



LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

Genetic Engineering Workshop Report, 2010

J. Allen, T. Slezak

November 24, 2010

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Genetic Engineering Workshop Report, 2010

Thomas R. Slezak and Jonathan Allen
Lawrence Livermore National Laboratory



This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.



Genetic Engineering Workshop Report, 2010

Jonathan Allen and Tom Slezak, LLNL

Introduction

The Lawrence Livermore National Laboratory (LLNL) Bioinformatics group has recently taken on a role in DTRA's Transformation Medical Technologies (TMT) program. The high-level goal of TMT is to accelerate the development of broad-spectrum countermeasures. To achieve this goal, there is a need to assess the genetic engineering (GE) approaches, potential application as well as detection and mitigation strategies. LLNL was tasked to coordinate a workshop to determine the scope of investments that DTRA should make to stay current with the rapid advances in genetic engineering technologies, so that accidental or malicious uses of GE technologies could be adequately detected and characterized. Attachment A is an earlier report produced by LLNL for TMT that provides some relevant background on Genetic Engineering detection.

A workshop was held on September 23-24, 2010 in Springfield, Virginia. It was attended by a total of 55 people (see Attachment B). Twenty four (44%) of the attendees were academic researchers involved in GE or bioinformatics technology, 6 (11%) were from DTRA or the TMT program management, 7 (13%) were current TMT performers (including Jonathan Allen and Tom Slezak of LLNL who hosted the workshop), 11 (20%) were from other Federal agencies, and 7 (13%) were from industries that are involved in genetic engineering. Several attendees could be placed in multiple categories. There were 26 attendees (47%) who were from out of the DC area and received travel assistance through Invitational Travel Orders (ITOs). We note that this workshop could not have been as successful without the ability to invite experts from outside of the Beltway region.

This workshop was an unclassified discussion of the science behind current genetic engineering capabilities. US citizenship was not required for attendance. While this may have limited some discussions concerning risk, we felt that it was more important for this first workshop to focus on the scientific state of the art. We also consciously chose to not dwell on matters of policy (for example, screening of commercial gene or oligo synthesis orders), as multiple other forums for policy discussion have taken place in recent years. We acknowledge that other workshops on

topics relevant to genetic engineering should be held, some of which may need to take place at higher classification levels.

The workshop moderators would like to acknowledge the enthusiastic participation of the attendees in the discussions. Special thanks are given to Sofi Ibrahim, for his extensive assistance on helping this report reach its final form.

Executive Summary

The genetic engineering workshop brought together a diverse mix of genetic engineering pioneers and experts, Federal agency representatives concerned with abuses of genetic engineering, TMT performers, bioinformatics experts, and representatives from industry involved with large-scale genetic engineering and synthetic biology. Several talks established the current range of genetic engineering capabilities and the relative difficulties of identifying and characterizing the results of their use. Extensive discussions established a number of recommendations to DTRA of how to direct future research investments so that any mis-use of genetic engineering techniques can be promptly identified and characterized.

List of Recommendations to DTRA:

- 1. DTRA should establish, fund and maintain a multi-disciplinary team to track and document advances in genetic engineering technologies. They should engage in the iGEM and DIY-bio communities, to foster awareness of ethical issues and encourage public education in the use of techniques that maximize biosafety. This could include an “iGEM for PIs” (instead of students) or other types of focused design challenges. Such challenges should have real money at stake, to foster interest from capable teams. The DTRA-funded genetic engineering team should also be tasked to develop software to identify and characterize the different methods of genetic engineering as they emerge.**
- 2. DTRA should work with researchers and industry to build a comprehensive library and database of fully-sequenced engineered organisms and establish databases focusing on the needs of genetic engineering and its detection. (Example: mutation/phenotype database, pathway alteration database, centralized GE “parts” database, etc.) Expert curation on an ongoing basis is a necessary part of these databases.**
- 3. DTRA needs to continue funding basic research on the biology of microbes with respect to genetic engineering (e.g., virulence, drug resistance, host immune modulation, the horizontal transfer of the corresponding genes, etc.) and applied research for detecting and characterizing the use of genetic engineering techniques.**
- 4. DTRA needs to fund bioinformatics tools for annotation and analysis that scale to the range of 1,000s of genomes. Sponsoring better Bio-CAD tools and “cloud” computing environments using these tools could also be a way to channel the bio-engineering communities towards safer experimental designs and more efficient analysis of engineered organisms.**

5. **Genetic engineering of viruses poses additional challenges for detection and characterization that requires focused research attention.**
6. **Efforts in mitigation of genetic engineering derived threats should include focus on methods that apply directed evolution.**
7. **A follow-on workshop should be held to focus on what “the art of the possible” in genetic engineering techniques will likely be in 5, 10, and 15 years.**

Workshop Format

Most workshops of this nature attempt to cram too many talks into too little time, with little if any opportunities for meaningful discussion. To avoid this problem, we chose 4 talks to frame the context of the workshop in the first morning, and utilized the rest of the time that day and the 2nd morning for discussion. We kicked off the discussion by asking the Federal agency people present to provide us an informal summary of the perspective their agency has on genetic engineering issues. The four formal talks were:

The synthetic cell story and what would Bin Laden Genomics Inc. learn from it

**John Glass:
J. Craig Venter
Institute**

Genetic manipulation of Salmonella, Yersinia and cyanobacteria to develop vaccines against infectious diseases, fertility and cancer and produce biofuels

**Roy Curtiss:
Arizona State
University**

Implications of Engineered Biological Chassis on Safety and Security

**Harris Wang:
Harvard University**

Challenges in genomic based approaches for GE detection

**Jonathan Allen:
LLNL**

Prior to the workshop, attendees were sent a short paper (see Attachment C) that framed 3 questions that the workshop discussions would focus on:

1. **What are the specific goals that research efforts should address and that funding agencies can use to support their customers?**
2. **What are the genetic engineering applications, and can they be precisely defined to prioritize mitigation efforts and ensure high priority threats are addressed?**

3. **What are the genetic engineering methods, their enabling technologies, and how do they affect the difficulty of the threat mitigation problem?**

Discussion of Recommendations

The workshop recommendations are derived from the over 5 hours of discussions at the workshop and notes taken by several attendees. A draft of this report has been circulated to all attendees to provide an opportunity for their editing suggestions.

1. DTRA should establish, fund and maintain a multi-disciplinary team to track and document advances in genetic engineering technologies. They should engage in the iGEM and DIY-bio communities, to foster awareness of ethical issues and encourage public education in the use of techniques that maximize biosafety. This could include an “iGEM for PIs” (instead of students) or other types of focused design challenges. Such challenges should have real money at stake, to foster interest from capable teams. The DTRA-funded genetic engineering team should also be tasked to develop software to identify and characterize the different methods of genetic engineering as they emerge.

Rapid developments in genetic engineering technologies mean that efforts to ensure the capability of detecting abuses are always in catch-up mode. Investing in an ongoing effort to track and document new advances in genetic engineering will provide DTRA-funded bioinformatics detection/characterization efforts with specialized expert research. As much of the genetic engineering new technology information comes from methods sections of publications about applications of genetic engineering, considerable expertise is required to keep current. An initial startup effort could include documenting existing genetic engineering tools and techniques. Longer term, an ongoing effort is needed to develop software that is capable of detecting the potential usage of these tools and techniques, by analyzing the altered genomes after they have been sequenced.

It became evident in the workshop’s discussions that many of the genetic engineering experts present had large amounts of specialized knowledge that is not readily accessible. Additionally, they expressed a strong interest in assisting DTRA’s efforts to better understand topics such as mutations affecting virulence, resistance, and host immune system alteration. Some have deep experience with engineering the organisms that might be the most likely targets for deliberate abuse of genetic engineering. DTRA should seek to establish ongoing consulting relationships with key SMEs to be able to leverage their knowledge and experience as needed. An example could be specialized workshops (or consultations) on specific organisms, GE design objectives and/or techniques to help analyze and interpret the genomes of unusual samples and provide recommendations to support development of detection and countermeasure methods.

