Working Paper – Sept. 17, 2009
Comments Strongly Encouraged

Making Commercial Biology Safer: What the Gene Synthesis Industry Has Learned About Screening Customers and Orders

Stephen M. Maurer,[1] Markus Fischer,[2] Heinz Schwer,[3] Cord Stähler[4], Peer Stähler,[4] and Hubert S. Bernauer[5, 6]

I. Introduction.

Scientists first learned how to make artificial DNA in arbitrary, gene-length configurations in the 1970s. [Khorana 1979] In principle, this immediately provided a new way to do genetic engineering. At first, however, the technology was difficult, expensive, and limited to very short (~ 200 nucleotide ("nt.")) molecules. For this reason it took almost twenty years for academic laboratories to improve the process to the point where synthetic DNA could begin to replace cloning in practical experiments. The turning point came in 1999 when several corporations began selling synthetic DNA commercially. Because the new firms served far more customers than any academic lab, they could afford to specialize and make large investments in R&D and equipment. This made artificial genes still more affordable, further boosted demand, and attracted additional investment. Today, this virtuous cycle is finally encountering diminishing returns. Even so, it will probably continue to run for another five years or so.

Falling DNA prices have conferred important benefits on society. First, traditional genetic engineering experiments have become vastly more affordable. This has boosted both academic and commercial research and will ultimately generate new products for consumers. Second, synthetic DNA has opened genetic engineering to many researchers (*e.g.* undergraduates, high school students) who lack advanced cloning skills. In this sense, synthetic DNA – like the microchip – has "brought computing power to the masses." Finally, synthetic DNA has made it possible to do many experiments that used

[1] To whom correspondence should be addressed. Goldman School of Public Policy and Boalt Law School, University of California, USA. smaurer@berkeley.edu.

[2] Entelechon GmbH, Regensburg, Germany.

[3] Sloning BioTechnology GmbH, Puchheim, Germany.

[4] Febit Synbio GmGH, Heidelberg, Germany.

[5] ATG:Biosynthetics, Freiburg, Germany.

[6] The authors wish to thank Damon Terrill, Tom Slezak, and Robert Jones for information and many helpful suggestions. Any errors or omissions are due to the authors alone.

to be forbiddingly expensive. This has spawned a new discipline ("synthetic biology") aimed at reengineering organisms to make everything from drug compounds to jet fuel.

Policymakers should be careful to preserve these enormously positive developments. That said, synthetic DNA could also be misused. While the size of the risk is uncertain [Maurer & Rutherford 2009], the potential consequences are large and justify reasonable precautions. This article looks at what the gene synthesis industry is doing to address these concerns. Section II introduces the basic gene synthesis technology and examines its implications for biosecurity. Section III profiles today's gene synthesis industry. Section IV examines support for screening among gene synthesis companies and their customers. Section V explores the sometimes subtle economics of screening. Section VI looks at the problem of identifying potential threats against a background of routine orders. Section VII looks at methods for confirming that dual use genes will not be misused. Section VIII describes current screening programs and Section IX presents various proposals for moving forward. Section X provides a brief conclusion.

II. The Technology

This section describes the main steps that gene synthesis companies typically perform[7] and their implications for biosecurity:

> *Step 1: Production Planning.* DNA is a custom product whose cost can vary by ± 50% depending on which sequence is ordered. Manufacturers must therefore design a new manufacturing plan for each order. This, in turn, means carefully examining orders for genes that code for substances (*e.g.* proteases) that can interfere with the bacteria that are used to amplify DNA or else pose worker safety issues (*e.g.* snake venom). In general, this examination has become steadily more comprehensive and automated over time as known manufacturing issues have proliferated and companies have learned that they cannot always trust customers' paperwork to disclose potential issues.
>
> *Step 2: Oligosynthesis.* Gene synthesis usually starts by making short building blocks of DNA called "oligonucleotides" or "oligos." [Baker et al. 2006][8] Ten years ago this required highly skilled technicians and was the biggest obstacle to gene synthesis. Since then, however, commercial R&D programs have successfully encoded the required skills into user-friendly machines. These routinely deliver the high quality (0.5%/nt error rate), appropriately sized (40-50nt.) oligos needed for gene synthesis. Furthermore, oligo prices have fallen to the point where they comprise just 30% of all total gene synthesis costs.

[7] There are, of course, important differences. For example, Sloning assembles genes from a library of standardized oligonucleotides that include just six variable base pairs. (J. Van den Brulle et al., 2008). These differences do not, however, materially affect our analysis.

[8] The exception are highly specialized synthesis techniques which require advanced custom-developed machines and high initial investment costs.

Steps 3-5: Assembling and Amplifying Genes. The next three steps include (3) assembling the oligos into a complete gene, (4) inserting copies of the gene onto a plasmid, and (5) inserting the plasmids into a bacterium. In principle, the work is straightforward and can be done in a high school biology lab using methods found in well-known textbooks like Sambrook and Russell (2001). In practice, however, our experience with undergraduate and high school interns shows that the required skills remain elusive. While interns usually grasp the underlying principles and protocols quickly, even good students have trouble reproducing manual operations correctly. This leads to repeated problems with contamination, incorrect concentrations of reagents, compromised enzymes and similar issues. Perhaps more importantly, interns seldom know how to work around unexpected obstacles that arise when, for example, bacteria fail to take up DNA. Most workers only acquire this knowledge after several years of practical lab work.

Step 6: Quality Control. Current synthesis methods typically deliver several flawed oligos for each correct one. Isolating perfect copies for cloning requires highly specialized skills and accounts for roughly 50% of all gene synthesis costs. Although basic methods are found in the literature, gene synthesis companies invariably use "tricks of the trade" that are nearly impossible for outsiders to learn. These are typically taught through a combination of graduate-level education and on-the-job training in gene synthesis labs. Gene synthesis companies are currently investing large sums to automate these steps. However, these solutions are expensive, complex, and demand highly trained workers. We therefore expect automation to remain unaffordable by all but the largest gene synthesis companies for many years to come.[9]

Steps 7: Producing Bulk Genes. The final production step involves growing and extracting genes from cloned bacteria. As with Steps 3-5, the work can theoretically be done in simple facilities using widely-described methods. In practice, however, tacit knowledge is important.

Implications for Biosecurity. None of the foregoing steps is beyond a well-trained microbiologist who routinely performs cloning experiments. For this reason, attempts to deny commercial gene synthesis services to nation states like Russia or even Iran have only limited value. Terrorists, on the other hand, have notoriously small budgets – Al Qaeda reputedly spends about $35 million per year [9/11 Commission 2004] – and are very sensitive to cost and technical risk [Ackerman and Bale 2009]. For this reason, barriers that add even modestly to cost ($100,000) or failure risk (10%) are typically worth considering.[10]

---

[9] This has not prevented some companies from trying to build more affordable automation solutions. For example, Febit designed a tabletop gene synthesizer in 2007 but decided not to release it for synthetic biology applications yet.

[10] We justify this benchmark by noting that Osama bin Laden came close to cancelling the September 11 conspiracy several times [9/11 Commission] despite its low cost (~$500,000) and high *ex ante* chance of success (~10 – 50%). Inevitably, this analogy is limited. On the one hand, terrorists would presumably be

By this standard, the lowest rung of the gene synthesis ladder – oligo production – is probably no longer worth protecting. Indeed, large commercial oligo suppliers exist worldwide and fill an estimated 25 million orders per year – 500 times as many as commercial gene orders. [Cassagrande (2008)] Furthermore, various manufacturers including Affymetrix, Illumina, Agilent, Roche, and febit have successfully encoded the skills needed to make oligos into highly automated machines. Furthermore, the Agilent, Roche, and febit machines provide especially flexible production platforms while the febit device integrates the entire process on a tabletop scale. These can be purchased from dealers and even on E-Bay. [Garfinkel et al. 2007] Unless and until these sources are regulated, attempts to limit access to the gene synthesis industry's own, comparatively modest, oligo production seems pointless.

