TWIST BIOSCIENCE

# Synthetic gene-length DNA: Evolving export control concerns

James Diggans, PhD
Director, Bioinformatics and Biosecurity

July • 2019

# Disclaimer

*Dr. Diggans is Director of Bioinformatics and Biosecurity at Twist Bioscience in San Francisco, California. He is also Twist's representative to the Board of the International Gene Synthesis Consortium (IGSC). Dr. Diggans is presenting today in his capacity as a Twist Bioscience employee and not as an IGSC board member. Accordingly, the views he expresses do not necessarily reflect the views of the IGSC.*

TWIST BIOSCIENCE

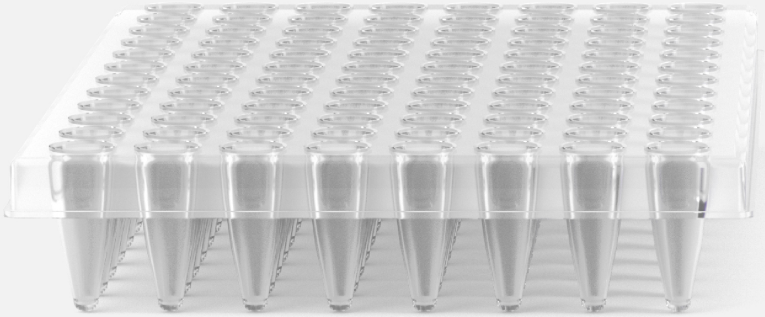
**Founded 2013**

**IPO in October, 2018**

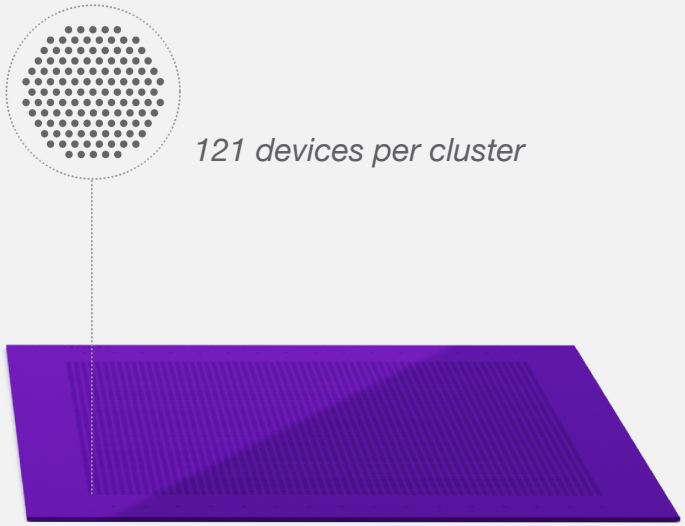
**~350 global employees**

- San Francisco
- South San Francisco
- Tel Aviv
- San Diego
- Singapore

# Rewriting DNA with the Power of Silicon



**96 WELL PLATE**
*makes 1 gene*

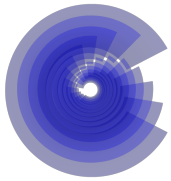



**TWIST SILICON PLATFORM**
*makes 9,600 genes*

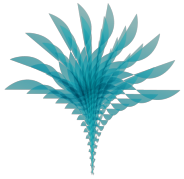
Developing Game-Changing Throughput and
Cost through Quality and Speed at Scale