Workshop attendees with iGEM experience noted that sponsors do not tell student teams what type of research challenges to undertake. They noted that the FBI has recently

engaged successfully with this community. Much could still be accomplished to aid the “bio-hacker” communities in terms of increasing awareness of ethical and safety issues. One suggestion was that DTRA could sponsor an “iGEM for PIs” contest, where perhaps the focus of the competition might be more readily set by the sponsor. It was also noted that the CAGEN (Critical Assessment of Genetically Engineered Networks) competition is also relevant for DTRA to be aware of and potentially involved with. (See: <http://openwetware.org/wiki/CAGEN>)

2. DTRA should work with researchers and industry to build a comprehensive library of fully-sequenced engineered organisms and establish databases focusing on the needs of genetic engineering and its detection.

Relatively few researchers fully sequence organisms following successful genetic engineering attempts. This leaves bioinformatics researchers tasked with detecting nefarious genetic engineering with scant real-world information to guide their algorithm development. Discussions at the workshop indicated that many researchers would be quite willing to let DTRA sequence some of their constructs for the purpose of furthering knowledge of GE detection. (Presumably, those with constructs with IP concerns would not want public disclosure of the sequence, but they might agree to a controlled-access agreement for specific DTRA needs.) Associated metadata should accompany construct sequence. (Purpose of construct, how made, biological effects, etc. as available.)

It was noted in the workshop discussion that much genetic engineering happens in industry. A suggestion was made that DTRA might consider outreach to the industrial microbiology society as one approach to enable future cooperation in the event of any genetic engineering adverse events that might have indications of industrial origins.

It was noted that there are no specialized databases relevant for the detection of genetic engineering (**Example: mutation/phenotype database, pathway alteration database, centralized GE “parts” database, etc.**). **Expert curation on an ongoing basis is a necessary part of these databases.** Workshop attendees discussed how mutations knowledge is not available in one database, and how a centralized GE “parts” database that includes sequence is not available. Other relevant knowledge, such as pathway alterations to affect phenotypes, simplifications of regulatory networks, etc. are scattered throughout published literature. This means that should the TMT’s sequencing centers encounter an engineered organism, the impact of certain sequence modifications are likely to remain a mystery, even though the biology has been described in a publication. Ongoing expert curation of these databases would be needed to keep them current with evolving knowledge. We note that many of these databases have other potential use for TMT, for countermeasure target design, etc.

3. DTRA needs to continue funding basic research on the biology of microbes with respect to genetic engineering (e.g., virulence, drug resistance, host immune

modulation, the horizontal transfer of the corresponding genes, etc.) and applied research for detecting and characterizing the use of genetic engineering techniques.

At multiple points in the workshop discussion it was noted that we do not yet have enough fundamental knowledge of microbial mechanisms such as virulence, resistance, host range selection, host immune modulation, etc. Similarly, we are only beginning to understand what the minimal requirements are for microbial life, or how microbes can utilize alternative pathways if some are disrupted. Most workshop attendees were likely unaware of the range of DTRA's current investments in these areas. It was noted that some of the knowledge that is required may be better obtained via some sort of a broader "Big Science" program than is possible with traditional single-investigator grants on single organisms or mechanisms. A corollary need mentioned in the workshop is that better systems for extracting and integrating knowledge from published scientific literature are desperately needed. This is not a DTRA-centric problem, but one that should be tackled by the wider sponsor community.

One conclusion from the discussion is that a focused program of applied research aimed at developing techniques to identify and characterize the use of known genetic engineering techniques should be supported by DTRA. This research is not likely to be supported by NIH/NIAID or NSF, primarily because it is not hypothesis-driven research. DOE is also less likely to support this work due to their focus on non-biosecurity related tasks. DTRA might be able to collaborate with DHS on genetic engineering detection and characterization, although DHS has yet to launch its own efforts in this area, which would likely focus on forensic applications.

4. DTRA needs to fund bioinformatics tools for annotation and analysis that scale to the range of 1,000s of genomes. Sponsoring better Bio-CAD tools and "cloud" computing environments using those tools could also be a way to channel the bio-engineering communities towards safer experimental designs and more efficient analysis of engineered organisms.

Recent dramatic advances in sequencing technology have completely outstripped the range of existing bioinformatics tools. DTRA is faced with the likelihood to need to be able to detect and characterize advanced genetic engineering techniques from metagenomic data in the near future. At the same time, we will likely have thousands of (mostly draft) genomes of strains of many of the organisms DTRA cares about. A focused program, coordinated with NIAID and other agencies, is needed to ensure that DTRA has the bioinformatics tools available to meet its mission needs.

In a related vein, it was noted that DTRA could consider sponsoring better Bio-CAD (Computer Aided Design) tools. This could help to focus the bio-engineering communities towards safer experimental designs. Some workshop attendees noted this could be done by ensuring that synthetic amino acids were needed for organism survival,

or that organisms designed could not live at human body temperature, or readily identifiable genetic signatures are included, etc.

It was also noted that “cloud” computing environments for biology researchers are needed, as explosive data growth coupled with network bandwidth, storage, and local computing constraints will preclude everybody possessing local copies and doing local processing. Making “safe” GE design tools available on a cloud environment could be one way to channel the wider GE community towards safer engineering practices.

Another scaling issue is that the explosion of sequencing and other data relevant to DTRA needs (including GE detection) are such that it soon may not be feasible for most DTRA researchers to have local copies of all the data they require for their research. Similarly, the up-scaled computational needs may exceed the resources available to many DTRA researchers. DTRA’s TMT program is currently establishing a central service at AFRL (Dayton, OH) that might be a nucleus for the much larger computational requirements of the near future. DTRA should begin planning now for the exploding computational needs that will become limiting factors in the very near future if not handled properly.

5. Genetic engineering of viruses poses additional challenges for detection and characterization.

Viruses were highlighted in the workshop as the most easily initiated threat. On the other hand they may be subjected to the least amount of alteration (other than just copying what has been found in nature) and as such their threat may present identical challenges to any emerging viral pathogen. Due to the high replication error rates of RNA viruses, deep sequencing of a viral quasi-species cloud might indicate numerous modifications that could appear suspicious, but might not result in viable constructs. Research on whether we could distinguish a cleverly engineered virus from a natural fortuitous random set of mutations does not appear to have been performed. Concerns were raised at the workshop that engineered viruses raised in cell culture should be identifiable from wild type. Efforts like the recent DTRA call to identify genetic markers of lab-passaged strains should be encouraged as ways to support GE detection.

6. Efforts in mitigation of genetic engineering derived threats should include focus on methods that apply directed evolution.

Several workshop attendees noted that all the methods of directed evolution should not be ignored in favor of recombinant DNA engineering. Indeed, as shown in Harris Wang’s talk, they have built a system to highly automate directed evolution, with random combinations of modifications in a small set of selected genes. We note that current DTRA calls are requesting proposals for new types of systems to more efficiently study viral evolution. As such systems emerge for legitimate viral research; they also raise the risk of potential misuse. It would be helpful if each DTRA call for advances in bio-

engineering technologies also required that developers adequately address the potential for miss-use and how to identify applications of their technique.

7. A follow-on workshop should be held to focus on what “the art of the possible” in genetic engineering techniques will likely be in 5, 10, and 15 years.

The focus of this year’s initial GE workshop was to understand the current state of the art of genetic engineering techniques and the challenges of detecting them. The next workshop should focus on what is likely to be possible in the short and mid-range future, when science fiction will suddenly become science reality. How will this further challenge the detection and characterization of miss-uses of GE technologies?

Attachment A: (Extracted from an earlier LLNL Report to TMT)

Background Information on Genetic Engineering

Genetic engineering is defined in this document as scientific alteration of the structure of genetic material in a living organism. It involves the production and use of recombinant DNA in order to produce a new trait in the engineered organism. Both bacteria and viruses can be engineered however the level of difficulty is usually considered higher for the engineering of viruses as their genomes are smaller than those of biothreat bacteria and therefore more easily disrupted by genetic manipulation. We note that genetic engineering, by our operational definition, seems distinct from “directed evolution” or other types of selective breeding that attempt to steer natural evolution through various types of evolutionary challenges. However, as shown in Harris Wang’s talk at the workshop, hybrid methods already exist, that permit directed evolution of a set of specific changes in a designated small set of genes. Thus, any distinctions between synthetic versus natural types of genetic engineering are likely to be arbitrary and increasingly irrelevant.