But oligo production is only the first of several steps needed to manufacture a gene. Based on our experience with interns, we estimate that even gifted undergraduates would encounter failure rates of approximately 20%/step for Steps 3, 4, 5, and 7 and significantly higher rates for Step 6. Naively, this suggests that the odds of inexperienced workers correctly assembling a gene from commercial oligos the first time are about one in ten thousand. More realistically, persistent workers could probably overcome an initial failure by trouble-shooting experiments and/or consulting outside experts. This, however, would probably take a year or more during which the plot's cost and security risks would steadily increase. Alternatively, terrorists could reduce – though not eliminate – the challenge of assembling oligos by dividing their desired sequence into several orders none of which was long enough to trigger an alert.

Based on this analysis, we conclude that reasonable steps to deny commercial gene synthesis services to terrorists are almost certainly justified. This conclusion is likely to hold as long as automation remains difficult and expensive. Most gene synthesis companies believe that this condition will likely persist indefinitely.[11]

III. The Industry

Private sector biosecurity initiatives are inevitably shaped by business realities. This section describes the evolving gene synthesis industry and the mature market that is likely to exist five years from now.

Prices, Products, and Volume. The first gene synthesis companies (including Entelechon and Geneart) were founded in 1999. During the industry's early years competition

---

willing to bear additional cost and risk for the ability to kill tens and even hundreds of thousands of victims. On the other, terrorists who possessed synthetic DNA would still face very significant costs and risks before they could make a working weapon. Very roughly, we expect these effects to cancel.

[11] There are, predictably, dissenting voices. For example, Carlson (2008) concedes that "practical implementation of [automation] has thus far proven challenging" but predicts that unspecified breakthroughs could produce a tabletop gene machine by 2013.

largely took place on quality. As a result, prices held steady at about € 5.00/nt. for several years. As the technology matured, however, price became increasingly salient. This led to steep declines of about 20%/year from 2002 – 2005 and a more moderate, 10% per year decline since then. As of 2009, typical gene prices stood at roughly € 1.00/nt. although these varied by up to ± 50% depending on which molecule was requested. Falling gene prices, in turn, made genetic engineering research much more affordable. This was one of the prime drivers in the doubling of biotech R&D budgets between 1999 and 2008. [Bio 2008] Barring unforeseen technology breakthroughs, we expect prices to continue falling for another five years or so before bottoming out between € 0.30 - 0.50/nt.

Today, there are roughly 50 dedicated gene synthesis companies worldwide. [Garfinkel et al. 2007][12] This industry produces an estimated 50,000 genes per year [Cassagrande (2008)] with individual firms producing anywhere from ten to 2,000 orders per month. Assuming that prices continue to decline, production will probably double by 2015. Meanwhile, the average length of requested molecules is also growing. As of 2009, most companies filled 50 to 70% of their orders in the 500-2,000 nt. range with a few percent above 5,000 nt. Orders above 20,000 nt. are still rare, although some companies have synthesized genes up to 35,000 nt. in length.

Estimates. We have seen that DNA is a custom product whose cost can vary by ± 50% depending on which molecule is requested. Customers almost always submit a Quote Request[13] to find out how much their particular order will cost before placing an order. This industry-wide procedure has important implications for security:

> *Cheap Automated Screening.* Gene synthesis companies have developed elaborate automated systems to analyze requested sequences, prepare production plans, and generate estimates. Extending these systems to include automated sequence screening is straightforward and inexpensive. By comparison, other forms of screening (*e.g.* researching unknown genes, investigating new customers) remain labor-intensive and costly.
>
> *Binding Commitment.* A gene synthesis company can only stay in business as long as customers believe that it will actually deliver DNA as promised. In practice, this means that companies must investigate and resolve all security issues *before* tendering a bid.
>
> *Uncompensated Work.* Gene synthesis companies recover the average cost of preparing estimates as overhead. However they do not – and as we will see probably cannot – impose surcharges on customers who submit particularly problematic orders. In a few cases, companies may decide that preparing a bid is unprofitable. These orders will never be screened at all.

[12] Prominent examples include ATG Biosynthetics, Biomax Informatics, Entelechon, Febit, Sloning, and Geneart in Europe; Blue Heron, DNA 2.0, and CODA Genomics in the United States; Geneworks in Australia; and Sangon and ShineGene in China.

[13] Geneart's quote request form is typical. http://www.geneart.com/english/quote-request/index.html

Market Power. Today, the synthesis of virtually all genes involves at least some manual steps and economies of scale are modest. This allows some gene synthesis companies to survive on revenues of just €1 – 1.5 million per year. On the other hand, automation will only cover its expected investment and maintenance costs if it is done on a very large scale. This suggests that the industry will be dominated by one or two very large, low cost firms by the middle of the next decade. There will probably also be five to ten smaller firms that specialize in technologies (including traditional manual methods) optimized around specific classes of genes that the dominant firms cannot readily mass-produce.

The new incumbents will possess considerable market power. On the other hand, their economies of scale – and survival – will depend on attracting and retaining a critical mass of customers. We therefore expect large customers to exert substantial leverage over their suppliers' prices and service terms. Indeed, most large and mid-sized customers have already developed elaborate strategies to exploit their buying power. These nearly-universal practices include (a) concentrating orders on one or two "preferred vendors" to maximize buying power, (b) centralized purchasing systems to monitor prices and service, and (c) close, ongoing working relationships between customer and vendor employees.

A Global Market? One would expect an industry that manufactures high value, easily shipped products to have a global market. However, most Western gene synthesis companies still export less than five to ten percent of all orders outside the US and Europe.[14] Furthermore, most of these sales are focused on Asia. By contrast, sales to the Middle East account for less than one percent of all revenues and most of these come from Israel. This situation undoubtedly reflects biotechnology's still-incomplete penetration in many parts of the world. On the other hand, low demand cannot be the whole story. Indeed, even the US and European markets are still significantly segmented with US firms doing about two-thirds of their business in North America and Europeans doing about 85% of their business in the EU.

Most observers still expect to see a global market emerge within the next decade. Indeed, many of today's largest gene synthesis companies (*e.g.* Sangon, ShineGene) are located in Shanghai. That said, current markets are still significantly segregated. The reason seems to be that practically all gene synthesis companies find it significantly cheaper to sell at home than abroad. For example, many Asian firms still lack efficient rich nation distribution networks. This forces them to sell through independent middlemen at a markup. [Anon., 2005] Conversely, we will see that US and European companies must obtain export licenses for so-called "dual use genes." Companies must typically invest about 20 hours of employee time to license each order. Furthermore, the process typically takes six to eight weeks. Knowing this, Asian customers that expect to place even a few "dual use" orders will often prefer domestic suppliers to US or European competitors.

[14] There are exceptions. One European company currently earns about 40% of all revenues from exports to the US.

Alternative Suppliers. Because of specialization, gene synthesis companies can provide genes at unmatched prices, lengths, and quality. On the other hand, the basic molecular biology skills needed to synthesize genes are widely distributed across commercial sector and academia:

> *Small Commercial Suppliers.* Several hundred companies currently offer gene synthesis as a low volume sideline. However, their prices are three to four times higher than those offered by gene synthesis industry and their ability to make molecules longer than 5000 nt. is currently limited. In practice, they almost always serve customers who (a) find it convenient to work with a familiar local supplier, and/or (b) expect the supplier to supply the DNA as part of a much larger project. While we expect falling gene prices to drive many of these firms from the market[15], some small companies will almost certainly offer synthesis services for the foreseeable future.
>
> *Academic Laboratories.* Several hundred university laboratories possess the required skills to make genes. While these organizations hardly ever perform work for hire, the possibility cannot be ruled out.

We have already explained why biosecurity precautions that impose material added cost, technical difficulty, and/or security risk on would-be terrorists are generally worth pursuing. By this test, the gene synthesis industry should improve its screening capacity whether or not small, high cost providers participate. That said, it would obviously be preferable to include these providers in the long run.

IV. Support For Screening

Commercial screening programs need funding to be sustainable. In practice, there are three possible sources: (a) legal mandates that require companies to practice screening as a cost of doing business, (b) vendor market power to impose screening fees that could not be sustained in a less concentrated market, and (c) the willingness of at least some large customers to demand (and pay for) company-wide screening policies.

