research findings and possible future collaborations in the development of measurements and standards were presented. In closing, the floor was opened to all participants to discuss the characteristics of a hypothetical "ideal" organism.

Industry Update

Industry representatives covered three main areas in their presentations: methods of organism development, approaches to cultivation and containment, and mechanisms for hazard assessment. Several questions were also raised without immediate resolution.

Organism development

- Researchers aim to first identify organisms naturally displaying desired characteristics, including through conducting bio prospecting ventures in varied environments around the world. Wide-ranging screening has been made possible through metagenomics and high-throughput analysis of promising strains.
- Follow-on genetic manipulation includes support of natural and directed evolution processes. Active techniques include radiation and pathway engineering.
- Once organism performance has been tested within the laboratory, the top performing strains are moved on to larger-scale trials (e.g., pond screening). A field-validated strain is one shown to grow robustly in the field and proves capable of cultivation; a production strain is further treated to increase pest tolerance. For example, one company reported taking a production strain, subjecting it to several rounds of mutagenesis, and ultimately finding that the evolved line grew better (noting that the parasitic fungi originally in question continued to grow). Pond crashes were noted as occurring over as short a time as one to two days.

Cultivation and containment

- Cultivation and containment methods vary depending on a company's biofuel production technique. For open pond production, cultivation is akin to farming, as crops must hold up against threats such as pests and weather. For photo bioreactor (PBR) systems, most external cultivation threats are controllable.
- Containment methods can be broken out as biological and physical:
    - **Biological.** The hazard assessment process (described in the following section) precedes strain scale-up and aims to eliminate the most overt of biological threats, such as invasiveness and toxicity. One pilot facility is beginning to test the invasiveness of strains in each type of water an escaped organism would encounter prior to reaching the ocean.

Preliminary findings have seen no evidence of invasiveness, although the screening has only newly begun and the adequacy of the mesocosms has not been verified.

- **Physical.** With PBR systems, physical containment is focused on structural soundness of the PBRs to prevent leaks; concrete pads and earthen berms to prevent spreading should a leak occur; and comprehensive treatment of effluent. One company cited physical containment levels designed to meet 500-year storm threats. For open pond systems, operations proceed under a general assumption of field release. One company noted that while birds and other creatures have been observed along the pond edges, it is only now moving forward in a partnership to develop monitoring tools to better characterize their presence. Finally, one open pond company is currently trialing the use of unlined (soil only) ponds.

- One company posited that metagenomic analysis could serve as a useful tool for studying an environment prior to release. Citing a study finding significant reductions in species diversity around power plant effluent as compared to in a mangrove swamp, the company noted that receiving waters could be tested for species diversity before, during, and after release as part of the general monitoring process.

Hazard assessment

- Once a species has been identified as of interest, a hazard assessment is conducted to ascertain its practicality as a commercial starting point. Several properties are instant disqualifications, including one company citing risk level rankings above Biosafety Level (BL) 1 and another screening through bioinformatics analysis to evaluate the presence of enzymes required to produce known toxins.
  - Multiple presenters noted that the hazard assessment process is regularly stymied by a lack of available information. While much information is available on a select few strains (i.e., those responsible for repeated harmful algal blooms and those already employed in commercial processes), little is available for others. Further, general taxonomy has become increasingly complex as actors have repeatedly shifted between "good" and "bad" groups.
- One company reported that of approximately 40 high level hazard analyses conducted at the genus level, only a handful have subsequently resulted in strain abandonment.
- In terms of hazard assessment, one company suggested that there was little value in identifying a strain as "native" or "non-native" based on state-level communications.
- Post initial strain selection, companies noted performing various types of horizontal gene transfer (HGT) studies prior to advancing strains further. Additionally, beyond the initial bioinformatics analyses, high-performance liquid chromatography (HPLC) is routinely performed to assure toxins are not being produced.

Other

- During the presentations, multiple companies highlighted data and/or knowledge gaps requiring further attention:
  - Additional information on taxonomy is needed, as the data are highly valuable for understanding risks yet are lacking in multiple areas.
  - No good exposure assessment model exists, though it is important and would be useful.
  - Does a well-characterized "bad bug" list exist? For example, a compilation of organisms and parts that should be avoided? One company stressed the importance of companies facilitating a collegial sharing of information so as to better advance the industry as a whole.
  - If a strain is completely non-toxic, should zero release be expected?
  - If an organism displays a three-fold increase in growth rate, what would be the implications upon escape into the wild?
  - If escape results in the replacement of one species by another, would that be considered harming the population in a substantial way?