# Twist Products

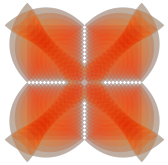
Precision DNA Synthesis at Scale



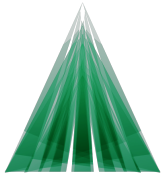
Genes



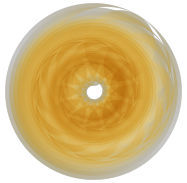
Oligo Pools



Libraries



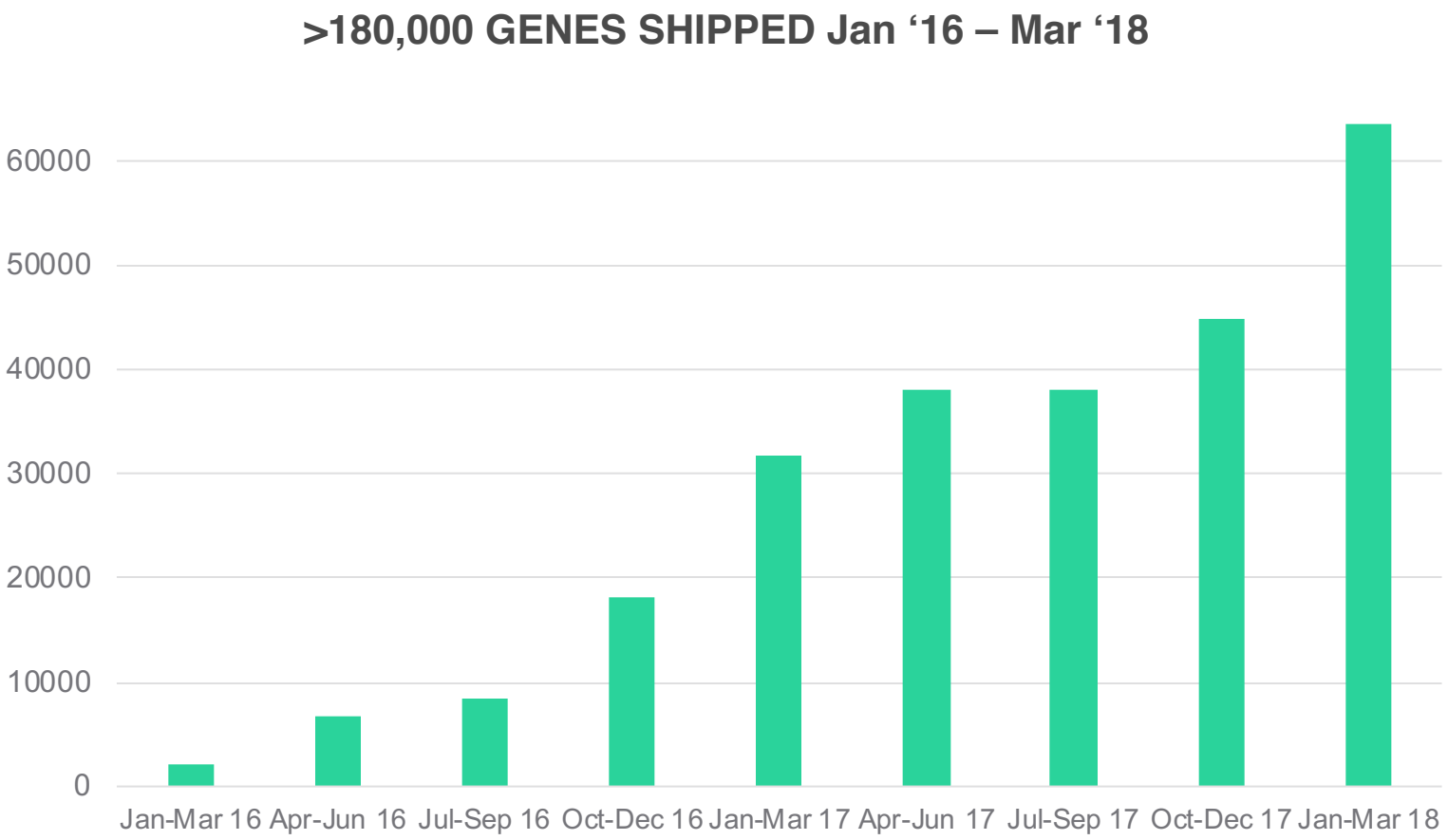
NGS



Data Storage

**Precision DNA Synthesis at Scale**

# Genes: Scale at Twist




| Apr – Jun 2016 | Jan – Mar 2017 | Oct – Dec 2017 | Jan – Mar 2018 |
|---|---|---|---|
| **2,000 genes per month** | **10,000 genes per month** | **15,000 genes per month** | **>20,000 genes per month** |

# Scale: horizontal and vertical

**Twist Bioscience to Provide One Billion Base Pairs of Synthetic DNA to Ginkgo Bioworks to Support Expansion into New Industries**

***Ginkgo Bioworks and Twist Bioscience Continue to Drive Industry Growth***

# Export Control at Twist

# When does a sequence require an export license?

U.S. Export Administration Regulations (EAR)

- Controlled human and animal pathogens and toxins are listed in 1C351; controlled plant pathogens are listed in 1C354.