Legal and Regulatory Requirements. To date, no government has specifically tried to regulate gene synthesis. However, many parts of the process are subject to a patchwork of older regulations that collectively encourage screening:

> *Manufacturing Safety.* German laws governing work that involves cloning, GMOs, toxins, and resistance genes usually require companies to know which

[15] The case of commercial oligo production is instructive. During the 1990s many companies and universities built centralized facilities to make oligos. Many of these have now been decommissioned in the face of falling oligo prices. [Garfinkel et al., 2007] However, other small businesses still offer high-cost oligos as a convenience to clients.

genes they are handling. In theory, these requirements could be evaded by developing production plans in which genes are never introduced into living bacterial host cells. This, however, would be expensive so that compliance -- and screening – is almost always cheaper in practice.

Biosafety. The US Toxic Substances Control Act ("TSCA") prevents companies from shipping orders to unknown addresses. Similarly, EU law limits shipments to customers who follow European safety standards. In practice, however, these standards seldom require customer investigation of known customers and universities. In most other cases, companies can usually meet the requirement by asking customers to provide written – but otherwise unverified – assurances before filling the order.

Biosecurity. US law requires gene synthesis firms to screen orders for (a) genome-length sequences which "are inherently capable of producing a select agent virus," or (b) code for select agent toxins. [NSABB 2006][16] However, these requirements do not extend to other select agent genes that confer virulence. US law also requires companies to obtain an export license when they make so-called "dual use" genes for foreign customers.[17] The situation for European firms shipping genes outside the EU is similar. There is widespread agreement that these laws extend to individual genes.

Legal Liability. Gene synthesis companies would face lawsuits if their products were ever misused. It is impossible to know in advance what duty of care would be imposed in this situation. However, there is widespread consensus that companies have an obligation (a) to search orders for select agent genes, and (b) conduct follow-up customer investigations if these turn out to be virulent.

Records Retention. The US TSCA statute requires gene synthesis companies to retain records for five years. Similarly, local law in most EU countries (including Germany) requires companies to retain records of all GMO experiments. According to a recent survey, "nearly all" gene synthesis companies retain data concerning customer orders for at least two years. [Cassagrande 2008]

Industry Support for Screening. Practically all gene synthesis executives recognize a social obligation to screen orders. This presumably implies some willingness to support screening from existing profit margins.

Customer Support for Screening. Large customers want their suppliers to operate ethically and are willing to withhold business from firms that do not. Indeed, at least one

[16] The preamble to the Select Agent list states that "It is incumbent upon the entities that manufacture substances to know what they are manufacturing."
http://www.selectagents.gov/resources/42_cfr_73_final_rule.pdf

[17] 7 Code of Federal Regulations Part 331, 9 Code of Federal Regulations Part 121, and 42 Code of Federal Regulations Part 73. See http://www.selectagents.gov/selagentRegulation.htm.

large pharmaceutical company has explicitly announced that it is "committed to working only with contractors, such as suppliers…who embrace standards of ethical behavior that are consistent with our own." [AstraZeneca 2008] To date, however, customer involvement has been limited by the relative obscurity of the screening standards issue. Indeed, most companies report that they currently have less than one conversation per year on the subject. That said, we are aware of several instances in which large, repeat customers have gone out of their way to confirm that vendors have screening programs in place even though they clearly understand that they will have to pay higher gene prices to support them. None of us has ever heard a customer ask for less screening, although it would be very surprising if such views did not exist in at least a few cases.

Beyond these direct statements, customer confidentiality expectations also limit vendors' ability to screen.[18] These restrictions usually focus on preventing gene synthesis companies from sharing information that could potentially confer economic advantage[19] with third parties and/or transmitting it over the Internet. Conversely, confidentiality places few if any restrictions on in-house use of the data.[20] Similarly, companies are not required to withhold information required by law or court order or to obtain an export license for dual use genes. To the best of our knowledge, all gene synthesis executives agree that they are prepared to report suspicious orders to the authorities.

V. The Economics of Screening

We have seen that gene synthesis executives and, especially, large customers support screening. At the same time, this viewpoint is unlikely to be unanimous. In general, we expect gene synthesis companies to follow the wishes of whichever customers command the most purchasing power. This situation is logically consistent with at least four distinct outcomes:

> *Universal Screening.* All gene synthesis companies practice voluntary screening in a single, worldwide market.

[18] While the figures vary somewhat from company to company, roughly one-half to two-thirds of all customers require their gene synthesis company to sign a written confidentiality agreement. This figure includes practically all corporations, most large customers, and some academic orders. A few gene synthesis companies (*e.g.* DNA2.0) offer a standard confidentiality agreement to all customers.

[19] Examples typically include (a) specific DNA sequences requested, (b) the encoded protein sequence, (c) specific combinations of genes, and (d) the existence of the order itself.

[20] Beyond these mainstream views we are also aware of occasional reports of customers who have asked providers not to query sequence identity internally (Imperiale 2008). This would seemingly prevent all but the simplest biosecurity screening and even normal manufacturing planning. At the moment, these positions are clearly outliers. Indeed, we have found that the overwhelming majority of our customers cooperate fully when security issues arise. In those few cases where customers did express trade secret concerns, signing a confidentiality agreement was sufficient to go forward. At the same time, there are no guarantees that these attitudes will persist. The fact that some companies try to exploit confidentiality as a marketing tool [Minshull 2009] could encourage some customers to press such positions in the future.

Tiered Screening. Some companies screen and are patronized by consumers who are willing to pay for this service. However, other companies openly adopt a "no screening" policy and serve customers at the lowest possible price.

Regional Screening. A global market never emerges. Companies and customers practice screening in some regions but not others.

No Screening. No companies screen orders anywhere in the world. The existence of robust, non-screening firms convinces customers that might otherwise favor screening that such measures are pointless.

Because screening must be industry-wide to be effective, only the first of these scenarios ("Universal Screening") offers useful biosecurity benefits. Remarkably, we will see that today's US and European gene synthesis companies closely approximate this regime.

Limits of Voluntary Screening. Given Universal Screening's potential instability, policymakers should be careful to avoid any intervention that might destabilize it. Here, we argue that attempts to impose very high screening standards could produce precisely this outcome. To see why, recall that Step 1 ("Estimation") can be subdivided into the following three steps:

(1a) *Computer Screening.* Screening begins with a computerized comparison of the customer's order against a list of potential threat genomes. This step is highly automated and costs very little.

(1b) *Human Screening.* This step requires follow-up investigation of any suspicious "hits" identified in Step 1a. This step requires human labor and is potentially expensive. Furthermore, companies can predict these costs based on the number of "hits" generated by Step 1a.; and

(1c) *Estimation.* This step involves submitting a written offer to the customer promising to make the requested gene at a specific price.

Now consider how a profit-maximizing firm would analyze this process. Starting with Step 1c (Estimation), there is a familiar economic maxim that "sunk costs are sunk," ***i.e.***, that a company contemplating new expenditures should only ask how much profit it can make going forward and *ignore* costs that have already been spent. This means that the price offered to customers at the end of Step 1c will not reflect the actual costs of performing Step 1b (Human Screening).[21] Knowing this, companies that have completed Step 1a (Computer Screening) will not proceed at all if there are so many "hits" that the expected profit from obtaining and filling the order does not cover the expected cost of

[21] Companies can, of course, charge an overhead item to reflect *average* costs over all customers.

Steps 1b and 1c. All else equal, we would expect this to happen more often when screening standards are high.[22]

How Stable is Universal Screening? We have argued that gene synthesis companies operating five years from now will need a customer base equal to roughly 10% of the market to remain competitive. This immediately suggests that a Universal Screening regime will be absolutely stable if more than 90% of all customers support screening. Furthermore, we expect a Universal Screening regime to be "metastable" even if this is not the case so long as no single customer can supply enough business for a new, non-screening company to achieve minimum efficient scale. Instead, the hypothetical company would have to engineer a *coordinated* move by l*arge numbers* of customers. This is highly unlikely so long as screening costs are low. At the same time, Universal Screening is not absolutely stable. Indeed, the existence of even one openly non-screening vendor could easily snowball if it persuaded customers who currently support screening that further efforts were pointless. In this case, initially small defections could rapidly degenerate into a self-reinforcing cycle in which customer decisions to abandon screening encouraged more firms to reduce their screening programs and *vice versa*.

In general, there is very little theoretical guidance on when this tipping point would occur. On the other hand, customers who were turned away under an existing industry-wide screening standard would be strongly motivated to find and place orders with a new, non-screening entrant. This suggests that a screening standard that turns away ten percent of all customers is very likely to collapse into a Tiered or Regional regime. More conservatively, standards that turn away even a few percent of all customers should be viewed with caution.