New Research Developments

An algal researcher and product developer presented a brief summary of recent advances in the field. In particular, the participant emphasized the following key areas:

- Whereas much of the original algal biofuels research focused on freshwater species, such advances are now being tackled with saltwater organisms.
- Regulators will need to consider a wide range of products from synthetically engineered algae in the future, as many applications beyond biofuels are advancing toward commercial production. For example, the researcher explained a successful trial algal production of nutrients traditionally found in colostrum.
- In developing countries, it is unlikely that products will be able to bear the additional costs associated with production in PBRs; therefore, it should be assumed that applications will be produced in open ponds in such locations.
- Presented data displaying the successful incorporation of a synthetically engineered gene into another organism (in this instance, involving sensitivity to high- versus low-light).
- Cited research by Susan Golden identifying four traits that successfully decrease grazers.

Tools for Developing Methods and Standards

A representative from a company specializing in measurement methodologies provided an overview of their technology development process alongside emerging technologies, and highlighted some possible areas for collaboration in the algal field.

- The representative emphasized the importance of identification of current unknowns and data needs within the industry so as to allow for targeted product

development, and noted the utility of public-private partnerships in such endeavors.

- As examples of current applicable technologies, microarrays—enabling the development of large libraries of sequences as well as for genome partitioning products—and oligo fluorescence in situ hybridization (FISH)—allowing high sequence specificity for targeting microbial applications—were described.

Characteristics of an Ideal Organism

In a discussion of ideal organism traits from a *production* perspective, workshop participants built off an initially prepared slide highlighting eight broad areas likely to be of focus. Following discussions, clarifications, and some points of debate, the following list was derived (with qualifiers noted):

- Enhanced photosynthesis
- Enhanced lipid production
- Rapid growth
- Enhanced nutrient uptake, production, or utilization
- Enhanced survival in pond monoculture (e.g., resistant to herbicides, pests, and pathogens)
- Increased tolerance to adverse environments
- Allelopathic [*allelopathy:* the inhibition of growth of one species of plants by chemicals produced by another species]
- Geared toward more cost-effective sections of supply chain
- Genetic malleability, or increased ease of modifying genomes (at least initially to facilitate further strain modification)

The topic of biological containment mechanisms was discussed, though few specifics arose owing to significant knowledge gaps remaining in the area. Additionally, participants debated the merits of ease of organism traceability, though no resolution was reached. Finally, many participants expressed concern that the list was not ideal when considered from the perspective of environmental concerns. However, the participants were reminded that these traits were only being gathered so as to be able to better focus discussions later in the day regarding understanding possible ecological endpoints of modified traits.

Summary of Session III: Overview of Current Review Process

Regulators from the Environmental Protection Agency (EPA) discussed how existing rules, laws, and mandates might define the agency's role in regulating synthetically bioengineered algal biofuels. While a number of different regulatory mandates give EPA potential jurisdiction in this area, none clearly defines EPA's role or establishes a set of activities and criteria to guide such a role.

The specific laws that were discussed as potential mandates for EPA regulation of synthetic algal biofuels include the following:

- Energy Independence and Security Act (EISA)
- Toxic Substance Control Act (TSCA)
  - TSCA may apply because "new" microorganisms (depending on how this is defined) can fall under the rubric of "new chemicals," which TSCA grants EPA jurisdiction over in the context of manufacturing, importation, and research and development for commercial purposes. TSCA would also require the submission of an Experimental Release Application (TERA) 60 days prior to the introduction of microorganisms to an uncontained commercial facility.
- Clean Water Act (CWA)
  - CWA might apply because engineered organisms could be considered "pollutants."
- Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA)
  - FIFRA would apply to any disinfectants or pesticides used in the commercial growth process. At present, however, despite the existence of these several potential regulatory mandates, EPA's role is relatively undefined.

EPA's regulatory role in this area will depend substantially upon the novelty of a bioengineered organism. This, in turn, will hinge on what definitions and standards are put in place for synthetic biology and the criteria for determining whether and how genetic modifications lead to an organism possessing "new" characteristics.

EPA would likely have to expand its assessment capabilities, develop new areas of expertise, and develop new standards in order to keep pace with the current and expected pace of innovation in algal biofuel research, development, and production.

Representatives from the Department of Energy (DOE) approached the issue of synthetically engineered algae from a very different perspective than EPA. DOE's mission in this area is driven by the government's priorities under EISA and other acts in promoting the development of new sources of fuel that can effectively substitute existing fossil fuels. DOE's primary focus has been to support industry initiatives to promote the

development of new technologies by offering sources of funding, and by partnering with industry to promote more effective research and development approaches.

In the regulatory area, this consists primarily of DOE offering guidance and assistance to private firms on maneuvering through government regulatory requirements, establishing research and development protocols that best minimize and address regulatory barriers to technological innovation, and cooperating with other U.S. government agencies to facilitate regulatory transparency and compliance.

DOE has also sponsored more limited work at the national laboratories on developing criteria for the assessment of new synthetic biology applications in the production of algal biofuels. This has consisted primarily of the development of a set of indicators focused on environmental and human safety.