Methods of Genetic Engineering

Numerous techniques have been developed for genetic manipulation of bacteria and viruses. They fall into two categories:

- Extrachromosomal genetic engineering (EGE) involves the transfer of an autonomously replicating DNA (plasmid, cosmid, bacterial artificial chromosome) to the engineered organisms. This DNA can have a range of sizes of genetic material from 2 to 200 thousand base pairs.
- Intrachromosomal genetic engineering (IGE) involves the transfer or modification of chromosomal DNA from the organism through an in vivo or in vitro process.

A number of approaches can be used for IGE. The common mechanism for most approaches is recombination between the genome of the engineered organism and a designed external fragment of DNA. Most of the methods differ in the makeup of this external DNA fragment. It can be a plasmid, a bacterial artificial chromosome, a transposable element, a virus or viral fragment, or even a short sequence of synthetic DNA. Table 1 summarizes many of the techniques used for genetic engineering.

Genetic Change	Base pairs (bp) or Kilobase pairs (kbp) changed	Mechanism of Change	Reference
Plasmid or λ , Cre, FLP recombination	2-20kbp	Extrachromosomal or Recombination	Schweizer H. 2008 Horzempa J. 2009
Bacterial or Phage Artificial Chromosome BAC or PAC	20-200kbp	Extrachromosomal or Recombination	Ioannou P.A. 1994
Transposable element	20-60kbp	Recombination	Fu J. 2008
Engineered virus	2-60kbp	Recombination	Jacobs B.L. 2009
Multiple SNPS	20-100 base pairs	Recombination	Wang H.H. 2009
Single SNP	Single base pairs	Recombination	Heerman R. 2008 Datsenko K.A. 2000
Synthetic genome	2-?kbp	Synthesis	Gibson, D. G. 2008 Cello, J. 2002

Table 1. Techniques used for genetic engineering of microorganisms

Because recombination is generally an inefficient process, the external DNA fragment usually contains a selectable marker to allow differentiation of the genetically modified organism from the high background of unmodified organisms. In many cases this selectable marker can leave a unique non native nucleic acid sequence (such as a gene for antibiotic resistance) in the genome of the modified organism that is later detectable as a marker of genetic modification. Genetically modified food crops are identified thorough the detection of such markers. In some cases a second step may remove most of the marker gene but still leave genetic remnants of the alteration (Datsenko and Wanner, 2000). While not all genetic engineering leaves behind a non native sequence (Sun et al., 2008), genetic engineering must by its nature result in an altered genome and by some means this alteration from the natural antecedent must be detectable. The level of difficulty of detection however varies depending on the laboratory process used to create the mutation and the availability of the antecedent strain.

Genetic Engineering detection overview

A microorganism with an engineered genome could have a genetic alteration of thousands of base pairs or a single base pair and the difficulty of detection will depend on the amount of genetic difference between the engineered and non engineered organism. This relationship between size of genetic manipulation and ease of detection is summarized in Fig 1 below.

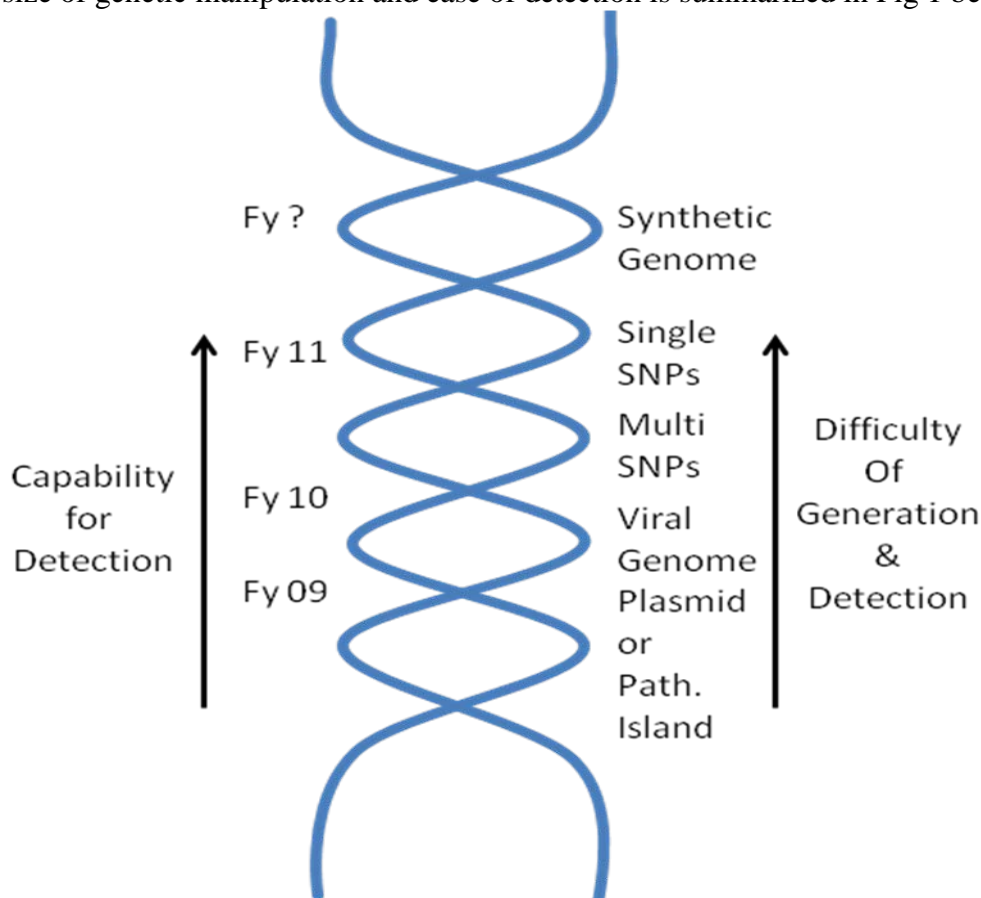


Figure 1. Hierarchy of genetic manipulation versus capability for detection

In addition to the impact of the size of the genetic change on the ability of detection, variation in the consensus genome sequence of the engineered organism can also impact the ability to detect genetic alteration. For example, when the complete genome sequences of many strains of *Bacillus anthracis* are compared there is very little genetic difference in the genome organization and sequence of these genomes. However, *Burkholderia pseudomallei* strains can have tens to hundreds of thousands of base pairs of sequence difference between individual isolates (Tuyanyok A et al., 2008). This wide latitude in core conserved sequence between different biothreat bacteria necessitates the assembly of a large database of sequence information for each of the bacterial agents. Only with a fairly comprehensive database of genomic sequence can one hope to discriminate a naturally occurring variation between individual isolates from an engineered change. Recommendations for additional genome sequencing for each of the

biothreat bacteria and hemorrhagic fever viruses has been covered in a previous TMT report on sequencing priorities submitted to DTRA in October of 2009.

Genetic Engineering detection

Work by the LLNL bioinformatics department under TMT and other funding has led us to define two distinct but complimentary approaches to detecting genetic engineering in microbes when using DNA extracted from a microbial sample as evidence. The first approach includes methods that look for direct evidence of genetic engineering. This applies to efforts to look for specific genetic markers associated with artificial genome manipulation. Examples of this include vector mediated engineering, and the various forms of site specific integration and homologous recombination where “scars” are left in the resulting genome from the vectors or primer sequences used to facilitate a genome manipulation event. The second category of genetic engineering detection approaches covers the methods that look for indirect evidence of genetic engineering by predicting that the presence of certain genetic sequences or sequence patterns are unlikely to occur from natural evolutionary processes.