The Screening Budget. The foregoing discussion suggests that screening practices should be set so that the number of customers rejected (R) is equal to, at most, a few percent. This can be expressed mathematically by the formula:

$$R = S_A S_M(c_1) C_A C_M(c_2), \qquad \text{(Eqn. 1)}$$

where S_A is the fraction of sequences deemed suspicious by automated sequence screening, S_M is the fraction of sequences that cannot be cleared by follow-up manual inspection of the suspect sequence, C_A is the number of customers that are deemed suspicious by some automated process, and C_M is the number of these customers who cannot be cleared by follow-up manual investigation. Automated screening steps are deemed to be costless while c_1 and c_2 represent the maximum budget for, respectively, follow-up manual investigation of suspect sequences and customers. The remainder of

[22] In principle, companies could avoid this catastrophe by offering to defer screening until after the customer signed a contract and then placing a bid based on the *expected* cost of screening. We would expect such a system to be significantly more stable than one which turned customers. Nevertheless, the end result would be similar. We would expect the system to become unstable shortly after the number of customers paying a security surcharge became large enough so that a new, non-screening entrant could produce DNA at prices low enough to undercut the surcharge.

this article will provide estimates for S_A, S_M, C_A, and C_M, c_1 and c_2 under both current and proposed systems.

Globalization and Screening. We have seen that synthetic genes are still sold in segregated regional markets. This could potentially produce a Regional Screening system in which Asian companies practiced noticeably less screening than the rest of the world. Conversely, the emergence of a world market would (a) encourage large Asian firms to compete for US and European customers who value screening, and (b) drive small, non-screening Asian firms from the market.[23]

VI. Sequence Screening

Perfect sequence screening would allow companies to fill most orders immediately and minimize the number of cases that require expensive human follow-up. This section describes the various design issues that face screeners:

The Logic of Sequence Screening. Any screening program must start with an express or implicit model of what to look for. In practice, most screening programs focus on "select agent" organisms that have been successfully weaponized in the past. However, many other threats are possible. These include organisms that have not previously been used as weapons (*e.g.,* 1918 influenza); weapons based on normally benign genes (*e.g.* genes that regulate human growth hormone); and so-called advanced weapons that have been so heavily reengineered that the original wild type source is unrecognizable. *See, e.g.,* Maurer (2009), Maurer & Rutherford (2009) and Dando (2001). Future screening programs will almost certainly have to search for some or all of these threats. Conversely, screening that uses the organism's identity as a threat proxy is almost always overbroad since it includes large numbers of genes are unrelated to pathogenicity.[24]

Compiling a threat list is only the first step. Companies must also link the threats to specific genes and, ultimately, customer-submitted sequences:

> *Linking Sequences to Specific Genes.* In principle, this can be done either by (a) BLASTing the submitted sequence against the NCBI database and manually examining the closest matches, and/or (b) screening against genes on a specially prepared threat lists such as the US Select Agent list.[25] In both cases, the problem is knowing how many of the top matches to examine. Given that similar

[23] Conversely, the emergence of a global market would expose US and European firms to Asian customers who might prefer less screening. This is unlikely to matter, however, as long as Western customers dominate the world market.

[24] The case of puffer fish toxin (tetrodotoxin) is illustrative. Researchers suspect that this molecule is produced by symbiotic bacteria within the fish. Pending conclusive evidence, however, screeners continue to compare orders against the entire pufferfish genome. This routinely produces false positive hits against sequences based on animal genes.

[25] Closely related lists have also been prepared by the Australia Group nations, the European Union, and virtually all national governments.

sequences can have radically different functions, establishing that a few top matches are benign can never completely demonstrate that an order is safe. Screening against known-threat lists partially compensates for this problem by focusing screeners' attention on known pathogens and may even provide some limited protection against threat genes from other pathogens that share common ancestors.

Threat Judgment. The next step is to make a judgment about whether the identified gene constitutes a threat. In practice, most companies rely on the biological concept of virulence as a proxy. For well-annotated Genbank entries, virulence status can usually be established in a few minutes. In other cases, workers must check Genbank's references by hand. This typically takes two hours per gene.

However sensible, the current use of select agents and virulence factors as threat proxies has several defects. These include:

Cost. Paying human screener to perform manual virulence research consumes resources that could otherwise be used to screen against a larger threat list and/or investigate more customers.

Virulence vs. Threat Potential. The biological concept of virulence is only approximately aligned with threat potential. For instance, genes that code for surface proteins rather than enzymes are not considered virulent because they cannot harm the host. Nevertheless, weapons developers might still find them useful in stabilizing pathogens against environmental stresses and evading the human immune system. Similarly, many genes only confer virulence in the presence of other genes and/or certain biological contexts (Imperiale (2008)). In this context, it is almost impossible to know when isolated genes pose a threat. Similarly, terrorists could potentially exploit combinatorial synthetic biology methods to engineer lethal interactions among individually harmless enzymes. Finally, the genes that confer virulence on some organisms (*e.g.Francisella tularemia*) are only beginning to be discovered.

Non-Traditional Threats. Many naturally-occurring pathogens have never been investigated by military biological weapons programs but could nevertheless appeal to terrorists. [Maurer and Rutherford 2009] Furthermore, some potential threat genes do not come from pathogens at all. Examples include genes that modify human growth hormone and code for plant and animal toxins. Finally, sequences for a so-called "advanced agent" would probably bear little resemblance to any preexisting organism.

Gene Fragments. An ideal system would screen DNA fragments down to the 50nt. lengths from which genes are made. In practice, however, the chances of an accidental match rise steeply as gene length declines.

Customer Orders. Gene synthesis companies have learned great deal about customer orders from production planning, security screening, consulting, and other services. This provides important information about how often even benign orders require investigation:

> *Naturally-Occurring Genes.* Current screening programs are able to identify progenitor organisms in roughly 75% of all customer orders. These sequences have almost always been genetically engineered although many of these changes are modest. For example about one-quarter are more than 95% identical to the parent sequence. On the other hand, 10-15% have been so heavily revised that less 35% of the original sequence remains.
>
> *Pathogen Genes.* Only about one percent of the orders which can be identified involve genes derived from bacteria or viruses. Furthermore, fifty to sixty percent of these genes turn out not to be virulent. Of the remainder, about one-half code for toxins. Most customers who request these sequences are engaged in drug and vaccine research. Here, the most common experiments involve using truncated viral antigens to develop vaccines, making bacterial and viral enzymes to test the inhibitory effect of candidate drugs, and making organisms resistant to antibiotics. Still other customers use viral and bacterial promoters to increase gene expression.
>
> *Artificial Sequences.* Roughly 10-15% of gene orders involve artificial sequences that are not modeled on any living template. Anecdotally, most if not all of these sequences are used for as calibration standards or tests for enzyme specificity. We are not aware of any instance in which customers have tried to design an artificial gene that coded for proteins that could be used in a living organism.[26]
>
> *Unknown Sequences.* Current screening practices fail to identify roughly 15% of all orders. Since most of these searches already include a complete Genbank search further effort is not likely to reduce this number. Most of the unknown orders are probably arbitrary constructs which have no correlate in nature.

Based on the foregoing, we expect automated select agent searches to identify "hits" about one percent of the time; broader but still reasonable threat lists which included, say, plant and animal toxins could increase these rates several times over. In terms of Eqn. 1, therefore, we estimate $S_A = 1\%$ for the select agent list and $S_A = 2\text{-}5\%$ for broader searches. In either case, follow-up manual investigation will usually demonstrate that about half of these orders are benign ($S_B = 50\%$).[27] This suggests that at most only a few orders in a thousand will require follow-up customer investigation. We therefore conclude a Universal Screening regime can safely support these standards. The case

[26] Current screening practices do This leaves at most 15% of all orders in the unknown category. BLASTing orders against the entire GENBANK database could presumably drive this number to within a percent or so of zero and would ultimately be limited only by GENBANK's own incompleteness. We have no reason to think that these orders are significantly more problematic than those we have

[27] We will see in Section VIII that the resources available for manual sequence screening c_1 are more than adequate for this task.

could be different, however, to the extent that screeners dispensed with predetermined threat lists in favor of examining homologous genes. Here, cost would depend on how many matches are reviewed and is potentially unbounded.