Summary of Session IV: Previous Workshops – Addressing Data Needs

The December 2012 workshop is not the first time these topics have been discussed by EPA. In 1994, a three-day workshop was run by EPA that focused on bacteria and fungi, but not algae. As explained Gwen McClung, EPA Office of Pollution Prevention and Toxics (OPPT), Risk Assessment Division, the goal was to work toward developing different testing schemes. Some of the main conclusions from the 1994 report include the fact that TSCA does not have any specific testing requirements. It has specific informational needs, but there is no hard and fast set of rules to acquire these data. This means that regulating a GMO under TSCA is different from regulating it under the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), which has specialized, required tests. The 1994 workshop also differentiated between environmental and ecological effects of an organism. Risk assessments should be based on an organism and its respective modifications that could alter its behavior in the environment (exposure x hazard = risk). These tests would be performed in microcosm tests in tier 2. When evaluating a GMO, EPA is limited to focusing on the ecological issues associated with the immediate environment in which the organism is introduced.

1994 Workshop Notes

In 1994, the U.S. Environmental Protection Agency's Office of Pollution Prevention and Toxics, Health and Environmental Review Division, Office of Research and Development, and Environment Canada, Commercial Chemicals Evaluation Branch, sponsored a workshop to develop ecological tier testing schemes for genetically engineered microorganisms. There was general agreement that the "potential ecological impacts of microorganisms released into the environment have not been well characterized." Ecological effects endpoints identified as potential areas of evaluation included:

1. Effects on Primary production
2. Effects on cycling of limiting nutrients
3. Effects on community structure and diversity
4. Effects on community function
5. Trophic level changes / effects on grazers
6. Effects on sensitive species

Based on the discussion, the group of participants developed a tier testing scheme (0-3) to evaluate genetically engineered microorganisms in closed, semi-closed, and open applications. Tier 0 contains preliminary information, taxonomic identification, proposed use, and site characterization. Tier 1 contains initial exposure and hazard assessment components such as persistence, dispersal, pathogenicity, toxicity, and basic ecological effects. Tier 2 addresses additional questions about exposure and hazard from Tier 1 and contained longer term and more complex ecological effects testing. And Tier 3 contains open or limited field tests in the selected environment.

The 1994 workshop was followed in 1996 by a workshop on testing methods. The goals of this workshop were to identify protocols to test GMOs prior to release and to develop standard operating procedures to guide data collection. Some resulting major questions included:

1. Is this organism temporary or will it persist in the environment?
2. Will there be a recovery in the organism's population after an unknown amount of time?

These questions were asked nearly 20 years ago. What is still relevant? What additional information is still needed for us to be comfortable releasing microorganisms into the environment?

More recently, the Wilson Center has hosted workshops on both rE. Coli and cyanobacteria. Some of the major ideas raised during these discussions included considerations of fate and transport of DNA, modeling gene transfer, and understanding the respective time lags involved.

Findings from the previous workshops highlight the still-existing broad areas of uncertainty in the field. Some of the topics discussed more recently are very similar to topics discussed in the 1990s. Overall, stakeholders must be aware that history will repeat itself if it is not sufficiently studied. It is important to now come up with a clear set of issues that need to be worked out in order to properly test these organisms.

The system must be thought of in totality as opposed to simply comparing organisms' interactions with one another. There must be objectives and ecological endpoints that are considered. Overall, how can these organisms be tested for the long term? How many "cycles" are needed?

Arizona State University (ASU) is developing and validating methods that deal with these questions. At least one annual cycle is needed, possibly more. However, these are all context-dependent and if there are more unanswered questions that arise, more data will be needed.

What about the by-products of these facilities? One industry representative noted that at present, all by-products are returned to the productions process. Another stakeholder noted that there is also an algae interagency working group that is currently looking at the use of algae in animal feeds.

Summary of Session V: Identification of Ecological Endpoints to be Assessed

This discussion considered immediate and long-term data needs for algae synthetically engineered for biofuels production.

Broad receiving environments: terrestrial, freshwater, marine

- Relevant questions: What are the options? What do we expect to be coming? Where will facilities be located and what is the effect of location?
- There are applications in bags, raceways lined and unlined in the desert, and potential applications near waterways.
- For companies based in deserts, which a number are, early applications will potentially be in the desert.
- There is a need to take things stepwise to gather information to fill in blanks before proceeding to environments that are more difficult from a knowledge or risk angle; i.e., not jumping right into deploying by the ocean.