Previous work detecting genetic engineering with microarrays

Progress has been made in detection of both forms of genetic engineering with the development of several microarrays. As an initial attempt to develop bioassays that find direct evidence for genetic engineering, a microarray was designed that hybridizes query samples to DNA probes specific for artificial plasmid vectors associated with genetic engineering in bacteria (Allen et al., 2008). Experiments have shown the ability to identify engineered *Bacillus subtilis* and *Bacillus thuringiensis* samples as well as to correctly not call natural *Bacillus* strains as being genetically engineered. A related effort demonstrated the potential to detect specific types of genetic engineering in plants using microarrays (Tengs et al., 2007). In this effort a tiling microarray was used to detect a small number of plasmid vectors used for genetic engineering in *Arabidopsis thaliana* and *Oryza sativa*. It may be the case, however, that the genetic engineering method leaves no evidence of an altered genome, other than the desired manipulation. For example, a site specific insertion event, where a gene is inserted but no signature sequence is left at the flanking sequence. In such cases, we are left to consider a second approach to detecting genetic engineering, which identifies the differences in a microbial sample that distinguish it from the collection of known microbial samples and determines whether the differences are due to natural evolution or human intervention. One approach, which has been undertaken, is to focus on the efficient detection of genetic differences in the microbial sample that are predicted to be of greatest interest for medical diagnostics and countermeasures. In particular, extensive testing of microarrays has been undertaken to detect genetic elements associated with virulence factors and antibiotic resistance (Jaing et al., 2008, Frye et al., 2006). For example, if a normally benign strain of *Bacillus subtilis* displays an unexpected phenotype, the query microbial sample could be interrogated with a microarray or sequenced to determine whether genes associated with virulence or antibiotic resistance are detected even when previously sequenced strains lack these genetic elements, and as a consequence these findings could infer a genetic engineering event.

Motivation for whole genome sequencing to detect genetic engineering

Of particular relevance to TMT is determining what additional information might be gained from sequencing a microbial isolate to identify potential genetic engineering. There are two relevant approaches to this type of sequencing, metagenomic and whole genome. Metagenomic sequencing involves directly sequencing the entire DNA isolated from a single sample.

Metagenomic sequence is considerably more difficult to analyze since only parts of the genome of interest may be captured along with partial genome sequences of other microbes.

Unfortunately, this may be the only form of sequencing possible if the microbe cannot be isolated and cultured for whole genome sequencing. This report focuses on the challenge of using whole genome sequence for genetic engineering detection with the understanding that additional future challenges must be addressed for dealing with metagenomic samples.

There are three challenges to using whole genome sequence for detecting genetic engineering:

1) identifying the genetic differences between the newly sequenced isolate and near neighbor reference strains, 2) functional characterization of the genome (in particular the differences with nearest neighbors), and 3) estimating the likelihood that genetic differences are due to a genetic engineering process.

Identifying the genetic differences themselves requires strong comparative genomics expertise, and the ability to build a comprehensive catalogue of genetic differences depends on how closely related the query sequence is to the reference genome and the degree of genomic variability within the family. For example, since *Bacillus anthracis* strains exhibit a more stable genome the likelihood increases that any new query *Bacillus anthracis* genome will map in its entirety to near neighbors. In contrast, it may be more difficult to map a novel *Francisella tularensis* sequence to its near neighbors, further increasing the number of reported genetic differences and thus making it more difficult to pinpoint the functionally important changes.

A sequence annotation pipeline applied to the genome sequence can give a functional characterization of the genome yielding results similar to microarrays but with the potential for greater precision. Thus, it may be possible to increase the sensitivity of detection of functionally relevant genetic elements, such as the presence of antibiotic resistance genes or other virulence markers that may be missed due to the limits on the number of virulence probes available for detection on the microarray. This stage of analysis could include comparing the newly sequenced genome to a database of genetic markers associated with known genetic engineering methods.

The first two steps to whole genome analysis (identification of genetic differences to near neighbors, and functional characterization of the differences) will be of principle importance for developing broad range diagnostics and countermeasures, irrespective of whether genetic engineering is a factor. Nevertheless, it is important to recognize that simply having access to the complete microbial genome and the first two steps of the analysis does not automatically give the ability to determine whether the differences are due to genetic engineering when known genetic engineering scars from a predetermined set are absent. Knowing whether changes are due to genetic engineering may in fact be of secondary importance, however, having the ability to make this determination could provide supplemental information in prioritizing which

differences with nearest neighbors should be analyzed more closely. For example, if the manipulation is the insertion of a motif or mutation, the genetic alteration may get lost in a sea of differences. Similarly, the significance of finding the presence or absence of certain genes may go unnoticed if their functional role remains unknown. Having additional methods to prioritize the analysis of a potentially large number of genetic differences could speed up the time it takes to identify and design effective countermeasures.

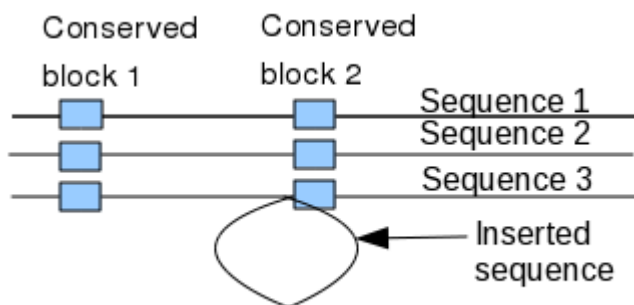


Figure 2. Schematic of a detected DNA sequence insertion event

Novel approaches to detecting genetic engineering with whole genome data

To begin to address these problems we have begun to explore a specific genetic engineering detection test case, to better understand the scope of the problem and determine if bioinformatics methods can translate TMT's sequencing capabilities into fast and accurate functional characterization of microbial genetic engineering events. To focus initial efforts, the question of site specific insertion and deletion events was considered. A simplified schematic is shown in Figure 2. The goal is to start with the (relatively) more simple detection problems, which can serve as a basis for developing additional detection approaches for other forms of genome manipulation such as gene swapping events, which more closely resemble naturally occurring lateral gene transfer events and nucleotide polymorphisms.

The broad question posed is, given the comparison of a newly sequenced isolate to nearest neighbor reference sequences; can genetic differences due to some process other than natural evolution be identified? To address this question, we first define a model for the natural process of evolution. In our initial examination, the definition of a “functional element” was left intentionally very broad to include detecting insertion or deletion of sequences that may contain promoters, genes, transcription terminators or parts of genes. There is likely no one model of evolution for the gain and loss of such sequence. It is important to note here that we are explicitly excluding in this definition mobile genetic elements. Two data sets were examined to generate preliminary results: 19 *Bacillus anthracis* genomes, and 12 *Yersinia pestis* and *Yersinia pseudotuberculosis* genomes and preliminary analysis on the *Yersinia* data set is given here.

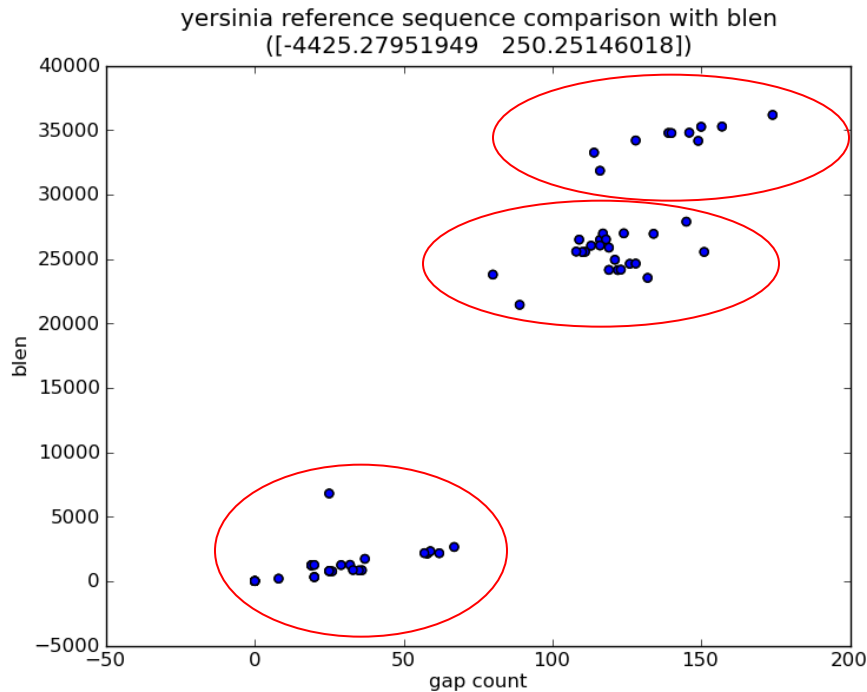


Figure 3. Pair wise nucleotide differences between *Yersinia* genomes (Y-axis) versus number of pairwise distinct insertion/deletion events