Prospects for Improvement. We have seen that screeners must currently determine the virulence status of select agent genes manually. Compiling a comprehensive, automated list of these genes would cut these costs (c_1) almost to zero. More generally, screening would also be improved by constructing broader threat lists to protect against non-select agent pathogens, genes that code for hormones or toxins, and/or advanced weapons. There are several ways to construct such a list:

> *Examining Top Matches.* Companies could require screeners to determine the virulence and/or threat status of a predetermined number of best matches for each customer order. Depositing search results into a common database would gradually build up a comprehensive list of threats and also benign genes.
>
> *Developing New Proxies.* In principle, it should be straightforward to compile extended threat lists that include potential (but non-traditional) that are known to the literature. These threats would presumably include a wide variety of pathogens, toxins, biopromoters, and the like. Furthermore, weapons based on new science would be difficult or impossible to predict. This new science could be quite simple, *e.g.* finding and sequencing a threat gene not previously reported in Genbank. Alternatively, weapons makers could potentially use advanced combinatorial synthetic biology methods to engineer lethal interactions among enzymes that are individually harmless. These latter threats would be almost undetectable based on the literature and scientific understanding available today.
>
> *Virulence Algorithms.* In the long run, researchers would like to replace specific threat lists with algorithms for detecting virulence in arbitrary sequences. Recent work showing that many virulence factors are located within so-called pathogenicity islands [Hacker et al. 2000] suggests that adding sequence location data to current homology scores could dramatically improve virulence prediction. Other promising avenues are based on examining predicted proteins against a Hidden Markov Model or similar statistical technique. In addition, textmining has already been successfully applied to extract known facts about genes from publications and can likely be used to identify and confirm genes associated with new threats (see for instance [Zaremba et al 2009]).

Expanding searches beyond the select agent list will inevitably require more human follow-up including customer investigations. However, there is no guarantee the savings from improved threat lists can cover these costs. In this case, government support could be needed to fill this gap.

VII. Customer Screening

We have seen that practically all problematic gene orders are dual use, *i.e.* may or may not be appropriate depending on what is done with them. This section reviews various methods for screening customers and their experiments.

The Logic of Customer Screening. Ideally, customer screening should perform three separate functions:

> *Identification.* Successful identification enables deterrence by establishing a "return address" that authorities can use to find and punish wrongdoers if a plot is detected. Identification also produces information about the customer (*e.g.,* nationality, residence, employment, and/or academic affiliation) that provides at least weak statistical evidence that the order is (or is not) legitimate.
>
> *Intended Use.* This step requires (a) confirming that the proposed experiment is genuine and not a cover story, and (b) documenting the proposed work in sufficient detail to reach a judgment about whether it should be allowed to go forward.
>
> *Legitimacy.* This step involves making a judgment that the requested gene will not be used (a) to perform illicit biological weapons research, and (b) to perform so-called "experiments of concern" that could potentially facilitate the design or acquisition of weapons.[28] Fortunately, Steinbruner *et al* (2007) have estimated that fewer than one percent of all academic research projects present serious "experiment of concern" issues.

Investigation Techniques. No search mechanism is perfect and we will see that gene synthesis companies must establish identification, intended use, and legitimacy using, at most, a few hours of employee time. Within these constraints, several lines of evidence are possible:

> *Direct Evidence.* The most obvious way to gather information is by direct contact with the experimenter and/or examination of the proposed experiment. However, gene synthesis investigations are inevitably brief and typically rely on telephone conversations, letters, and e-mails. Assessing trustworthiness in these circumstances is often difficult.
>
> *Independent Verification.* Companies can consult trusted contacts who know the researcher and his work. Examples include entities and people who are already known to investigators (a) from past interactions, (b) by reputation, or (c) have been randomly selected from trusted web sites and industry directories.

[28] More formally, the US National Science Advisory Board for Biosecurity has defined "experiments of concern" as "research that, based on current understanding, can be reasonably anticipated to provide knowledge, products, or technologies that could be directly misapplied by others to pose a threat to public health and safety, agricultural crops and other plants, animals, the environment, or materiel." [NSABB 2007]

Redundant Information. Customer investigation provides very little added value in cases where it simply confirms work that is already known to authorities. Gene synthesis companies should normally deliver genes to researchers who can produce select agent licenses, evidence of government grant support, and/or Institutional Biosafety Committee approvals.

Signaling. Many lines of evidence depend on what economists call "signaling," *i.e.* requiring the experimenter to do things that would be difficult or impossible for terrorists to accomplish. This can include showing that the experimenter commands resources that are large relative to any known terrorist budget and/or operates openly in a way that invites scrutiny from large numbers of people. Examples of the former include evidence that the experimenter has placed large numbers of benign DNA orders in the past or is performing work for a large established company. Examples of the latter include showing that information about experimenter's work is available from the scientific literature, business press, or within the experimenter's host institution. Given that experiments of concern by their very nature tend to be performed by senior researchers, we expect publication data to be available in most instances.

Institutional Controls. Gene synthesis companies have very little ability to monitor and control experimenters. For this reason, it is often cheaper and more effective to rely on the researcher's home institution to monitor the work. In these cases, it will usually be sufficient to confirm that the home institution is aware of the DNA order, has reviewed the proposed experiment, and has made a facially-reasonable determination that the work should go forward.

Direct evidence is almost always needed to establish identification. Reliable methods include (a) confirming the experimenter's affiliation through independent web research, (b) confirming the experimenter's contact information with an entity, researcher, or shipping address that the investigator knows from past experience, or (c) requesting select agent licenses and other evidence of government approval. Other methods – including, for example, consulting credit reports or examining the return addresses on e-mails – are notoriously subject to fraud and should be used, at best, as confirmatory evidence.[29]

Arguments based on signaling, redundant information, and institutional controls become increasingly important once identification is complete. This is particularly true for legitimacy judgments. In theory, gene synthesis companies have three possible choices: (a) make an independent judgment that customer's proposed experiment should proceed, (b) make a judgment that allowing the proposed experiment to go forward is facially compatible with mainstream views, and (c) confirm that the experimenter's home institution has taken reasonable steps to review the experiment and subsequently

[29] Company identification also includes confirming that customer and vendor employees are acting in their employers' interests and have not been suborned by outsiders. Requiring several employees to approve each order is a reasonable precaution in this regard. [Garfinkel et al. 2007]

approved it. In practice, however, the first option – which requires gene synthesis companies to second-guess customers who almost certainly know much more about the proposed experiment – is seldom realistic. In practice, therefore, we expect most gene synthesis companies to adopt the second or third options. Probably the hardest case is where customers are not affiliated with sophisticated host institution. In these cases, gene synthesis companies should consider asking customers to submit their experiments to outside reviewers. This can be done by retaining outside biosecurity consultants [Friedman (2008)] or else consulting on-line advice portals.[30]

Customer Base. We have seen that many of the cheapest and most effective customer screening methods depend on connecting the experimenter to a known institution. Remarkably, this condition is often satisfied:

> *Corporate Customers.* Most gene synthesis companies earn between one-half and three-quarters of their revenue from corporate customers. Most of these firms are large businesses like pharmaceutical companies, mid-sized biotech firms, and, in some cases, chemical or agriculture companies. These customers almost always (a) possess elaborate information systems to authorize, track, pay for, and receive orders at clearly defined delivery addresses, (b) use "preferred vendor" strategies that encourage ongoing employee-to-employee relationships between suppliers and customers, and (c) operate formal Institutional Biosafety Committees. Small companies, particularly biotechs have much less elaborate in-house controls and are to that extent harder to screen.
>
> *Repeat Business.* Most gene synthesis companies earn about one-third of their revenues from just three customers and two-thirds from the top ten. Vendors usually know these institutions intimately and are expected to bring irregularities to their customer's attention.
>
> *Academic and Government Orders.* The remaining one-half to one-fourth of all orders typically comes from smaller, non-commercial buyers. Here the biggest groups include university laboratories (10-15% of orders) and researchers within government research institutes. In the US and Europe, these institutions typically operate centralized accounting systems that cover all orders and payments. However, the systems tend to be less intrusive than their private sector counterparts and frequently permit shipments directly to the requesting laboratory. They are nevertheless almost certainly sufficient to detect imposters or unauthorized purchases. Furthermore, confirming the customer's institutional affiliation and/or Institutional Biosafety Committee status is usually straightforward.[31]

---

[30] Berkeley recently launched an on-line advice Portal where researchers contemplating so-called experiments of concern can obtain expert, impartial, and independent biosecurity advice. The site is open to all researchers regardless of nationality or institutional affiliation. *See* http://gsppi.berkeley.edu/EoC/uc-berkeley-synthetic-biology-security-program/experiments-of-concern/.