- DNA sequences from controlled items in 1C351 and 1C354 may require licensing if they meet criteria defined in ECCN 1C353.
  - All genes specific to controlled viruses require a license.
  - Bacterial genes that do not '**endow or enhance**' pathogenicity or are not unique to controlled species do not require a license.

a.1. Any gene or genes specific to any virus controlled by 1C351.a or .b or 1C354.c;

a.2. Any gene or genes specific to any bacterium controlled by 1C351.c or 1C354.a, or any fungus controlled by 1C351.e or 1C354.b, and which:

a.2.a. In itself or through its transcribed or translated products represents a significant hazard to human, animal or plant health; *or*

a.2.b. Could endow or enhance pathogenicity; *or*

a.3. Any toxins, or their subunits, controlled by 1C351.d.

https://www.bis.doc.gov/index.php/regulations/commerce-control-list-ccl

# The Australia Group (AG)

U.S. export control for DNA sequence is identical to that of 43 other nations

The AG harmonized national export control for dual-use chemical and biological materials to promote cooperation and uniformity internationally.

- Founded in 1985 after use of chemical weapons in the Iran/Iraq war
- Participant nations implement AG control lists for export licensing
- Currently 43 participant nations including the U.S. and EU, Canada, India, Korea and Japan. China is not currently a participant nation.
- U.S. EAR is updated to reflect changes to AG control lists and language

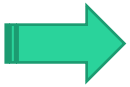
Australia Group

**Genetic Elements and Genetically-modified Organisms:**

Any genetically-modified organism[1] which contains, or genetic element[2] that codes for:

1. any gene or genes specific to any listed virus; or
2. any gene or genes specific to any listed bacterium[3] or fungus, and which
   a. in itself or through its transcribed or translated products represents a significant hazard to human, animal or plant health, or
   b. could endow or enhance pathogenicity[4]; or
3. any listed toxins or their sub-units.

https://australiagroup.net/en/participants.html



U.S. EAR CCL 1C353

a.1. Any gene or genes specific to any virus controlled by 1C351.a or .b or 1C354.c;

a.2. Any gene or genes specific to any bacterium controlled by 1C351.c or 1C354.a, or any fungus controlled by 1C351.e or 1C354.b, and which:

a.2.a. In itself or through its transcribed or translated products represents a significant hazard to human, animal or plant health; *or*

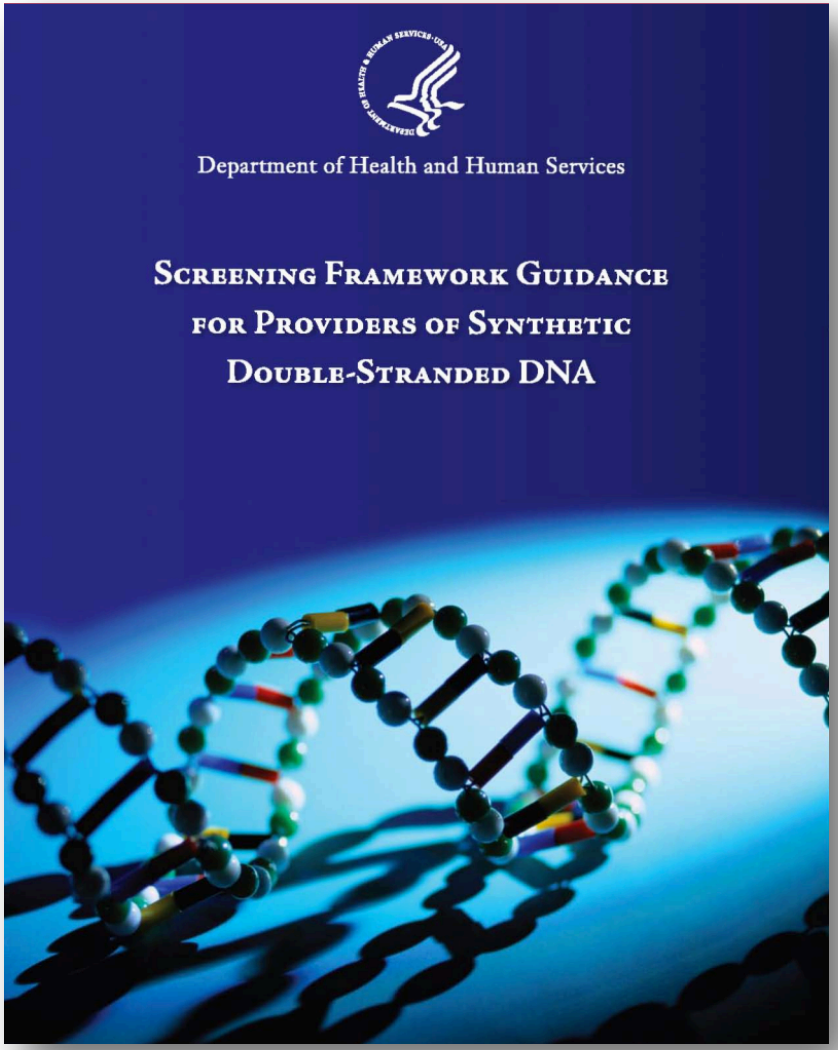
a.2.b. Could endow or enhance pathogenicity; *or*

