ThreatSEQ™: An Advanced DNA Screening Platform for Gene Synthesis Biosecurity

Helping the synthetic biology community perform research responsibly

- Battelle's ThreatSEQ™ web-based analytical software tool and service provides a powerful new approach to reduce the potential for unintended production of DNA sequences of concern.
- Built on more than a decade of research, ThreatSEQ quickly and accurately detects DNA sequences of concern and provides this information to the user in a concise report.
- ThreatSEQ analysis enables biotechnology companies to make faster, more informed decisions to protect themselves, their clients and society from potential biosecurity risks.

BATTELLE

HOW IT WORKS

ThreatSEQ detects and classifies sequences of concern in genomic data. The foundation of the platform is four databases, including a proprietary highly-curated sequence of concern database. These databases are coupled to an aligner, and all alignment results are processed for hits to databases consistent with current HHS guidelines.[1] The hit profiles across the databases are subsequently processed for a threat ranking that is displayed to the end user through a graphical user interface.

Sequence of Concern Database

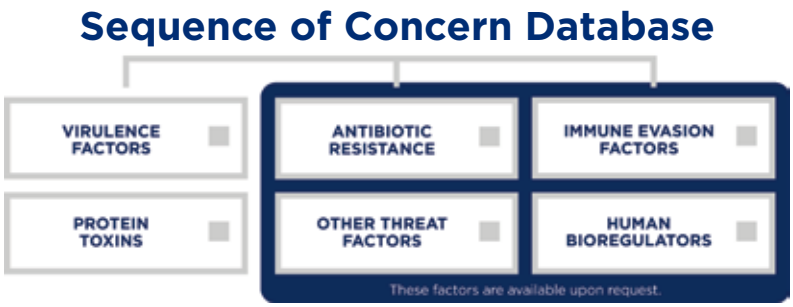
The gene-level database was built on more than a decade of research into the factors that make pathogens dangerous. It compiles more than 10,000 sequences of concern comprising 850 sequence types of concern from 75 species of bacteria, 96 viruses, 12 eukaryotic pathogens and other contributors to pathogenesis.

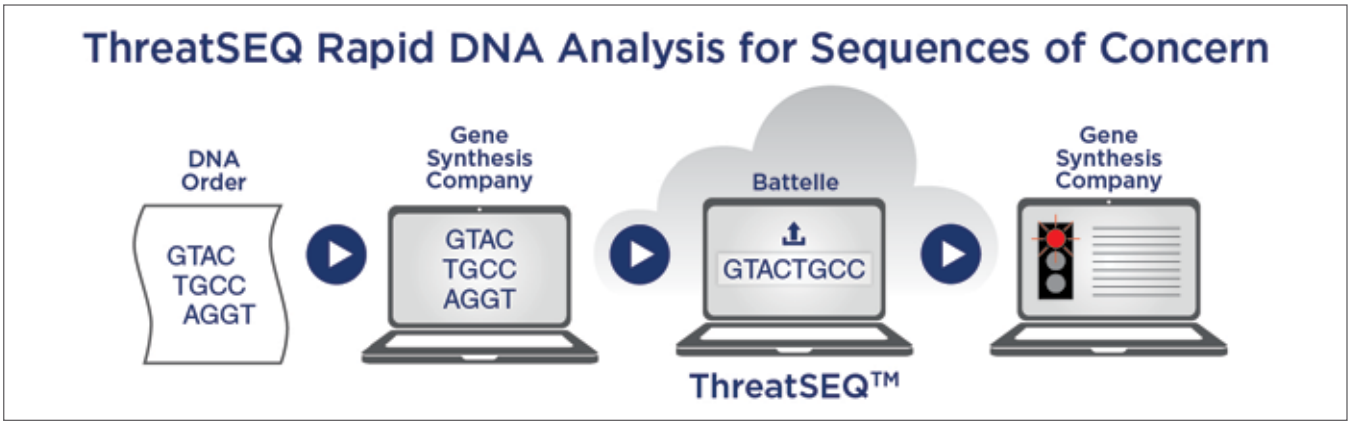
The database covers 100% of human U.S. Select Agents and Australia Group Lists (Tier 1) and virtually all known bacterial human/zoonotic pathogens.[2, 3] To ensure that the most comprehensive comparison is performed, the user also is provided with results against the full genomes of organisms derived from the National Center for Biotechnology Information (NCBI) database and from select agent registries around the world.