Identification of locations and environments that participants viewed as "good" for deploying the technologies

- An ideal location may have: water, abundant sunlight, CO_2 from power plants or other sources, nutrients, land.
    - Resource availability is important: Lands must meet certain criteria, such as availability of saline water in sustainable supply. Availability of CO_2 is important: CO_2 is one of the most expensive inputs today. Nutrients are also important (N, P, K).
    - Co-location may be an option. Nutrients and other inputs can be sourced from farming.
    - Pacific Northwest National Lab (PNNL)/National Renewable Energy Lab (NREL)/Argonne have techno economic analysis work on availability of resources, standards, fresh and salt water, specifically for algae. It may provide a framework for making good decisions.
    - An industry member noted that his company is focused on inland solutions, and that coastal applications would involve very different technologies.
- One good option is land that cannot be used because of former pollution:
    - People at DOE have tried to develop those lands for other renewable energy (e.g. wind, solar).
    - What are issues there compared with a pristine (e.g. desert) site? The polluted site is already polluted.
    - There are also lands that have previously had agricultural activity but are now degraded to the point of no longer being useable for such purposes

(e.g. salted too much). The environmental effects have already happened. The USDA does not have a program for treating such land.
    - Use of polluted/degraded lands would not satisfy environmental groups, due to concerns about the organisms getting out. The organism would need to be shown to be safe.
- Identification of other examples of sites where algae facilities could be beneficial or less harmful:
    - In Florida, citrus groves are dying near a CO_2 source. There are thousands of unused acres.
    - Some participants noted that water is being removed from the Salton Sea and argued that as the water recedes, dried material will blow to populations and become a medical liability, and thus an algae pond on top of it would be a benefit to mitigate environmental disaster. Some participants representing environmental concerns argued otherwise, particularly if there are endangered species nearby (e.g., pup fish in nearby hot springs).

Identification of locations and environments that participants viewed as "bad" for deploying the technologies, and limitations

- Identification of types of locations to be avoided:
    - Algae facilities should avoid collocating with another algae facility, e.g. cultivating algae for fuel production near algae for food or dietary supplement uses. Even if the genes do not integrate, if the food facility tests its pond and finds algae for fuel production, it loses all of its sales for a period of time, until it can prove that there is no more contamination. This is a tort issue and a location question.
    - A "bad" location is near where there is an endangered species, one that is water dependent, or worse, algae dependent. Environmental groups would want to know how the producer would verify no harm done to the endangered species.
- Identification of other examples of poor sites:
    - A participant identified an "unqualified entrepreneur's garage" as a worst possible location.
    - A participant identified Minnesota, as an area with many lakes and lower light, as a worst possible location.
- Identification of location limitations:
    - Land use rules in general act as limitations: Where can one put a hundred thousand square foot facility?
    - Open or semi-open ponds today are limited to areas without extreme cold. Look at where everyone is locating facilities: Arizona, Florida, warm areas. They have to be able to run year round to make it economically feasible. If can make it yield twice as much, only half as much land is needed. NREL and PNNL are conducting a study on locations, overlaying many local conditions.

- There may be international treaty issues, for example: There are ducks regulated by treaty with Canada and Mexico that eat algae. They may want to know whether those ponds are replacing ponds that ducks normally eat.

Risks of engineered/synthetic biology vs. natural strains

- If natural strains used for biofuels production are located in areas identified as "problematic," are the natural and engineered strains so different that alternative conclusions on siting would be reached? Are there things about engineered algae that have stakeholders worried in comparison with natural algae?
  - Natural species are only regulated through land use laws. Now added regulatory/safety constraints are being introduced due to synthetically engineered traits.
  - The natural ones have been reproducing and living there for thousands of years.
- Most current-generation algal biofuel strains are generated using directed evolution. Are there concerns about directed evolution strains vs. natural strains? Should directed evolution be added to methods of creation to be concerned about?
  - Some participants argued that yes it should be added; it is not clear how concerned to be about it, but concerns exist. There are concerns about the effects of speeding up the rate of evolution.
  - A participant argued that directed evolution means speeding up one organism in relation to others, so directed evolution could result in a mismatch of competition, by creating an ecological imbalance in how systems react to change.
  - However, another participant argued that directed evolution is a means of finding "the needle in the proverbial haystack": Making the organism through directed evolution is just easier than finding it naturally.
- A participant argued that the idea that humans can make an organism that is better than what nature can produce is false. There are millions of organisms all competing with one another. Humans are the only organism that has outcompeted everything. Whatever humans could do is to benefit people, not provide the organism with a competitive advantage. There is no organism that can outcompete everything. The ones that produce damage are already there, producing harmful algal blooms. One wants to avoid making more of those, and must consider whether an algae that had never been in an environment before will cause perturbances when introduced.

Invasiveness, escape, and effects of natural dispersal on risk

- Some participants argued that algae have already been transported everywhere (e.g. by wind) so there is no need for concern about invasion.
  - There are examples of where algae are not as cosmopolitan, but as a rule they tend to be well distributed.