Figure 3 shows a comparison of pairwise genetic distances in the form of phylogenetic tree branch lengths using neighbor joining tree inference. The y-axis is marked “blen” for the branch length and the x-axis marked “gap count” is the number of observed insertion/deletion sequences (length > 30 nucleotides) between two *Yersinia* sequences. A linear fit of the data is used to estimate a gap insertion/deletion “rate” relative to the genetic distance observed between two sequences. The rate is estimated to be 1 insertion/deletion per 250 branch length units. However, it is clear from the figure that there are three clusters, with potentially distinct insertion/deletion rates. Figure 4 and Figure 5 show the breakdown of the two different clusters, which correspond to *Yersinia pestis* (Figure 4) and *Yersinia pseudotuberculosis* (Figure 5). The third cluster represents distances between the two subgroups. The *Yersinia pestis* rate is estimated to be higher with 1 insertion/deletion event per 37 branch length units. In contrast, the *Yersinia pseudotuberculosis* genomes appear to have a much lower site specific insertion/deletion rate at 1 per 274 branch length units. Looking at two of the more closely related *Yersinia pestis* strains, KIM and Nepal 516 as an example, predicts that we should see 5 distinct insertion/deletion gaps (the actual observed number is 8) given the number of nucleotide substitutions inferred between the two sequences. If accurate upper bounds on the rate parameter can be determined, this gives a powerful tool for determining the level of maximal sequence divergence for which a query strain can be compared to a reference strain and still easily identify an event attributable to genetic engineering. For example, if the maximal upper bound rate is a branch length 50 for each insertion/deletion event, the relative genetic distance between the KIM and Nepal 516, would expect to see 4 natural insertion/deletions between the two sequences. Therefore, engineering events that led to repeated insertions above 4 would stand out, but

gauging the significance of 4 or less insertion/deletions would require comparing the sequence to a closer near neighbor to determine whether natural evolutionary divergence is the cause of the differences. This information could prioritize, whether new *Yersinia pestis* strains are need be sequenced to fill in gaps that would facilitate genetic engineering detection. Although a lower insertion/deletion rate is estimated for *Yersinia pestis*, the minimal branch length between *Yersinia pseudotuberculosis* strains is 6794, implying that at least 25 insertion/deletions are observed among pairwise comparisons of the existing set of sequenced strains. Thus, this would indicate the need to consider filling in specific sequence gaps for *Yersinia pseudotuberculosis* in the tree.

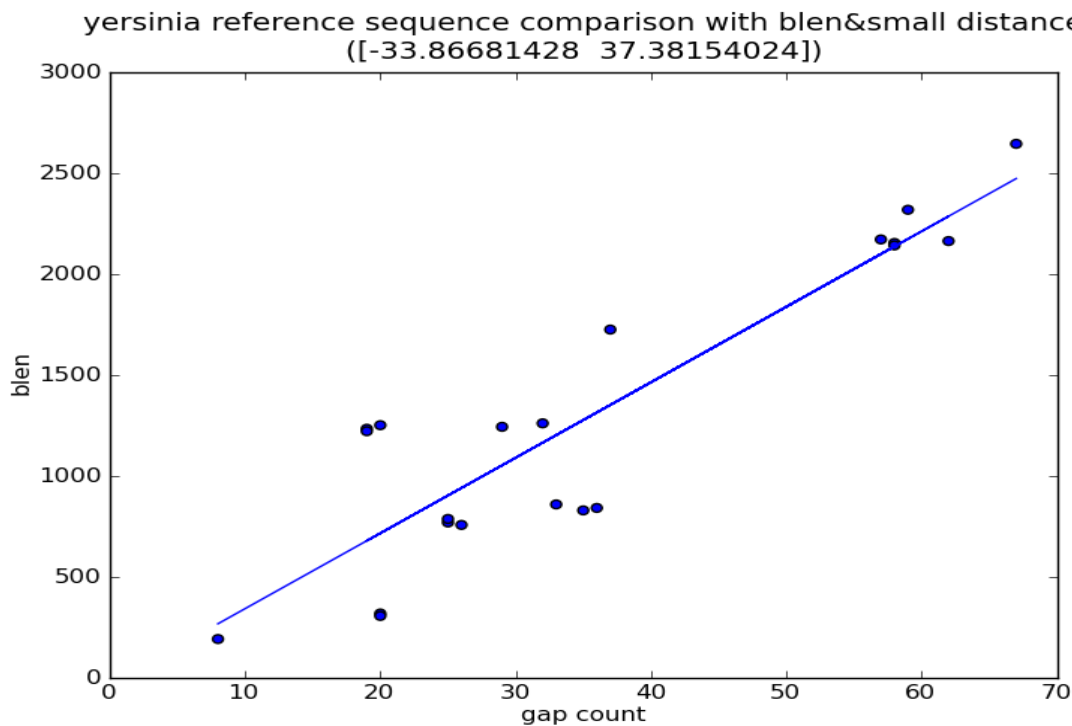


Figure 4. *Yersinia pestis* insertion/deletion rate

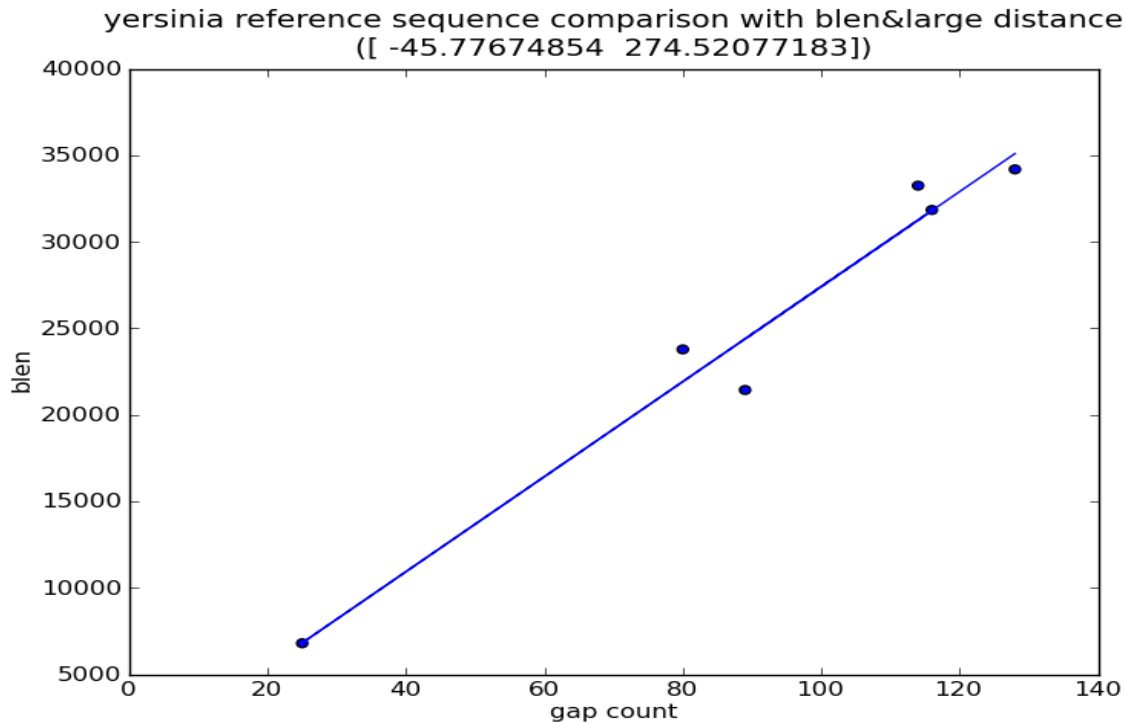


Figure 5. *Yersinia pseudotuberculosis* insertion/deletion rate

The results for the *Yersinia* data set presented here are only preliminary and require follow up verification. However, they illustrate the potential utility of applying bioinformatics analysis to support prioritizing sequencing of near neighbors for the purpose of rapid identification of genetic engineering. These preliminary results suggest the need may arise in the future for a larger sequencing effort to achieve the goal of a rapid, accurate and comprehensive detection system where more near neighbors are sequenced. Absent follow on sequencing to support improved detection, bioinformatics analysis will help quantify the confidence levels for asserting whether differences in a novel sample are attributable to genetic engineering or are the product of natural evolution.