a.3. Any toxins, or their subunits, controlled by 1C351.d.

# 2010 HHS Guidance

In 2010, the U.S. government published guidance to synthetic DNA manufacturers that defines best practices for:

- Customer screening
- Sequence screening
- Customer ‘red flag’ follow up
- Records retention




https://www.phe.gov/Preparedness/legal/guidance/syndna/Pages/default.aspx

# International Gene Synthesis Consortium Members

International
**Gene Synthesis**
Consortium

| | |
|---|---|
| ATUM | GenScript |
| Battelle | The Edinburgh Genome Foundry |
| BGI | IDT |
| Bioneer | SGI-DNA |
| Blue Heron | Thermo Fisher |
| Ginkgo Bioworks | Twist Bioscience |

https://genesynthesisconsortium.org/

# Biosecurity at Twist

## Ensuring safe, secure DNA synthesis

Upon submission to Twist, all gene synthesis orders are subject to:

- Customer screening
  - Verifies legitimacy with respect to an institution and identity with respect to an individual
  - Is the customer or institution on any lists maintained by the Departments of Commerce, State and Treasury?
  - Is the customer licensed to carry out work on regulated pathogens?
  - Ensure the shipping address is not a PO Box or private residence

- Sequence screening
  - Is the sequence from a regulated bacterial or viral pathogen?
  - Is the sequence from a gene that can endow or enhance pathogenicity?
  - Is the sequence legal to manufacture and ship within the United States?
  - Does the sequence require an export license to ship overseas?

These are industry best practices as codified in U.S. government guidance and the IGSC Harmonized Screening Protocol.

# Recent Challenges in Export Control

# 2010 Guidance - Challenges

## Customer screening

- Relies on name matching to government lists of proscribed entities
- Can be difficult internationally to resolve addresses as non-residential
- Defining the end user can be complicated