  - Some participants noted the uniqueness of the case of using a very localized/endemic algal strain for such applications. Does the potential uniqueness of localized algal strains factor into risk characterizations? Regulators may base a history of being able to use an organism safely on the scientific literature. However, a company's one submitted paper may be the only scientific literature on that organism. It would have to pass other tests: Could it be grown in a suitable environment, could the molecules of interest be made with it? A localized organism is probably too unique; it could not be grown in different places. Industry would want to find something easier to use. But lacking information, companies would need to study about it. This relates to discussions of the idea that the first-in-class gets more scrutiny.
  - Rock snot is an example of a plant exhibiting invasiveness in a new environment. There are more examples of that. A participant noted that rock snot is a naturally occurring species, unlike the engineered algae being discussed. Some participants argued that animals may be less widely distributed, so concern about animal invasives such as cane toads could be different from concerns about widespread organisms such as algal strains.
- Some participants argued that invasion is a concern.
  - There are whole algal genera that are common in the southern hemisphere but not found here. Common general ones are everywhere, but more specifically evolved ones are not as widespread. There are some very restricted environments and microorganisms that have not spread everywhere.
  - Even if a population of algae is found everywhere, does having a large number of a species in one area change the risk? If there are hundreds of thousands of times more of those algae in ponds than in the surrounding ecosystem, does that affect its ability to survive and affect the ecosystem? One participant argued that it may affect dispersal events, but that if the organism is local, if it could be in that environment then it would be.
  - Dispersal events are important. Invasion rate matters regarding turnover of species populations. That problem is occurring now: As climate change occurs, it is opening new environments to different organisms. It is important to think about where the evolutionary constraints are.
- Florida has adopted an Invasiveness Index, first developed by Australia. Is this for consideration of synthetic organisms?
- An algae industry participant noted that they were using advanced technologies to sample and study organisms that cannot be cultured, and are culturing organisms better; this is where further development of tools and instrumentation would be important to understand existing biodiversity.
- Some participants compared synthetic biology to historical domestication of crops and argued that domestication is not the process of making organisms more fit for the environment. Others argued that there are many examples of organisms becoming more environmentally fit and invasive due to domestication, so an analogy with domestication does not suggest that one should not worry about domestic algae.

- These questions will look different to the environmental community.
- Need to consider fitness, genetic stability, and gene transfer.

Additional location considerations

- Participants mentioned dispersed production facilities, which people put where they wish.
  - The distributed approach includes only environment types already discussed for centralized facilities (cropland, urban, deserts, freshwater, saltwater).
  - Distributed systems still have to be big and near adequate infrastructure. They will have oil to be refined, which will need to be piped/trucked, so such facilities will still be located near refineries, with the ability to move oil/fuel to where it can be refined/distributed. They will need to be near existing infrastructure and labor pools to avoid creating a shantytown in the desert.
- The discussion has dealt with variables one at a time (siting, etc.), but the factors interact; for example, an organism's intended use may affect the ideal location.
  - For example, is making lipids dependent on availability of customers? Should they be grown in urban environments to use available CO_2? How should we separate the variables?
  - Techno economic approach to decisions. Social, labor impacts. It comes down to techno economic. It is all economics, particularly in energy, which is a low-margin commodity.
- A participant argued that the organism should be contained. If it is not contained, the environmental parties would oppose it because of concerns about its getting out, unless the organism was shown to be safe through better testing that was public and not deemed Confidential Business Information (CBI).
  - A participant argued that there is a need for a better tool than TSCA for regulating these things.
  - A participant stated that testing is not CBI. All testing must be publicly disclosed. EPA has a mechanism for looking at that and making it open under Freedom of Information Act (FOIA) requests.
  - Some participants stated that all participants could agree that they were looking for a set of data that shows release is safe.
- Facilities are now located in Hawaii, China, and other places. They are in places participants have noted as not being ideal, but some are in the "best" locations as well.
- What if more people were doing this? There was a problem with chemicals because of synergistic effects. No work was done on interaction effects. What if there were lots of GM or synthetic biology organisms out there? How would that affect things and how would it be regulated? Other than foods, assume they are making chemicals or other things that are in EPA's regulatory space.
  - Some argued that such a situation exists for bacteria now, and that an analogy between bacteria and algae is valid because algae from different facilities do not go and mate/cross with each other.

    - Some argued that the existing bacteria are in reactors. The potential for interaction there are much less than when everyone is making their own "boutique bugs." What if there were many Bioprocess algae -type locations, in which one bug was making one product, and another was making another? Would it make a difference? It could be organized as an eco-industrial park, but it could just be co-located.
    - Some argued that industry is not going to be making new molecules that have never been seen before, so that type of combinatorial interaction would not take place.
- A business plan could be to take CO_2 from a highly polluting ethanol facility to make fuels, not using GM plants. Cannot answer the question on what that would mean for possible synergistic effects. A participant stated that what industry is coming up with now is the most efficient way of making an existing industrial product (for now).
- One could make a species to take advantage of a niche, but more likely as the niche changes, new strains/characteristics develop, e.g. the tar sands produced a new environment, organisms have evolved to take advantage of it. There are many considerations in siting that have nothing to do with these considerations. The cases of dispersal and of evolution to take advantage of a niche are difficult because there is still a lack of understanding of what those endpoints are and how they can be addressed.
- It is important to be sensitive to variations in the release environment.
- Siting issues become connected with facility design issues. EPA does consider design of facility in approval.