The initial work reported here points to important areas for developing genetic engineering detection methods using whole genome comparisons which are listed below:

- Evaluation of more bacterial strains using existing sequence is needed. Initial work has started on *Bacillus anthracis*, but many more bacterial genomes should be analyzed. There may be additional difficulties encountered for bacterial genomes with a high level of genome plasticity.
- The accuracy of the insertion/deletion rate must be improved to ensure that a tight upper bound is achieved.
- Automated detection of different insertion/deletion rates within a bacterial group may be needed.

- Methods can be extended to consider sequence replacement events, rather than insertion/deletion events, which would be similar to other forms of lateral gene transfer.
- Better quantify the role natural mobile genetic elements (such as transposons) may play in genome diversification.
- Develop distinct rates for specific functional events. For example, distinguish gene gain/loss rates, from intergenic insertion/deletion and partial gene insertion/deletion.
- Experiment with different similarity cutoffs to decide what constitutes a non-homologous (laterally transferred or engineered) sequence.
- Evaluate the insertion/deletion patterns of multiple sequences simultaneously to more precisely determine, which insertion/deletion events are more likely to be due to genetic engineering when multiple candidates are available.
- Look at variation in observed nucleotide substitution rates across the genome to consider the potential to detect more subtle forms of engineering (i.e. point mutations). Evaluate the different roles draft genome sequence versus finished genome sequence can play in genetic engineering detection.

Other methods for detecting genetic engineering

The comparative genomics method discussed above provides a powerful tool for detecting various forms of laterally transferred genes, which may be naturally occurring or due to genetic engineering. The utility of this method will decrease, however, as the distance to the nearest neighbors increase beyond a certain threshold where the genome alignments become error prone and incomplete. More work is needed to better characterize what these distance thresholds are and the values will depend on the particular organism family in question. Our initial work with *Bacillus anthracis*, *Yersinia pseudotuberculosis* and *Yersinia pestis* shows the ability to generate alignments that cover most of the sequences, but the quality of the alignments may not hold for other organisms with greater levels of divergence.

A second approach to get around this problem is to consider a gene comparison, or more generally a genome segment comparison, where smaller sub regions are considered for their potential for lateral gene transfer independent of surrounding levels of conservation. This could give another method for targeting regions of interest absent a comprehensive functional characterization of the region and could provide evidence for genetic engineering. One approach is to build a consensus phylogenetic tree for a bacterial family of interest from a concatenation of orthologous genes, and then build individual gene trees to search for regions of the tree where disagreement in tree topology gives evidence for lateral gene transfer (Beiko et al. 2005). A detailed method for implementing an analysis pipeline is given in Beiko and Ragan, 2008. This is an approach that is closely related to methods for detecting genetic shift in the form of Phylogenetic networks (Huson et al. 2006).

One source (Ragan 2001) gives a good overview of different methods for detecting lateral gene transfer including the use of unexpected codon patterns. Codon usage patterns have been recently used to aid functional characterization of genes in term of their predicted expression levels (Bailly-Bechet et al 2006), and codon usage patterns could provide an additional source of evidence for detecting genetic engineering.

Future directions and needs of GE detection

A large number of GE techniques can be adapted from other microorganisms to biothreat bacteria and viruses. Detection of GE in a microorganism increases in difficulty as the size of the genetic change decreases. In addition to the size of the engineered change, the consensus genomic sequence of the microorganism can impact the difficulty of detecting an engineered change in genomes that are highly variable from strain to strain. This problem will require comprehensive databases of whole genome sequences for more genetically variable biothreat bacteria and viruses.

Given the myriad of approaches that can be used for genetic engineering and the widely varying forms of genome alteration (from single SNP to whole genome synthesis), it is suggested that at a minimum, a database should be developed and maintained that would track the various forms of genetic engineering along with the corresponding genomics based detection tools (where such tools exist). This will allow TMT to maintain a dynamic resource of risks and responses to potential genetic engineering that can be rapidly accessed as well as identify where the capability gaps are expected to persist when characterizing newly identified samples. As a first step LLNL has assembled a collection of the commercially available plasmid vectors and designed vector specific DNA signatures, which can be used on a microarray or *in silico* to screen bacterial genomes for evidence of genetic engineering. A logical next step would be to expand this database to include diagnostic genetic signatures for other forms of genetic engineering that leave behind specific alterations in the genome. The challenge will be to maintain a comprehensive survey of genetic engineering techniques in light of the expanding toolbox of genetic engineering techniques under development in laboratories around the world. To guard against missing unpublished methods or methods that leave no distinct genetic signature, it is worth including sequence analysis tools that look for genetic engineering methods by detecting (potentially unexpected) differences to near neighbor sequences. Here too, considerable work remains to determine what genomic consensus set of sequences are needed for each biothreat microorganism in order to precisely identify a change in the genome sequence as a product of genetic engineering. The comparative genomics tools that are developed and applied specifically to the genetic engineering detection task will help guide what sequence set of bacterial genomes should be used or developed, and will determine the limits of future rapid detection and characterization capabilities for novel genetically engineered biological threats.

References

- Allen JE, Gardner SN, and Slezak TR. DNA signatures for detecting genetic engineering in bacteria. *Genome Biol* 9:R56, 2008
- Bailly-Bechet M, Danchin A, Iqbal M, Marsili M, and Vergassola M. Codon usage domains over Bacterial Chromosomes. 2:4, 2006
- Beiko RG, Harlow TJ, Ragan MA. Highways of gene sharing in prokaryotes. *PNAS* 102(40), 14332-14337, 2005
- Beiko RG and Ragan MA. Detecting lateral genetic transfer a phylogenetic approach. *Bioinformatics*, Volume 1: Data, Sequence Analysis, and Evolution, vol 452, 2008
- Cello J, Paul AV, and Wimmer, E. Chemical Synthesis of Poliovirus cDNA: Generation of infectious Virus in the Absence of Natural Template. *Science* 297, 1016-1018, 2002
- Datsenko KA, and Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *PNAS* 97(12), 6640-6645, 2000
- Fu J, Wenzel SC, Perlova O, Wang J, Gross F, Tang Z, Yin Y, Stewart AF, Muller R, and Zhang Y. Efficient transfer of two large secondary metabolite pathway gene clusters into heterologous hosts by transposition. *Nucleic Acids Research* 36(17), e113, 2008
- Frye JG, Jesse T, Long F, Rondeau G, Porwollik S, McClelland M, Jackson CR, Englen M and Fedorka-Cray PJ. DNA microarray detection of antimicrobial resistance genes in diverse bacteria. *International Journal of Antimicrobial Agents* 27:2, 2006
- Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, Zaveri J, Stockwell TB, Brownley A, Thomas DW, Algire MA et., al. Complete Chemical Synthesis, Assembly, and Cloning of a *Mycoplasma genitalium* Genome. *Science* 319, 1215-1220, 2008
- Heermann R, Zeppenfeld T, and Jung K. Simple generation of site-directed point mutations in the *Escherichia coli* chromosome using Red/ET Recombination. *Microbial Cell Factories* 7(14), 2008
- Horzempa J, Shanks RMQ, Brown MJ, and Russo BC. Utilization of an unstable plasmid and the I-SceI endonuclease to generate routine markerless deletion mutants in *Francisella tularensis*. *Journal of Microbiological Methods* (in press), 2009
- Huson, DH and Bryant D. Application of Phylogenetic Networks in Evolutionary Studies. *Mol Bio and Evol*, 23:254-267, 2006
- Ioannou PA, Amemiya CT, Garnes J, Kroisel PM, Shizuya H, Chen C, Batzer MA, and de Jong PJ. A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nature* 6, 84-89, 1994
- Jacobs BL, Langland JO, Kibler KV, Denzler KL, White SD, Holechek SA, Wong S, Huynh T, and Baskin CR. Vaccinia virus vaccines: Past, present and future. *Antiviral Research* 84, 1-13, 2009