[31] Interestingly, current gene synthesis efforts occasionally detect entities performing defensive bioweapons research for US and European governments.

First Time Customers. Orders from completely new and unknown customers are average a few percent of all orders. That said, they contribute a disproportionate number of problematic sequences. The reasons for this are unclear, but may reflect the fact that these orders often involve long-term problems (e.g. vaccine development) where a small set of genetic constructs lasts for many years.

Developing World Customers. Host organizations in the developing world often have minimal or unknown procedures for monitoring and controlling their researchers. In these cases, institutional affiliation provides relatively little assurance that a particular order is both genuine and legitimate.

Unaffiliated or Anonymous Individuals. Individuals (and very small companies) who lack institutional affiliations, accreditation, or publication records are among the hardest to investigate. Fortunately, we have received only a tiny handful of such orders.[32] We are also aware of a few cases in which researchers have sought to remain anonymous, including one customer who tried to obtain HIV-related sequences using a credit card. Our companies invariably refuse and report such orders to the authorities.

Based on the foregoing discussion, customer screening systems that rely on signaling, redundant information, and institutional control evidence are currently sufficient to clear most customers. This suggests that C_M is no more than 15-20%. In principle, this figure can be further reduced by more labor-intensive direct methods. Depending on how standards are set, however, this may not be affordable for the roughly five percent of all cases that involve individuals and first-time customers. This is particularly likely to be true where the customers in question are located in foreign countries and, especially, the developing world.

Prospects for Improvement. We have argued that commercial screening budgets cannot support extensive direct methods. If these are required government will have to fund them through, for example, a customer licensing program. On the other hand, gene synthesis companies typically learn a great deal about their customers from repeat information. Systematic attempts to capture and share this information would cut customer screening costs dramatically. Here, the main difficulty would be protecting customers' legitimate confidentiality expectations.

---

[32] Commentators have long predicted the rise of biohackers or do-it-yourself ("DiY") biologists. To date, however, none of us has ever received orders from such amateurs. Furthermore, most DiY projects would probably not be legal in Germany and most of Europe without permits. Similarly, nothing in the DiY literature suggests that US hobbyists currently conduct anything remotely resembling gene scale experiments. [McKenna, 2009; Bousted, G. 2008; DIYBio n.d.]. In the words of one community leader, "The DIYbio community currently does negligible amounts of synthetic biology, if any at all. To the best of my knowledge no DIYbio project has ever used commercial gene synthesis services, and such projects are unlikely in the near term." [Jason Bobe, personal communication]

VIII. Current Practice

We have seen that all gene synthesis companies examine submitted sequences to protect worker safety, design manufacturing plans, and prepare estimates. However, relatively few companies screened for biosecurity threats *per se* before the 2001 Washington Anthrax attacks and 2002's synthetic polio paper. Even then, progress was slow. For example, *New Scientist* found that nine out of 12 surveyed companies – including two out of three gene synthesis companies – still did not screen incoming orders routinely as of late 2005. Today, the situation is very different. Based on our own experience in workshops and bilateral contacts, all IASB companies operate screening programs that include significant human follow-up.[33] Despite this, rumors persist that some Asian companies do not screen. These should be very easy to confirm or deny.[34]

Current Practice. Gene synthesis companies have shared relatively little information about their respective screening programs until very recently. This explains why different companies tend to use slightly different methods:

> **Company A.** This company searches against a list of select agents and some additional threats. These searches require follow-up customer investigation in approximately two percent of all cases. Customers must produce a select agent license where high-homology matches are found.
>
> **Company B.** This company performs an automated search against a curated list of virulent select agent genes but does not examine the results manually. Instead, it sends customers a letter asking them to disclose the gene name and sequence, the "purpose of the construct (*e.g.* vaccine development/diagnostic test development, etc.)," and the biosafety level required for handling the completed gene. Customer responses which have been signed by a corporate vice president are accepted without further investigation. Furthermore, high homology matches require evidence of a select agent license.
>
> **Company C.** This company BLASTs customer orders against a local copy of the entire NCBI list and examines all homology matches up to an expect value of 10. It also BLASTs orders against the Select Agent, Australia, and German national lists. These genes are considered positive hits that require manual review if their homology score is greater than 50% of the best NCBI candidate. Customer

[33] Even today, it is not entirely clear how many companies specifically practice biosecurity screening or at what level. For example, Gryphon scientific survey found that "about 85%" of gene providers screen both customers and sequences. [Gryphon 2008] These figures, however, include an unknown number of oligo manufacturers.

[34] Companies often ask friendly customers to solicit quotes in order to learn about competitors' prices and capabilities. This method also provides clear evidence when companies do not screen. Similar red team methods could easily be adapted to find out whether companies screen incoming orders against, say, select agent organisms.

screening determines whether the the user exists and is a known organization, but otherwise is limited to biosafety for shipments within the EU.

Company D. This company BLASTs customer orders against local copy of the NCBI database which is updated quarterly and inspects the top ten homology scores by hand. It also checks orders against an in-house database of genes whose pathogenicity status has previously been investigated. Suspect sequences are forwarded to two specialists – both of whom hold PhDs in medical microbiology –for a case-by-case threat determination. The company also investigates all dual use gene orders with telephone calls and internet research to confirm that an end user exists, is a known organization, and has a legitimate use for a given gene. The company also screens all customers against a government-provided database of prohibited organizations and individuals. [BAFA, 2007] Finally, EU law requires the company to obtain export permits for all dual use shipments outside Europe. This even includes cases where the customers is, for example, a well-known US company.

Company E. This company BLASTs customer orders against the entire NCBI list and examines the top ten hits. This requires follow-up virulence research in 3-10% of all cases. This leaves some safety (10%) and biosecurity (5%) issues. Clearing these issues requires customer contact roughly one percent of the time. The company generally accepts orders from known institutes and companies but has refused one anonymous order and five toxin orders in the past. The company estimates that screening effort absorbs roughly 0.5% of its revenues

Company F. This company uses Black Watch software to screen customer sequence requests. The curated database includes all DNA and Protein sequences from GenBank for organisms considered hazardous biological agent5s by CDC, USDA, or US Commerce Department Regulations. Positive hits are reviewed and and verified manually. Every order is also checked for indications of possible misuse. Steps include checking the order's country of origin, customer institution, and confirming that the invoice and delivery addresses match. Customers are also required to disclose the gene name and sequence, the "purpose of the construct," and the biosafety level required for handling the completed gene. All order data are indefinitely archived in case subsequent retracing is required.

Company G. This company checks all incoming inquiries to make sure that the person who submitted the order is listed on the official webpage of a known laboratory, public institution, or company. If no such listing exists the company takes additional steps to confirm the affiliation by contacting professors and/or other high-level personnel within the customer's organization. The company will not fill orders absent such confirmation. Additionally, confirmation may sometimes reveal that the person placing the order is a citizen of state subject to export restrictions. In such cases, the company takes additional steps to confirm that these individuals are licensed to work with the requested gene. The company

also asks customers to explain the nature of their proposed projects. Suspicious orders are then checked against Genbank and SwissProt. This comparison is done at the protein level where function is usually recognizable even for heavily optimized genes.

Significantly, all providers except Company B recognize the need for substantial human follow-up investigation to determine virulence and/or contact customers. However, the number of orders that require this effort ranges from less than 1% (Company A) to approximately 5% (Company D). Typically, companies are prepared to invest up to 10 hours per order in this process. In terms of Eqn. 1, we therefore estimate that $c_1 + c_2 = 10$ hours under current conditions. This figure depends, however, on profit margins, customer preferences, attitudes of gene synthesis industry executives, and/or formal legal requirements change that could easily change in the future.