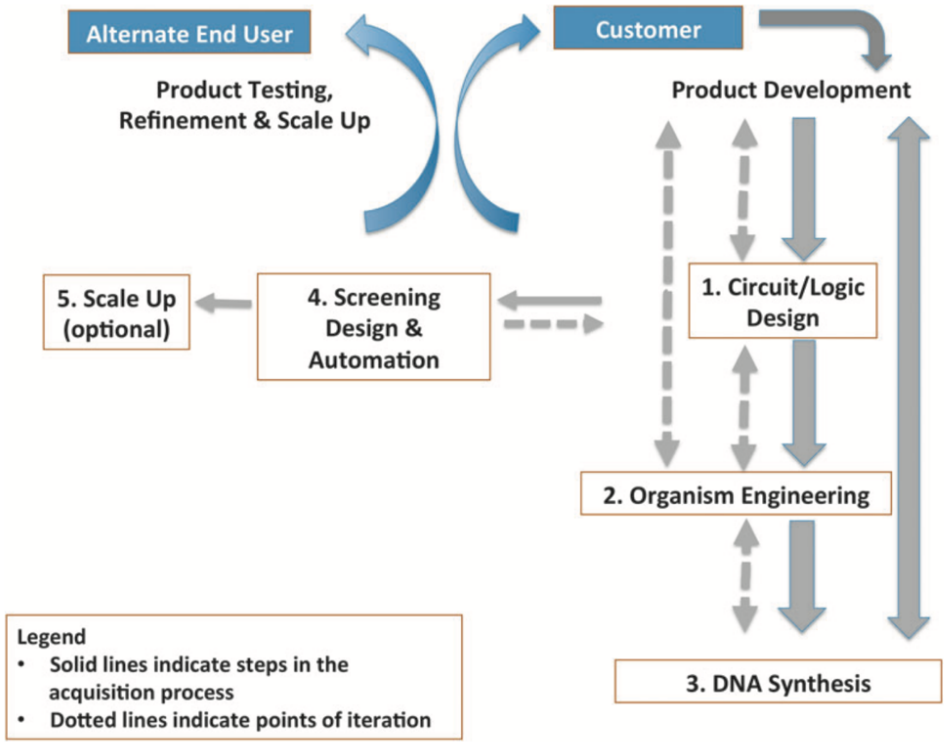


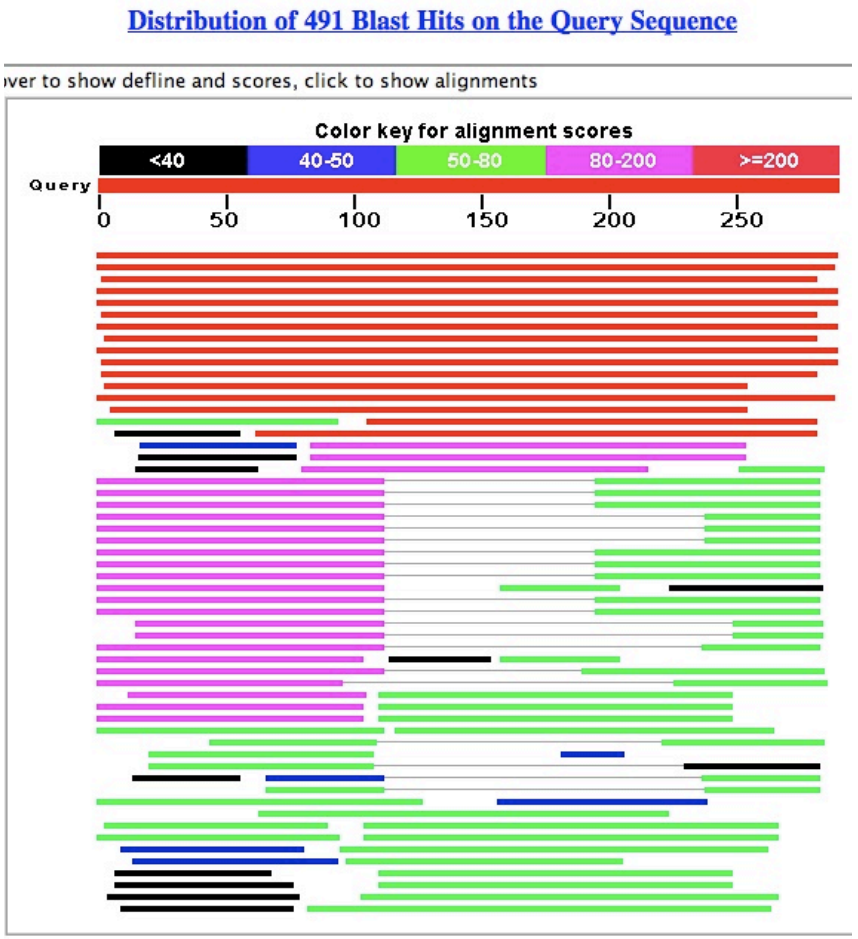

Figure 2. Synthetic Biology Supply Chain

https://www.liebertpub.com/doi/abs/10.1089/hs.2016.0083

# 2010 Guidance - Challenges

## Sequence screening

- Mapping/maintenance of species names to NCBI taxon IDs
- What about fragments intended for assembly?
  - E.g. the entire ebola genome in overlapping, non-ORF fragments.
  - Current practice: send letter re: re-export license requirement after assembly
- How do we estimate the performance of screening implementations?



# 2010 Guidance - Challenges

## Export control

- ‘Endow or enhance’ pathogenicity
  - **Much improved over ‘associated with pathogenicity’**
- Mapping/maintenance of species names to NCBI taxon IDs
  - USG lists could use the taxon ID as the unit of control to remove ambiguity
- Biologically-centered control language e.g. avian influenza / Newcastle disease
- Significance of open reading frame
  - E.g. ‘capable of encoding a protein’

a.4. Avian influenza (AI) viruses identified as having high pathogenicity (HP), as follows:

a.4.a. AI viruses that have an intravenous pathogenicity index (IVPI) in 6-week-old chickens greater than 1.2; *or*

a.4.b. AI viruses that cause at least 75% mortality in 4- to 8-week-old chickens infected intravenously.