Jaing C, Gardner S, McLoughlin K, Mulakken N, Alegria-Hartman M, Banda P, Williams P, Gu P, Wagner M, Manohar C, and Slezak T. A functional gene array for detection of bacterial virulence elements. *PLoS ONE* 2008

Ragan MA. Detection of lateral gene transfer among microbial genomes. *Current Opinion in Genetics & Development* 11:620-626, 2001

Schweizer HP. Bacterial genetics: past achievements, present state of the field, and future challenges. *BioTechniques* 44, 633-641, 2008

Sun W, Wang S, and Curtiss III R. Highly efficient method for introducing successive multiple scarless gene deletions and markerless gene insertions into the *Yersinia pestis* chromosome. *Appl Environ Microbio*, 74(13), 4241-4245, 2008

Tengs T, Kristoffersen A, Berdal K, Thorstensen T, Butenko M, Nesvold H, and Holst-Jensen A. Microarray-based method for detection of unknown genetic modifications. *BMC Biotechnology*, 7(1):91, 2007

Tuyanyok A, Leadem BR, Auerbach RK, Beckstrom-Sternberg SM, Beckstrom-Sternberg JS, Mayo M, Wethiekanun V, Brettin TS, Neirman WC, Peacock SJ, et. al. Genomic islands from five strains of *Burkholderia pseudomallei*. *BMC Genomics*, 9:566, 2008

Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 460, 894-899, 2009

Attachment B: Workshop Participant List

(Note: One or more Federal agency attendees requested that their names not be listed in this report.)

Person	Institution	Email
Harris Wang	Harvard	hhwang@genetics.med.harvard.edu
Pam Silver	Harvard	pamela_silver@hms.harvard.edu
Adam Arkin	UCB	aparkin@lbl.gov
Chris Voigt	UCSF	cavoigt@picasso.ucsf.edu
Ryan T. Gill	U Col Bo	rtg@colorado.edu
J. Christopher Anderson	UCB	jcanderson@berkeley.edu
Jean Peccoud	VBI	peccoud@vt.edu
Karl E. Klose	UT SA	karl.klose@utsa.edu
Roy Curtiss III	ASU	rcurtiss@asu.edu
Herbert P. Schweizer	Col. State	herbert.schweizer@colostate.edu
Lynno Thomason	NCI Frederick	lynno.thomason@gmail.com
Jim Sawitzke	NCI Frederick	sawitzkej@mail.nih.gov
Barry Wanner	Purdue	blwanner@purdue.edu
Eva Top	U. Idaho	evatop@uidaho.edu
Kenan Murphy	Umass Med School	Kenan.Murphy@umassmed.edu
Dave Rasko	U MD SOM	drasko@som.umaryland.edu
Peter A Carr	MIT	carr@media.mit.edu
R P Shetty	Ginkgo BioWorks	rshetty@mit.edu
Sandy Gibbons	ECBC	henry.gibbons@us.army.mil
Nick Bergman	NBACC	bergmann@nbacc.net

Shanmuga Sozhamannan	NMRC	shanmuga.sozhamannan@med.navy.mil
Traci Pals	DHS	traci.pals@dhs.gov
John Mulligan	Blue Heron	mulligan@blueheronbio.com
Michele Garfinkle	JCVI	mgarfinkel@jcv.org
Mihai Pop	UMD	mpop@umiacs.umd.edu
Owen White	UMD	owhite@som.umaryland.edu
Alan Christian	Monsanto	allen.t.christian@monsanto.com
Yuriy Fofanov	U Houston	yfofanov@bioinfo.uh.edu
Jim LeDuc	UTMB/GNL	jwleduc@UTMB.EDU
Mike McIntosh	USDA/PIADC	Michael.T.McIntosh@aphis.usda.gov
Julie Fruetel	SNL	jfruet@sandia.gov
Frank Schaefer	EPA	Schaefer.Frank@epamail.epa.gov
Dan Drell	DOE	daniel.drell@science.doe.gov
Charles Cooke	USG	charllc@ucia.gov
Maria Giovanni	NIAID	mgiovanni@niaid.nih.gov
Dave Shepherd	DHS	dave.shepherd@dhs.gov
Kevin Anderson	DHS	Kevin.anderson@dhs.gov
John Dileo	MITRE	jdileo@mitre.org
Sofi Ibrahim	USAMRIID	sofi.ibrahim@us.army.mil
Greg Meyers	TAURI	gregory.meyers_anser@taurigroup.com
Tom Brettin	ORNL	brettints@ornl.gov
Randy Kincaid	DTRA	randall.kincaid@dtra.mil
John Glass	JCVI	jglass@jcv.org
Lynne Gilfillan	SRI	Lynne.Gilfillan@sri.com

Mark Gaillard	DTRA	Mark.Gaillard@dtra.mil
Dennis Imbro	TAURI	dennis.imbro@taurigroup.com
David Klaasse	TMT	david.klaasse@dtra.mil
Scott ReMine	TMT	scott.remine@dtra.mil
Paula Imbro	TAURI	paula.imbro@taurigroup.com
Rhea Brooking-Dixon	DHS	rhea.brooking-dixon@associates.dhs.gov
Scott Jackson	FDA	scott.jackson@me.com
Jonathan Allen	LLNL	allen99@llnl.gov
Tom Slezak	LLNL	slezak@llnl.gov
Jerry Blutman	DTRA	

Attachment C: Challenges of Genetic Engineering (focus document provided to participants prior to workshop)

Suggested topics for discussion at the genetic engineering workshop, September 23-24, 2010 Jonathan Allen, Tom Slezak, LLNL

The workshop is intended to focus on identifying the challenges presented when analyzing samples where genetically engineered microbes could be present. The samples could come from a clinical, environmental or lab isolate setting where the means of sample interrogation are likely to be genetic based (e.g. sequencing or some other variant of genetic testing). Challenges associated with other aspects of the problem, for example, preventing the proliferation of biological materials and technology will likely be better addressed in other forums.

The workshop will lead to a report based on the workshop discussions and will include recommendations for suggested follow on work. In writing a report recommending how DTRA and ideally other agencies might collaboratively fund and manage a research program to address the program of mitigating bio-security threats due to genetic engineering there is a need to consider the broad issues that should inform future funding decisions. Three broad questions are presented here to suggest ways to focus the workshop discussion:

- 1. What are the specific goals that research efforts should address and that funding agencies can use to support their customers?**
- 2. What are the genetic engineering applications, and can they be precisely defined to prioritize mitigation efforts and ensure high priority threats are addressed?**
- 3. What are the genetic engineering methods, their enabling technologies, and how do they affect the difficulty of the threat mitigation problem?**

These three questions are addressed in more detail in the next sections. The goal is to start by thinking broadly about the problems but hopefully drill down to begin identifying the key knowledge gaps that will drive progress in the field.

Question 1: What are the specific goals that research efforts should address and that funding agencies can use to support their customers?

Developing genomic based methods for detection and characterization of genetic engineering in microbes could require addressing broad technical challenges. The technical/knowledge gaps could provide a basis for defining what gaps the funding agencies need to fill to support bio-

threat mitigation. For example, the problem could be stated as solving a - Detection, Identification, Characterization and Eradication (DICE) ‘grand challenge’.

1. **D**etection of a genetically engineered microbe in a potentially complex sample. It is likely that a “pure” cultured isolate would not be available to work with. Potential challenges include:
 - Detecting the presence of genetic engineering. This could require the use of a signature. Thus, some *a priori* knowledge may be required that implies an upper bound on the degree of novelty in genetic engineering methods that can be detected.
 - Detection of the organism may challenge existing assumptions used for detection when the target genome sequence is subjected to manipulation. (e.g., signature avoidance)
 - The ability to efficiently process large numbers of samples may be needed, thus placing constraints on methods for sample interrogation.

What kind of data collection is possible, what are the minimal requirements? Methods likely would include: Microarrays, Metagenomic sequencing of complex samples, Multiplex PCR, Antibiotic resistance/sensitivity data, PFGE (pulsed field gel electrophoresis), MLVA (Multiple Loci VNTR Analysis), Immuno assays.