Reductions in organism fitness

Assuming the "ideal organism" traits discussed in Session II, participants discussed claims of decreased fitness. What measures or tests would they look for to identify reliable evidence of reduced fitness?

- Experimental data should be obtained to determine how an organism survives in the environment and how it competes against local organisms. For example, one might put the organism into ponds and see how it survives, or put algae into samples from nearby water and see how they do. It is important to include markers to be able to track the organism.
- One may also create microcosms and test fitness. These are not expensive, and are doable.
- Genetic stability: What if the organism loses a trait that had lessened its fitness?
    - Dependent on method. In vitro is point mutation, so it will drift immediately back when the selection pressure is relieved. If one does a stable transfer, then the genes are more stable. So methods affect stability. Process becomes important.
    - Reversion data exist, but are just never seen as worth publishing?

- With drug resistance experiments, as long as the selective pressure remains, only the ones that survive continue. GFP (green fluorescent protein) has existed forever.
- It would be valuable to test the degree to which traits shed. For example, tests conducted with plants and stack traits, but used to test one-by-one. Tests are only as valuable as the settings acknowledged. How to design tests for this? To check stability of attributes? How to test for hazard and likelihood?

- Bloom algae are rapid reproducers, strong against grazers; they accumulate, and then form harmful algal blooms.
- A participant suggested that lessons may be learned from earlier soil microbiology: When rhizobia were first being introduced, everyone said that they would die out rapidly in the soil. Later, someone planted sensitive legumes in the area where the rhizobia could no longer be found, and found them a decade after they had disappeared. Similar studies were conducted in Oregon and there too, it was found that organisms can be undetectable until the perfect conditions arise.
  - For algae, do the organisms die or do they just become undetectable?
  - A participant stated that algae do form spores. Large proportions of algae are rare and become abundant only once in a while.
  - How would the knowledge gained about rhizobia have changed the risk assessment decision in that case? A participant argued that the situation with rhizobia had been anticipated, but that the question at the time was what the hazard consequences are, and stated that risk assessments for algae would be approached in the same way.

Considerations of horizontal gene transfer (HGT)

- To put this in context, the question has been do they survive or not, but the real question needs to include a time variable. How many generations? 50? 100? Some participants argued that it should be assumed that the organism will survive long enough to transfer. Much more is known for cyanobacteria, but data are still needed for eukaryotic gene transfer.
- There has been some indication of rates and environment, but it would be good to know information specific to eukaryotic algae. How do they work, and are there appropriate recipients out there? This is an information gap that requires more data.
  - Dick Sayre's paper suggests there may be more transfer than expected, but there has been much less sequencing. Data are expected to come out within the next year or so; the extent of transfer has not been seen yet, just because not enough sequencing has been done yet.
  - One company showed no HGT against local organisms. Is that useful? Should companies be conducting such studies?
  - One reference is looking at gene transfer between alga and viruses. Is this really rare, or does it require further study?
- What is the significance of a high probability of naturally occurring transfer? How much has changed, how unique is the gene? Not all genes require further study;

how should it be determined which do? Transfer is not always hazardous, so the consequences of transfer need to be known. How much transfer is too much? Some participants argued that the limit cannot be kept to zero. Some organisms will take anything (e.g., rotifers take on everything they eat), some will not. There is a lack of sufficient data.

Who should be performing the needed studies and how should data be made available?

- Models for who funds and who performs studies:
  - Some participants argued that it is the role of government to perform these basic science studies. There are many experts in these areas, so the government should spend some of its basic research dollars on this. Private companies doing this research are always seen in a tainted light. The companies want to know what tests to do and they do not want to wait.
  - Petroleum Environmental Research Forum (PERF) model: a number of petroleum companies have the same questions, so they pool money, DOE matches (-ish), then national labs do the research and publish the answers. Some proprietary data are kept confidential, but at least the answers become available, and everyone benefits.
  - National Institutes of Health (NIH) model: "we fund, you publish, we post findings and data." Some participants stated that this is the type of model they would like to see.
- Many questions are being explored, but by proprietary entities. Who owns the data and how public are they? If they are not presently available, how can they be made public?
  - A participant FOIA'd DOE, and some labs gave everything while others kept nearly everything confidential. The participant argued that the agency needs a clear policy. The Department of Defense, on the other hand, gave lots of information (the Defense Advanced Research Projects Agency (DARPA), however, is entirely secret).
  - CBI and protecting firms, vs. protecting the public interest; is this a "collective interest free-rider problem"? Without scrutiny, is trust lost?
  - A biofuels industry member asked other participants what information they would like to see. Participants responded that they would like to see data from studies, and to know where variables were modified and how, etc.
  - In pharmaceuticals, the European Union has shifted to public dossiers.
  - Floating around Congress are revisions to TSCA that put more of a burden on the submitter for why their data should be confidential. How that will play out is unknown.