Commercial Screening Support Services. We have seen that screening programs rely heavily on initial, automated sequence screens to detect threats and limit the need for costly follow-up investigation. Before 2007, each gene synthesis company wrote these programs for itself at a typical cost of about € 25,000. However, this figure fell significantly after CRAIC Computing began licensing its "Blackwatch" program for commercially. Blackwatch is now freely-available to any company that requests a copy and simple installations can be completed within two hours or so. However, most customers find it in their interest to customize and integrate the program with the rest of their workflow software. For this reason, the total fixed cost is about €5,000. CRAIC is currently developing a second generation program called "Safeguard" which will automate a large number of virulence judgments that are currently performed manually.

An Industry Code of Conduct. In April, 2008, members of seven leading gene synthesis companies agreed to develop an a consensus definition of responsible screening. IASB members took the lead in preparing this code of conduct and presented a first draft text to the Biological Weapons Convention's States Parties Conference in December, 2009. Its main features include:

Sequence Screening. Companies should screen against the Australia Group, Select Agent, and national export lists at both DNA and protein levels. Searches should include all known genomic and plasmid transcripts.

Customer Screening. Companies that find matches should not accept orders until they have determined the customer's identity/institutional affiliation and confirmed that it is a legitimate users. Companies should also check shipping addresses to make sure that they correspond to registered businesses, internationally-recognized academic institutions, or similarly legitimate organizations.

Records and Reporting. Companies should retain records of orders that trigger biosecurity and biosafety investigations indefinitely. Companies should

immediately inform authorities of inquiries that indicate clearly suspicious behavior (*e.g.* attempted concealment of a non-business delivery address).

IASB will hold a workshop to finalize the code in November 2009.

IX. The Way Forward

We have seen that most US and European companies practice significant screening. However, it is still not clear whether a global market will emerge and inherit this regime. Current threat lists will also have to expand over time to remain effective. We examine these challenges in turn:

Consolidating the Current US/European Regime. Policymakers can take several steps to reinforce the Universal Screening regime that currently exists in the US and Europe:

> *Jawboning.* Current screening programs are supported by the good will of gene synthesis companies and their customers. Government should reinforce this support by publicly urging ("jawboning") community members that practice responsible screening. The emerging gene synthesis industry is particularly susceptible to jawboning, which works best in concentrated industries (McConnell 1962).
>
> *Defining "Acceptable Screening."* Customers cannot enforce screening norms without a clear definition of "responsible" screening. IASB is currently developing a "code of conduct" to fill this gap. Policymakers should participate in and encourage this process. In the long run, practically all observers agree on the need for official, government-approved screening lists. [NSABB 2006; Cassagrande 2008; Fischer 2008; Garfinkel et al. 2007]
>
> *Making Markets More Transparent.* Customers need to know when a company screens responsibly. IASB plans to award a "seal of approval" to any company (including non-members) that meets its code of conduct. Companies seeking the seal would submit, *inter alia*, to anonymous testing of their screening and order procurement systems. In the long run, it may also be important for customers to know when companies do *not* screen. This can be done by testing company systems with "red team" orders for dangerous sequences. Government is the most natural provider for this kind of testing.

Establishing a Worldwide Regime. Truly Universal Screening presupposes a worldwide market. Policymakers have only limited ability to promote this goal:

> *Reforming Dual Use Review.* We have seen that gene synthesis companies are seldom willing to invest the two months and 10 to 20 hours of employee time needed to obtain dual use licenses. This has slowed their entry into the

developing world. Government should either relax these requirements or else commit sufficient resources to process these licenses within five days.

Investigating Overseas Customers. US and European gene synthesis companies find it prohibitively expensive to investigate customers in the developing world. Government can potentially fill this gap by investigating and licensing customers [Fischer (2008)] Such a system would be similar to the "Expert Traveler" lists currently found in US airports.

Improved Sequence Screening (A): Threat Lists. Extending searches beyond the select agent list will inevitably increase the number of "dual use" cases that require follow-up manual investigation. Since the industry's ability to support screening is finite, this additional burden must be "paid for" with offsetting savings. In the near term, the most promising way to do this is to automate virulence searches that are currently performed by hand. This, however, will require a comprehensive database of threats and/or virulent genes.

Building this database will require approximately $2 million of effort. However, some of these resources already exist. In particular, we have already seen that gene synthesis companies already spend up to five hours per week researching virulence. Furthermore, companies report that they have seen 3-5% of all select agent genes before. This suggests that a consortium that pooled data from n companies could reduce their search burden by

$$S = 1 - [0.95]^n \qquad \text{(Eqn. 2)}$$

where S is the percentage savings.[35] Furthermore, our companies have recently seen a noticeable increase in orders requesting the same genes from pathogenic organisms. This suggests that, if anything, the benefits of pooling data are growing over time.

Berkeley and IASB have already announced plans to launch an open-source style virulence database (VIREP) in early 2010. This effort will be further bolstered by (a) software that uses VIREP as a "training set" to perform automated searches of the virulence literature, and (b) an on-line Forum where members can share experiences from their respective screening programs. [Maurer 2009] At the same time, VIREP cannot collect data faster than the rate at which current gene synthesis companies screen orders. If government decides that faster progress is needed, it will have to provide some form of grant support.

Improved Sequence Screening (B): Advanced Software. The case for sharing advanced software is more ambiguous. Economically, gene synthesis companies can pool their development efforts *either* (a) by paying royalties to commercial companies that develop screening software for many customers, or (b) open source user collaborations. Both

[35] The fact that S increases rapidly for small n is encouraging since it suggests that collaborations are worthwhile even when the number of participants is small.

methods suffer from inefficiencies[36] and we expect individual companies to adopt whichever solution is most cost-effective under the circumstances. This may not, however, be socially efficient if high royalty prices deter even one company from adopting the software. CRAIC Computing's recent decision to make its industry-standard Blackwatch software freely available as open source has resolved this dilemma for current, first-generation screening programs. Government or trade association support for second generation commercial software projects should similarly include "access pricing" terms to ensure affordability.[37]

Improved Customer Screening. In the long run, broader threat searches will almost certainly require more efficient customer screening as well. This will mean automating steps that are almost always performed manually today. Probably the best strategy would be to have a third party to operate a web site where trusted customers could be verified using Internet certificates similar to the current "Verisign" system. [Garfinkel et al. 2007] Trusted status, in turn, would depend on pooled data about the customer. This could include information that the customer (a) had purchased large numbers of benign genes, (b) had purchased "dual use" genes from a gene synthesis company known to practice responsible screening, and/or (c) had received Select Agent licenses, dual use certificates, or similar approvals from one or more governments in the past.

Beyond The Gene Synthesis Industry. Finally, nothing the gene synthesis industry does can prevent terrorists from purchasing synthetic DNA from other sources. It therefore makes sense to encourage small commercial providers that provide gene synthesis services as a sideline to adopt the same sequence screening tools that large companies use. At the same time, solutions that work in our industry should not be blindly applied elsewhere. Given that small companies almost always know their customers personally, industry initiatives that focus on encouraging companies to report suspicious contacts could yield substantial benefits.

Finally, we have seen that current sequence screening technologies are impractical for oligos. On the other hand, many researchers think that second generation software may permit screening down to just 50 nt. At that point, it should be feasible to organize the large oligo companies – which are already highly concentrated and serve global markets – around a Universal Screening regime.

---

[36] For example, economic theory suggests that proprietary pricing only extracts about 50% of the revenues that consumers would be theoretically willing to pay for a particular product. This means that copyright incentives cannot support every R&D project that should theoretically be funded. On the other hand, open source output is often suppressed by game theory considerations in which each player strategically waits for someone else to contribute value. [Maurer and Scotchmer 2006]

[37] The concept of "openness" would also have to be carefully rethought in a biosecurity collaboration. For example, Blackwatch reports that some users have tested its on-line software to see which inquiries are detected. While these incidents probably reflect innocent curiosity, a system that let terrorists perform similar "dry runs" would not be acceptable. At the same time, many researchers outside the gene synthesis community are interested in virulence data. For this reason, any secrecy provisions should probably allow legitimate researchers to request access to VIREP.