Threat Identification Algorithm

The algorithm rapidly scans a requested DNA sequence and looks for matches within the databases. Matches may be entire gene sequences or parts of gene sequences. ThreatSEQ has been designed to operate efficiently and to detect complex patterns (e.g., tiled mosaic DNA fragments).

In addition to identifying positive matches, the ThreatSEQ algorithm provides information about the types of threats that a gene sequence may present by providing information about the structures of the genes and their resulting functions. The existing analysis is supported by more than 4,400 peer-reviewed papers, and Battelle continues to update the database and algorithm as new research emerges.





[1]https://www.phe.gov/Preparedness/legal/guidance/syndna/Documents/syndna-guidance.pdf.
[2]Federal Select Agent Program 2017. Select agents and toxins list. www.selectagents.gov/SelectAgentsandToxinsList.html. Division of Select Agents and Toxins, Centers for Disease Control and Prevention, Atlanta, GA.
[3]The Australia Group 2017. Australia Group common control lists. www.australiagroup.net/en/controllists.html.

THREAT REPORTING

ThreatSEQ provides an easy-to-understand report that flags each identified sequence of concern and characterizes the potential hazard associated with the sequence. The report provides the reviewer with actionable information to easily determine the threat level of a sequence and to understand the rationale behind the threat determination.

Threats are ranked for more efficient review:

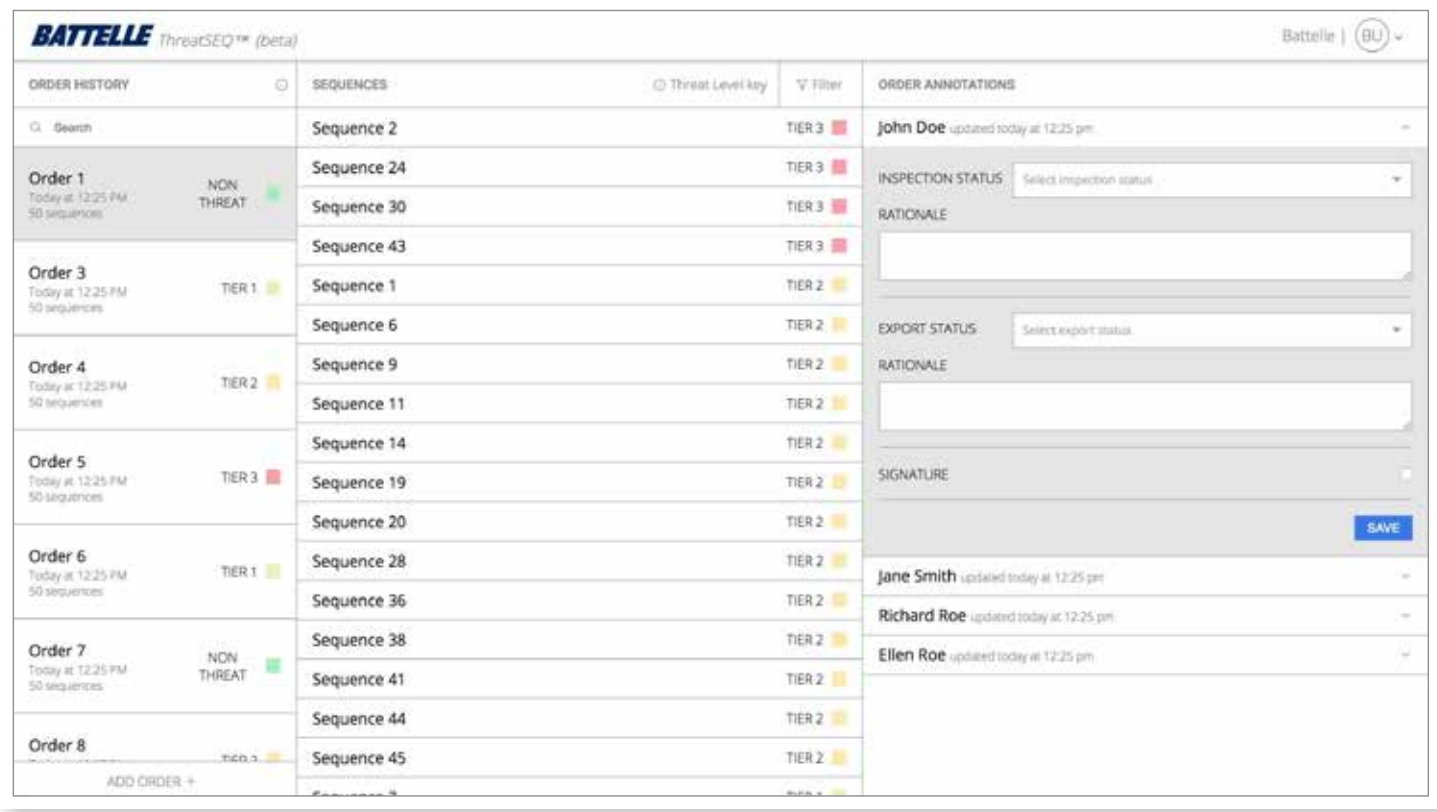
TIER 1 = matches a known sequence of concern or Tier 1 virus