Summary of Session VI: Methodology and Protocols

Participants were asked to discuss what kind of advances in instrumentation and measurement methods might be required to evaluate the potential or actual environmental impact of GM algae. The discussion focused on four areas: mesocosm experiments, ways of measuring genetic stability and gene transfer, the need for data on a "base set" of organisms, and the need for modeling, validation, and reproduction of all experiments. Many participants repeatedly emphasized the need for cooperation, data sharing, and replication/parallelization of experiments between industrial, government, and academic labs.

1. **Technical Aspects of Mesocosm Experiments**
   The first question is the choice of which environments to simulate. Sterile water is not a representative environment, although it could be used in a control experiment. Participants emphasized the need to choose a representative sample of the environments that an escaping organism is most likely to encounter near the site of cultivation. Although earlier discussions of land-use and siting focused on polluted and degraded land, one participant described a practice of choosing to sample healthy local environments, rather than those already perturbed by human activity.

   Self-contained mesocosms must be technically sophisticated enough to accurately simulate the natural environment. They must include diurnal variation of conditions, water flow, and replenishment of resources, and should not allow unnatural chemical buildup. Any mesocosm experiment should be run long enough to be meaningful (possibly as long as multiple years). The appropriate duration of an experiment will depend on the effect being studied.

   There is great desire for validated, standardized mesocosms; one participant said, "I'll buy a dozen of those reactors if it'll help me standardize." There is an existing field of mesocosm studies, but its methods and apparatuses may need to be modified to fit the needs of algae researchers. ORNL has developed flow-through systems that are germane to mesocosm development. Existing and proposed test bed facilities should be built with an eye to scaling up from test-tube to microcosm to mesocosm experiments. JBEI reportedly has a full range of production models (labs, greenhouses, open ponds).

2. **Measuring Genetic Stability and Gene Transfer**
   Several participants stated that as the study of horizontal gene transfer (HGT) has advanced, researchers have found that far more HGT is taking place than was previously thought. This is due partly, but not wholly, to improvements in sequencing and metagenomics techniques. There are good data on transfer between bacteria, but not for eukaryotes. It is now known that gene transfer can

cross kingdoms. Thus, there is a great need to narrow the search space. It is infeasible to do pairwise HGT tests of all organisms in the environment. With current methods, researchers can establish whether or not HGT took place, but there is currently no way to predict whether one organism will transfer genes to another. Given the large amount of HGT already occurring in nature, experimenters must also consider exactly which genes are being transferred. If a particular HGT event would have occurred in nature even without human intervention, then that event may not be of concern if it occurs due to a GMO. Presumably, novel/engineered genes are of greater concern than naturally occurring ones, but not all novel genes are equally problematic.

The stability of genes introduced into an organism is often tested by growing the organism for several generations without selection pressure for those genes, and seeing if the genes are retained. The stability of an introduced gene depends strongly on the method by which it was introduced. Plasmids are lost relatively rapidly, while genes inserted by chromosomal integration are far more stable; one participant cited stability of up to "decades in the lab." Chromosomally integrated genes are not perfectly stable forever, but it would be difficult to detect reversion events due to their rarity. The current EPA regulatory approach does take into account the method of insertion of genes, and treats plasmids differently than genes integrated into the chromosome.

Several participants stated that the group attending the current workshop did not have all the necessary expertise to have a comprehensive discussion on the topic of genetic stability. They suggested holding another workshop with evolutionary biologists, specifically to address this question, and named some potential attendees.

3. **A "Base Set" of Organisms**

The published literature on algae is very thin. Most potential commercial biofuel organisms have not been thoroughly studied in a way that addresses the concerns brought up at this workshop. A few food-relevant algae, such as Spirulina, have been so studied, but they may not be suitable analogues for biofuel-producing algae. One participant called attention to ORNL's extensive study of vascular plants for bioenergy applications in which the group examined hundreds of species in cooperation with USDA and others. The participant suggested that the algae community could do similarly wide-ranging studies.

Many participants agreed that it would be useful to conduct comprehensive studies on a "base set" of algae species, evaluating their safety in specific environments. The establishment of this "base set" could go hand-in-hand with the earlier suggestion of establishing a "bad bug list" of organisms to avoid using. Members of the base set could be considered as analogues when evaluating a novel GMO similar to one in the base set. The eventual goal would be to make the leap from specific conclusions like "organism X causes harmful outcome Y in

environment Z," to general statements like "engineered feature A will likely present problems, and feature B will not."

One participant proposed a basic experiment: take a wild-type strain, modify it to include green fluorescent protein (GFP) or to be traceable in some other way, and release it into the environment. (This strain would be relatively safe to release, but would still require a TERA.) Experiments like this one would be part of the "base set" species evaluations, but this particular experiment would only be a starting point.

4. **Experimental Design / Need for Modeling, Validation, and Reproducibility**
Measurement criteria should be defined ahead of time, and include consideration of factors like how long the experiment must be run. In every case, the experimenters must decide: what precisely is the effect of interest, and how long must they wait for it to occur? For example, the concept of evolutionary fitness includes much more than "yes, the organism survived" or "no, it did not survive." It also includes the relative growth/success of different species, and their prowess at nutrient utilization.