Sustaining Confidentiality. All government efforts to improve biosecurity within the gene synthesis industry should respect customers' and providers' strong desire for confidentiality. Companies and government agencies must protect exchanges of information about orders and customers using the same high standards that communications within companies do today. In some cases, government-supplied infrastructure could even improve the protection that gene synthesis orders receive.[38] Such improvements would significantly increase customers' willingness to accept appropriate information-sharing

X. Conclusion

The emergence of the commercial gene synthesis industry has made genetic engineering cheaper and more convenient than ever before. Any reasonable security policy must recognize and preserve these benefits to society. At the same time, it is worth remembering how the Soviet Union took advantage of 1970s-era commercial biotechnology to build the world's most advanced biological weapons. So far, we have absolutely no evidence that terrorists – whose budgets are roughly 1,000 times smaller than the old Soviet programs – are pursuing similar ambitions. At the same time, the risk is not zero and the potential consequences are large. For this reason, steps that materially increase the costs or risks of a hypothetical terrorist program are generally worth considering.

The industry has made a good start. Following early discussions, there is now broad agreement among US and European companies and their principal customers that screening is both necessary and desirable. Furthermore, the private sector has shown that it is prepared to invest significant manpower to investigate orders. The challenge over the next few years will be to reinforce the current screening regime and extend it to embrace a truly global market. Prudent government "jawboning" participants to encourage good behavior and selective investment in those cases where industry resources are inadequate. Above all, government should be careful not to impose so many unfunded mandates that the current voluntary system collapses.

Even assuming that current screening practices are adequate, this will not be true for all time. The terrorism threat is bound to increase and will eventually force screening programs to extend their searches beyond the select agent list. Here the key will be developing improved lists of threat genes. We have argued that the industry's open source-style VIREP initiative will go some distance toward meeting this goal. However, faster progress will require government support. This investment will ultimately pay dividends not just for biosecurity but also for basic research aimed understanding pathogenicity. A similar list of approved customers will be more controversial but is also potentially worth pursuing.

[38] For instance, government-operated computers could provide suitably-encrypted online screening support to any company that needed it. The service would be particularly appealing to small companies that provide gene synthesis as a side line and cannot afford to build secure IT infrastructures in-house.

References

Anon. "Editorial: The Peril of Genes for Sale." *New Scientist* (Nov. 12, 2005).

AstraZeneca. 2008. "Code of Conduct." http://www.astrazeneca.com/responsibility/code-and-policies/?itemId=5716829

Bundesamt für Wirtschaft und Ausfuhrkontrolle ("BAFA"). 2007. "Export Controls – Brief Outline." http://www.bafa.de/bafa/en/export_control/publications/export_control_brief_outline.pdf.

Baker, David, George Church, Jim Collins, Drew Endy, Joseph Jacobson, Jay Keasling, Paul Modrich, Christina Smolke, and Ron Weiss. 2006. "Engineering Life: Building a Fab for Biology." *Scientific American* (June).

Bousted, G. 2008. "The Biohacking Hobbyist," *Seed Magazine* . Dec. 11. http://seedmagazine.com/content/article/the_biohacking_hobbyist/

Van den Brulle, J, M. Fischer, T. Langmann, G. Horn, T. Waldmann, S. Arnold, M. Fuhrmann, O. Schatz, T. O'Connell, D. O'Connell, A. Auckenthaler, and H. Schwer. 2008. "A novel solid phase technology for high-throughput gene synthesis." *Biotechniques.*45(3):340-3.

Carlson, Robert. 2008. "Tracking the Spread of Biological Technologies." *Bulletin of Atomic Scientists* (Nov. 21)

Cassangrande, Rocco. Remarks in US Department of Health and Human Services, "US Government Roundtable on Development of a Synthetic Nucleic Acid Screening Framework," (2008).

Cello, Jeronimo, Aniko V. Paul, and Eckard Wimmer. 2002. **"**Chemical Synthesis of Poliovirus cDNA: Generation of Infectious Virus in the Absence of Natural Template." *Science* 297: 1016 – 1018.

DIYBio Home Page, http://diybio.org/

Dando, Malcolm. 2001. *The New Biological Weapons* (Boulder CO: Lynne Rienner)

Fischer, Julie. Remarks in US Department of Health and Human Services, "US Government Roundtable on Development of a Synthetic Nucleic Acid Screening Framework," (2008).

Friedman, Robert. Remarks in US Department of Health and Human Services, "US Government Roundtable on Development of a Synthetic Nucleic Acid Screening Framework," (2008).

Garfinkel, Michele, Drew Endy, Gerald Epstein and Robert Friedman. 2007. "Synthetic Genomics: Options for Governance." http://www.jcvi.org/cms/fileadmin/site/research/projects/synthetic-genomics-report/synthetic-genomics-report.pdf

Gibson, Daniel et al. 2008. "Complete Chemical Synthesis, Assembly, and Cloning of a Mycoplasma genitalium Genome." 2008. *Science* **319**, 1215

Gryphon Scientific. 2008. "Sequence Screening." (Presentation Slides for US Govenrment Roundtable on Development of a Synthetic Nucleic Acid Screening Network"; Bethesda MD., Sept. 12).

Hacker J, Kaper JB. 2000. "Pathogenicity islands and the evolution of microbes." *Annual Review of Microbiology* 54:641-79

Imperiale, Michael, in US Department of Health and Human Services, "US Government Roundtable on Development of a Synthetic Nucleic Acid Screening Framework," (2008).

International Association – Synthetic Biology. 2008. "IASB (Draft) Code of Conduct." (Nov. 24)

Khorana, Har G. 1979. "Total Synthesis of a Gene." *Science 203*: 614-625.

National Science Advisory Board for Biosecurity, "Addressing Biosecurity Concerns Related to the Synthesis of Select Agents." (December 2006). http://oba.od.nih.gov/biosecurity/biosecurity_documents.html

Maurer, Stephen. "Grassroots Efforts to Impede Bioterrorism," *Bulletin of the Atomic Scientists* (March 5, 2009).

Maurer, Stephen, "Technologies of Evil: Chemical, Biological, Radiological, and Nuclear Weapons" in S. Maurer (editor) *WMD Terrorism: Science and Policy Choices* at pp. 47-110. MIT Press: 2009.

Maurer, Stephen and George Rutherford, "The New Bioweapons: Infectious and Engineered Diseases," in S Maurer (editor) *WMD Terrorism: Science and Policy Choices* at pp. 111-138. MIT Press: 2009.

Maurer, Stephen and Suzanne Scotchmer, "Open Source Software: The New Intellectual Property Paradigm," in T. Hendershott (ed.), *Handbook on Information Systems* (Elsevier: 2006).

McConnell, Grant. 1963. *Steel and the Presidency – 1962* (WW Norton: New York)

Randerson, James. "Lax Laws, Virus DNA and the Potential for Terror." *Guardian* (June 14, 2006)

Sambrook, Joseph and David S. Russell *Molecular Cloning* (3rd ed. 2001).

Tumpey, Terrence M., Chirstopher F. Basler, Patricia V. Aguilar, Hui Zeng, Alicia Solorzano, David E. Swayne, Nancy J. Cox et al. 2005. "Characterization of the Reconstructed 1918 Spanish Influenza Pandemic Virus." *Science* 310: 77-80.

McKenna, Phil. 2009. "Rise of the Garage Biohackers," *New Scientist* (January 7) http://www.newscientist.com/article/mg20126881.400-genetic-manpulation-now-becoming-a-hobby.html?full=true

Make Magazine, http://www.makezine.com/.

Minshull, Jeremy. 2009. "Does Offshoring Put Patents in Jeopardy?" *Genetic Engineering and Biotechnology News*. http://www.genengnews.com/articles/chitem.aspx?aid=2798.

National Science Advisory Board for Biosecurity. 2007. "Proposed Framework for Oversight of Dual Use Life Sciences Research: Strategies for Minimizing the Potential Misuse of Research Information." http://oba.od.nih.gov/biosecurity/pdf/Framework%20for%20transmittal%200807_Sept07.pdf

Zaremba S, Ramos-Santacruz M, Hampton T, Shetty P, Fedorko J, Whitmore J, Greene JM, Perna NT, Glasner JD, Plunket G, Shaker M, Pot D. 2009. "Text-mining of PubMed abstracts by natural language processing to create a public knowledge base on molecular mechanisms of bacterial enteropathogens." *BMC Bioinformatics*10:177.