TIER 2 = matches any sequence in a non-annotated select agent's genome

TIER 3 = matches a non-concerning sequence in an annotated select agent's genome

NON-THREAT = matches a sequence in a non-select agent's genome

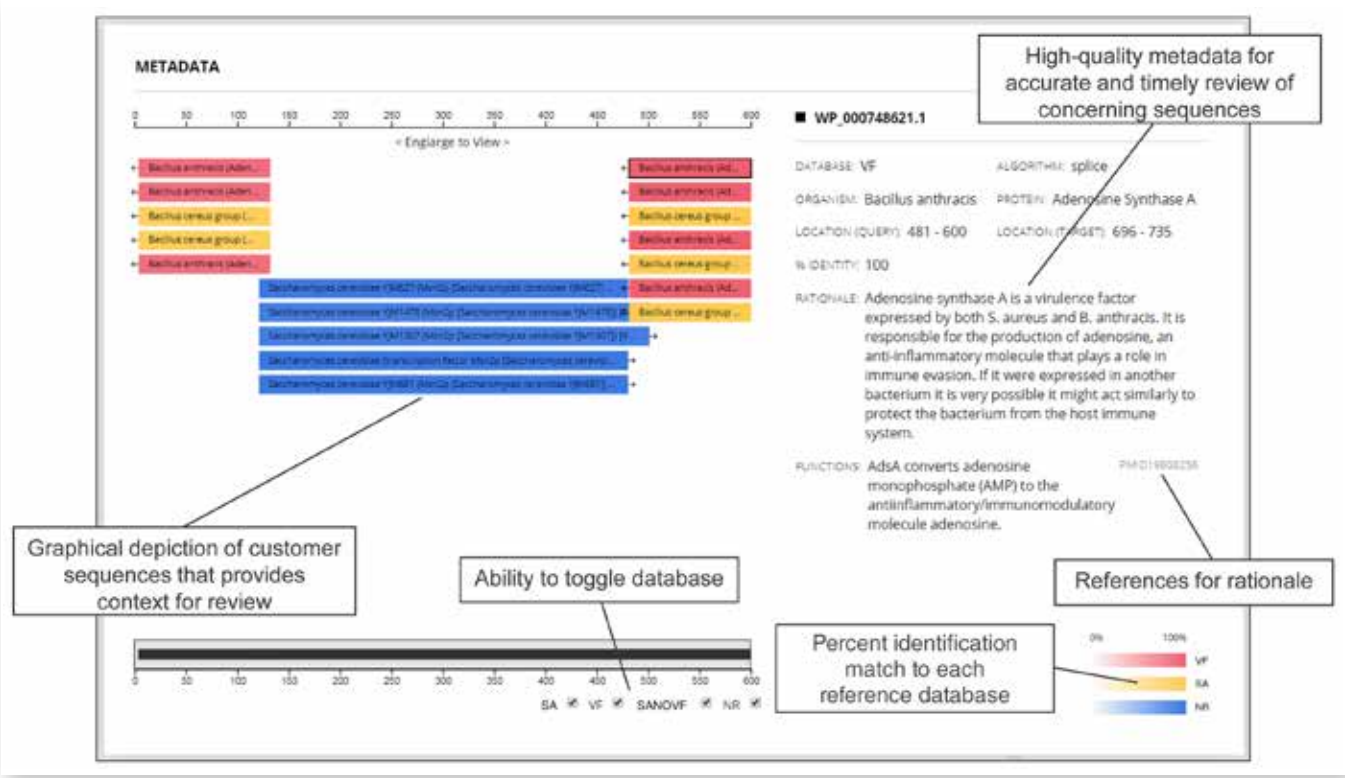
UNIDENTIFIED = does not match to any known database



Reviewers can run reports of order history and threat status for easy tracking and auditing of screening results.