A virulent Newcastle disease virus (avian paramyxovirus serotype 1) has an intracerebral pathogenicity index in day-old chicks (Gallus gallus) of 0.7 or greater or has an amino acid sequence at the fusion (F) protein cleavage site that is consistent with virulent strains of Newcastle disease virus. A failure to detect a cleavage site that is consistent with virulent strains does not confirm the absence of a virulent virus.

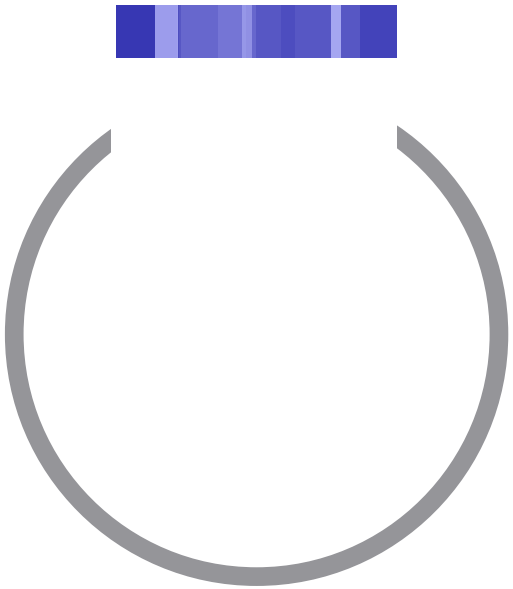
# ORF Definitional Challenges

When is a sequence controlled?

- Full open reading frame (ORF) is controlled – so …
  - Customers ordered ORF - 1 amino acid or -1 base
  - Customers ordered shortened ORF but onboarded a custom vector with the remainder of the ORF in the vector homology arms
  - Customers iteratively order shortened ORFs until no trigger occurs

AG controls are (in a round-about way) targeted at *function*

- 49% of a controlled ORF can still preserve a functional domain
- Challenge: How can a synthesis company make rapid determinations of preservation of function?



# Turn around time (TAT)

To compete, we must deliver DNA quickly

- DNA synthesis companies compete on TAT
  - Scientists want to carry out their experiments immediately
  - Time spent on synthesis is a barely-tolerated delay

- Twist TAT for non-clonal DNA is ~6-9 days; for clonal DNA, it is ~11-17 days
  - plus time spent for international shipping, clearing customs, etc.

- BIS suggested remedy for uncertainty around license requirement is a classification request
  - BIS turns these around heroically – usually within two weeks!
  - That still nearly *doubles* our TAT – customers usually use domestic (i.e. non-US) providers instead

# Twist Innovation

# Red Teaming

Estimating performance of complex systems

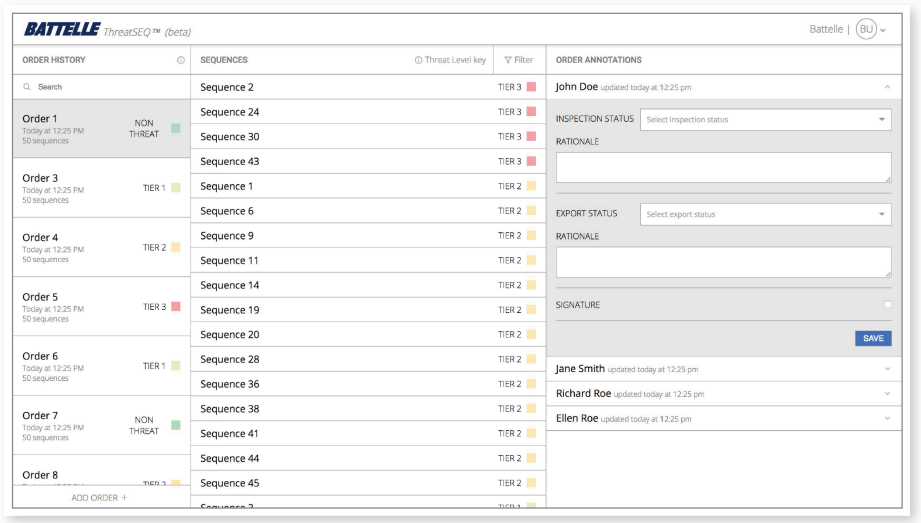
- IGSC onboarding process includes test data analysis
- Twist hired a third party to attempt to subvert our screening process
- We learned *a lot* by doing this
- Key findings
  - blastx alignments vs. NCBI's nr database were sensitive
  - **Sequence annotation quality and coverage was the primary weakness**