Mesocosm experiments need to be extensively replicated, and parallelized between different laboratories – academic, industrial, and governmental. This point was made repeatedly by multiple participants, one of whom told an anecdote about losing an entire summer's worth of data from open mesocosms in Lake Erie because a great blue heron defecated in one replicate and not in another.

Once mathematical models have reached an adequate level of sophistication, they could help ease the burden of replication. Models and replicable analyses would also help experimenters plan their measurements, controls, and duration of testing. Besides the inherent usefulness of models, as one participant noted, the algae community needs to reach the level of sophistication needed to produce relevant models in order for anyone to have confidence in the results of their experiments, whether virtual or physical.

Both models and physical experiments must always be validated against field observations. Although comparison between a mesocosm and a natural stream is the first step of validation, the community must decide on specific validation criteria beyond "it seems similar to nature." One proposed experiment was to see if natural environments and self-contained mesocosms react similarly to the introduction of a "somewhat exotic," but non-GM, organism. Following this experiment, a GMO could be introduced to the mesocosm, and its effects extrapolated to nature.

Participants also mentioned the possibility of genetic manipulation having unintended side effects. Metabolic networks are highly complex and redundant, and a modification to one gene may change the behavior of many others, and/or have biotrophic effects. In general, EPA expects applicants to know what changes they have made in a GMO

relative to the parent organism, but these side effects present a difficulty. Applicants might address this difficulty with a comprehensive set of gene microarrays, protein, RNA, and metabolite measurements. TSCA requires applicants to submit all data relevant to health or environmental consequences, and does not set a standard list of testing procedures. This allows EPA to be flexible, in case an applicant presents a highly unusual organism.

Finally, one participant asked whether the current discussion was taking place within an appropriate conceptual framework. There was some agreement that a good framework is needed for any such discussion in order distinguish which questions are most important and why. Several potentially applicable frameworks already exist:

- The TSCA statute and regulations themselves constitute a framework, but it may not be the most useful one and should certainly not be the only one used.
- The EPA is using a comprehensive environmental assessment framework for the evaluation of nanomaterials, and a prior Wilson Center workshop on synthetic biology used this same framework to guide the conversation on possible hazards from cyanobacteria.
- There also exist tools specific to the field of risk assessment, such as fault trees, which could provide guidance.

Summary of Session VII: Wrap-Up

To close the workshop, participants were asked: "If you had $100 million to spend on research to address the questions raised today, how would you allocate it?" Answers focused on three areas: meta and organizational efforts; basic algal studies; and specific experiments to do or technologies to develop.

Meta / Organizational Goals

- Carefully planning the studies to be done, in accordance with an agreed-upon set of priorities.
- Rigorously defining concepts like "fitness"—which has so far been used in a vague way—in terms of measurements, baseline conditions, environmental contexts, safety targets, and best management practices to reach those targets.
- Performing literature reviews, and examining the work of other countries on other types of GMOs such as fish.
- Examining algae-farming endeavors from the perspective of social/economic sustainability, as well as environmental sustainability.
- Supporting education and outreach to ensure that the algal research community serves the public as well as the expert audience.
- Wrapping up results from across the range of future studies into a coherent body of knowledge.

Basic Algal Studies

- Studying general algal biology to increase the community's knowledge and tool-set, ultimately endeavoring to approach the level of characterization currently available for other industrially useful organisms.
- Establishing environmental reference data on natural algal communities so that in the future it will be apparent if a change has taken place. This could include taking baseline data on natural algae, their effect on the environment, and their natural lipid production. One participant considered whether baseline environmental monitoring should be a condition of DOE awards or other grants for biofuels development.
- Establishing a "base set" of useful organisms—and accompanying comprehensive characterizations—as well as a list of organisms to avoid using.
- Studying what role the "base set" organisms play in natural microbial communities in order to know what one might expect to see, or what one should plan to measure, in a mesocosm experiment.
- Studying the ability of harmful algal blooms (HABs) to produce neurotoxins, and the ability of engineered organisms to do the same; sequencing genomes to look for toxin-producing or allergenic gene products.

Specific Experiments or Developments

- Developing replicable and realistic mesocosm apparatuses and protocols for use.
- Modeling relevant phenomena that are already well understood, such as the airflow over a typical open-pond facility.
- Taking an engineered organism, knocking out the inserted genes one at a time,

and measuring how well it competes in a natural population to simulate the effect of genetic instability.

- Studying GMOs in the context of "non-natural" environments, such as depleted farmland, and the combined effect of farming/depletion and algae culture over time.
- Expanding previous work on probabilities of various adverse outcomes to incorporate the full range of spatial and temporal scales. (See: Martin Alexander, "Ecological consequences: reducing the uncertainties". Issues in Science and Technology 1:57-67 (1985).)