2. **I**dentification of genetically engineered features. This challenge is made distinct from the detection problem largely due to the presence of genetic engineering methods, which lack explicit genetic markers.
 - To what degree can we distinguish marker-less genetic engineering from the results of various natural variation processes?
 - Does identifying the important novel functional elements become more difficult with engineered organisms than with natural variation?
 - What role do reference genome sequence databases play in helping to infer genetic engineering (via comparative genomics), what data is currently missing, what is needed?

What are the data collection requirements? Presumably with less *a priori* knowledge about the genetic engineering method comes the need for more detailed genetic data in the form of draft genomes and ideally (but unlikely) finished whole genomes.

3. **C**haracterization of the functional impact of the engineered features.
 - When genomics data is used as input, this may heavily overlap with the general genotype to phenotype challenges.
 - What can we infer about functionality from raw metagenomic reads or microarray hits?

- To what degree could we predict potential functionality of an unexpected component, from assembled draft or finished sequence? Is there prior evidence that the identified virulence gene has restored biological activity when transferred to a non-natural organism?
4. **Eradication of infection** - This may lie outside the scope of most genomics based tools and is not expected to be the primary focus of the workshop. Nevertheless the challenge of developing the final tool that actually can be used to eliminate an ongoing disease/outbreak cannot be overlooked since this is one of the important end goals.
- Note that there is a somewhat related problem with the remediation of environmental releases of pathogens: most genomics based methods cannot distinguish live from dead pathogen DNA. RNA viability assays have been attempted, yet are problematic. The “how clean is clean” problem remains challenging.

Each of the four ‘DICE’ challenges has overlapping and distinct features that apply to two distinct tasks of bio-threat mitigation: attribution and countermeasures. Attribution refers to the ability to accurately determine that a microbial sample is indeed the product of human mediated engineering and ideally the process of the engineering methods can be unraveled and traced to their source. This type of threat mitigation assumes the power of deterrence to the extent that if unethical applications of genetic engineering can be determined and traced, this would serve to discourage their use by those who wish to remain anonymous. Countermeasures refer to the efforts to treat an outbreak once the microbial source has been released. Strictly speaking, these two tasks can be considered independent in that efforts to treat an outbreak or disease do not necessarily require determining whether the source was genetically engineered, and similarly determining that a microbe is genetically engineered does not necessarily require determining the purpose of the genetically engineered feature. Note that the primary motivations of those performing attribution and countermeasures are quite different (e.g., apprehending a perpetrator versus saving lives), which can lead to differing priorities for both operations and research. Of course, information gained from either effort is likely to complement the other and in some cases may be essential. For example, information about mutations in a virulence gene or toxin that give rise to the same structural motif may be useful to detect natural novel mutations or engineered genes.

The four ‘DICE’ challenges could most effectively be addressed in parallel, not sequentially. For example, identifying engineered features is necessary to develop detection capabilities. Also, characterization of the functional impact of the engineered features may point to other engineered features causing the same effect, and prioritize detection strategies.

Question 2: What are the genetic engineering applications, and can they be precisely defined to prioritize mitigation efforts and ensure high priority threats are addressed?

Given that the number of applications of genetic engineering will likely continue to grow it is worth considering the specific bio threat implementation objectives one might hope to achieve with bacterial or viral genetic engineering. It is our presumption that we have to focus on specific objectives that an adversary might hope to achieve – but it may be important to understand what might be missed and what needs to be prioritized by making simple lists such as the one here:

1. Robustness to countermeasures (e.g. antibiotic resistance, antiviral resistance)
2. Detection evasion
3. Optimize dispersal strategy. This could vary from introducing stimulus sensors used to trigger initial infection to ways to change nutrient requirements to increase robustness of an 'attack' colony, or to better survive packaging for more efficient delivery.
 - a. Alter host range
 - b. Alter host immune response
4. Spoofing – altering an organism's genome to give the appearance of 'engineering' without introducing a novel function, or introducing a non-functional modification.
5. Biochemical processing that ideally piggy backs on the technology applied for commercial use and has been converted to generate 'dangerous' chemical materials.

Question 3: What are the genetic engineering methods, their enabling technologies, and how do they affect the difficulty of the threat mitigation problem?

A bio threat implementation strategy that addresses genetic engineering will in part be determined by ease of use and accessibility of the enabling tools. For example, a long term goal of many synthetic biology efforts is to make genetic engineering more accessible to a wider population by providing standard molecular tools that can be used for genetic manipulation. This may enable wider access, but the broader set of users may rely on a central repository of 'molecular biology parts', which may better support detection and identification strategies and efforts. At another extreme end of the spectrum would be complete 'in house' *de novo* synthesis of a genome, which could preclude any potential for tracking from a central repository. There may be differing opinions on the timeline of wide accessibility to this technical approach. There are many other methods that appear to fall in between these two extremes. It could be important to enumerate over this area with some degree of specificity as to if, and when, different methods may be of concern. It could be useful to understand how the field may evolve over time (1, 5, 10, and 30, years) to help determine how to balance funding existing threat mitigation work versus anticipated future threats.

Given that there is little opportunity for public (or private) research projects to directly investigate and publish on developing biological threats to humans, it is likely important to focus on how the dual-use elements will likely be the primary drivers of the threat potential. A closely related factor will be the expanding knowledge base of microbiology, which should continue to allow for implementation of new applications using existing genetic engineering methods.

What follows are two additional follow on questions, which could be addressed as a part of the previous questions.

Follow on question 2a: Can the feasibility of genetic engineering application objectives be assessed by evaluating the different implementation approaches?

Below is a list of the objectives highlighted in Question 2 and possible broad approaches that could be taken to achieve the objective. This is meant to be a preliminary list. Can this approach be taken to “drill down” to specific applications and enabling methods that are likely threats to consider?

Objective: Detection evasion

1. Removing or altering known signature genes
2. Amino acid scrambling using protein structure modeling to identify amino acid alterations that may not affect function.
 - a. Some regions may have divergent sequence but conserved structure, thus still be very sensitive to amino acid changes
3. Introducing a known pathogenic gene/pathway to a non-pathogen so that the vast majority of the genome appears to be of a ‘benign’ nature.
4. De novo novel gene/pathway/chromosome

Objective: Robustness to countermeasures

1. Introduction of known antibiotic resistance genes
2. Introduction of mutations that render antiviral resistance

Objective: Optimize dispersal strategy

1. Introduction of human cell receptor recognition proteins (e.g. engineered zoonosis)
2. Deletion of cell surface antigen genes (reduce host immune response)
3. Introduce cell packaging pathways to ensure robust growth in diverse environments
4. Introduce cell signaling pathways to trigger cell function
5. Introduce genes to survive temperature changes

Objective: Spoofing

1. Introduction of known engineering vectors, minus the ‘payload’

2. Gene deletion/insertion/alteration with no phenotypical effects

Objective: Biochemical processing

1. Introduction of known genes for chemical production
 - a. Possible gene tuning
2. Introduction of de novo genes

These all fall broadly into one of four categories: introduction of a known gene(s)/pathway(s), deletion of a known gene(s)/pathway(s), alteration of a known gene(s)/pathway(s), De novo synthesis of novel gene(s)/pathway(s) or any combination of the four.

Follow on question 3a: How will the different genetic engineering tools affect the feasibility of the different implementation approaches?

This question follows from question 3 and question 2a. As an example, implementations of some application objectives will require specific genetic engineering tools, which could in part determine the degree of difficulty involved in achieving the objective. Below is an example (incomplete) list of possible tools:

- Self replicating plasmids with ‘payloads’
- Site specific integration
- Site specific recombination
- General homologous recombination
- Gene synthesis and assembly
- Directed evolution with phenotype screening

Conclusion

Developing the technical capability to mitigate genetic engineering based threats will likely be a major challenge, and the decision to direct resources at addressing this problem will have to be balanced against the needs to address other bio-security threats. This workshop will serve as an opportunity to begin to define the technical challenges and determine how these challenges can be addressed within the framework of a strategic vision that draws on the knowledge of a diverse array of experts working in fields related to genetic engineering. We will use this document as a basis for starting our unclassified discussions at the workshop. It is likely that additional discussions will be indicated, and we will urge the workshop attendees to include their suggestions as part of our workshop summary.