# ThreatSeq Adoption

## High-quality, curated metadata drives screening performance

- Twist incorporating Battelle ThreatSeq into our biosecurity screening backend

- Advantages
  - Reduced false positive hit rate
  - Immediate contextualization of hit to known virulence factors
  - Maintained and expanded over time by a team of curators

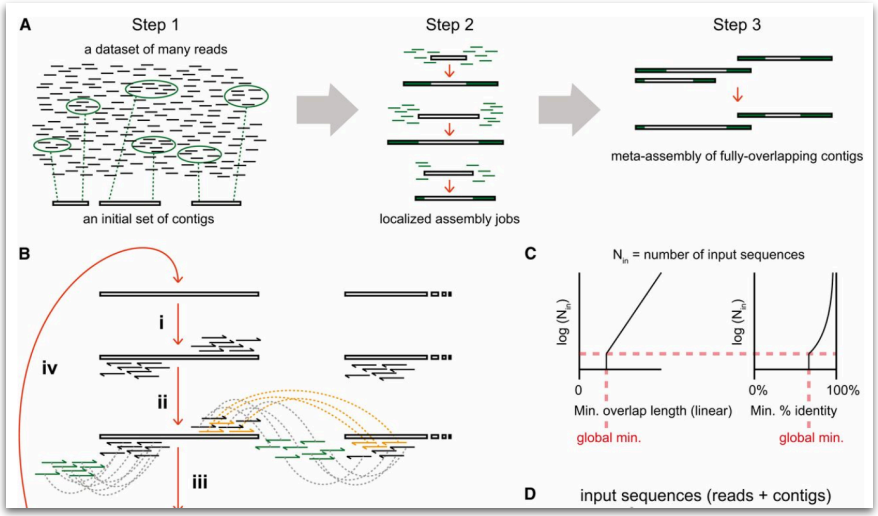



https://www.battelle.org/commercial-offerings/industry-solutions/threatseq-dna-screening-web-service

# Screening oligos and oligo pools

NGS assembly routinely analyzes millions of oligo-length sequences

- De novo assembly of 100M+ reads is common/tractable in metagenomics
- 100M oligos is … a big oligo pool.




http://www.g3journal.org/content/3/5/865

- Twist currently prototyping methods for recovering intended gene assemblies from large oligo pools
  - Wheat vs. chaff
  - Multiplex assemblies
  - Varied assembly strategies

# Changing Scientific Landscape

# Emerging Challenges

https://cen.acs.org/biological-chemistry/synthetic-biology/Synthetic-biology-enable-bioweapons-development/96/i26

*“Making ineffective biological weapons is easy, making effective biological weapons is not easy.”*

– Margaret Kosal, GA Tech

Addressing emerging challenges

- ‘Species’ is a less and less useful concept
- Protein design is ever more capable and powerful
- Oligo pools and easy gene assembly protocols
- Data storage in DNA
- S&T efforts – e.g. IARPA FunGCAT
- The synthetic biology supply chain is complex

# Recent Twist paper on next steps

- Expand customer screening recommendation across synbio supply chain
- Regular red teaming to estimate system performance
- Screening of oligos and oligo pools via NGS-derived methods
  - Applies both within- and between orders
  - Subject assembly results to gene-length sequence screening; follow-up screening on hits
- Additional S&T investment:
  - Expanded FunGCAT-like tooling for screening including open source / publicly available screening tools
  - Homomorphic encryption methods for centralization of screening results


https://www.frontiersin.org/articles/10.3389/fbioe.2019.00086/abstract

# How can Commerce help US companies compete?

- Continued Australia Group advocacy – export controls drive a need for innovation
  - Evolution of the unit of control: Species -> protein sequences -> function?
  - Control of oligo pools that can be used for gene assembly?
  - Control of fragments trivially assembled into controlled genes?

- API for SNAP-R request submission
  - Would allow company systems to submit classification and licensing requests directly without web interface point-and-click
  - Reduce cost and speed TAT – allow for more classification requests

# Questions?

jdiggans@twistbioscience.com

TWIST BIOSCIENCE