FEATURE OVERVIEW

ThreatSEQ threat assessment algorithms have been designed to detect sequences of concern that are in either chimeric or mosaic form. The platform provides the end user with an advanced graphical user interface to review the results for the submitted sequences, which includes high-quality metadata and a context mapping tool that allows the user to see how alignments pair to the query sequence.



FEATURES	THREAT SEQ	Current Practice
High-Security	✓	✓
Company-Specific	✓	✓
Virulence Factor Focused	✓	--
Provides Linked Metadata	✓	--
Continuous Updating	✓	--
Standardized and Objective Threat Status Score	✓	--
Complex Pattern Identification	✓	--
Comprehensive Reporting and Interpretation	✓	--
High Accuracy	✓	--
Web-deployed	✓	--
Focused Input	✓	--

ThreatSEQ delivers:

Periodically updated database: established pipeline for routine database updates.

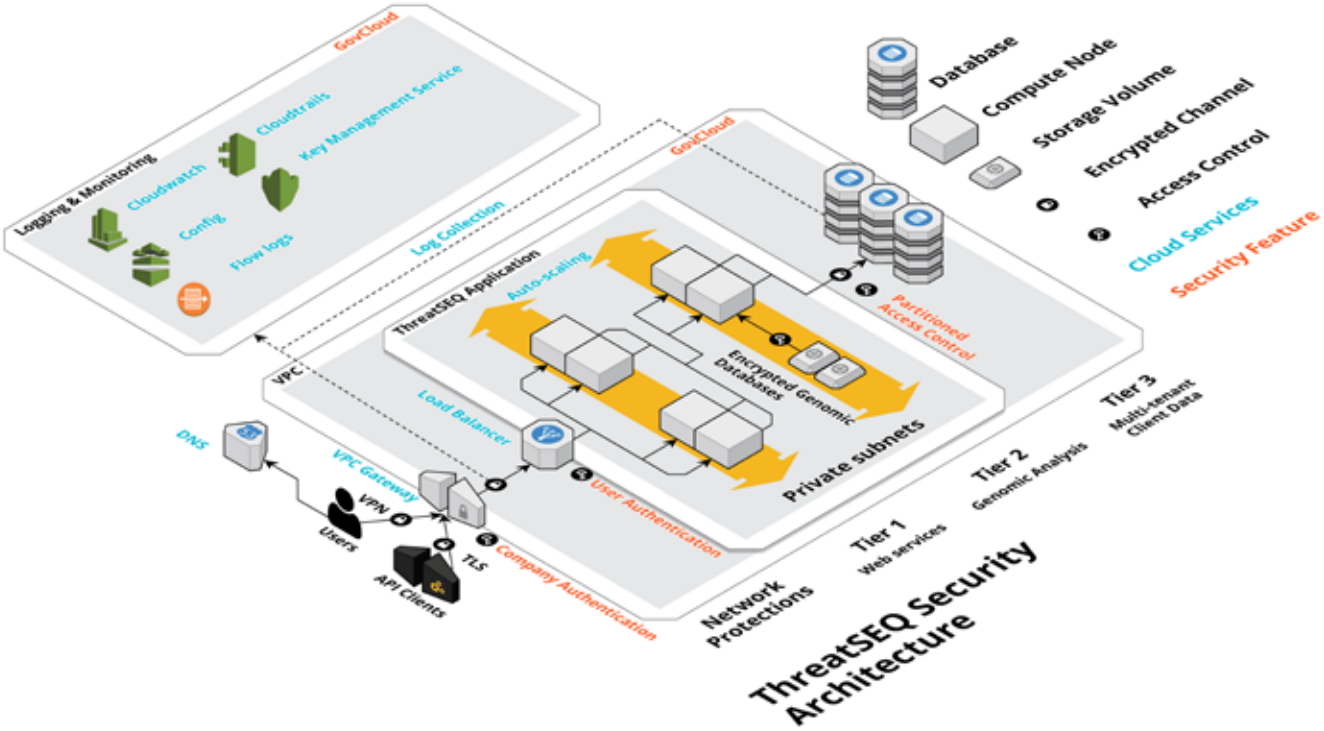
Efficiency: highly-curated database and custom algorithms reduce false positives compared to current screening methods

Accessibility: secure cloud-deployed solution with automated sequence-specific threat detection significantly reduces required infrastructure and SME labor

Reduces customer follow-up time: consolidated reporting and links to original data sources and rationale

HIGH-SECURITY SYSTEM ARCHITECTURE

ThreatSEQ has been designed to protect the confidentiality and integrity of client data and has been implemented in accordance with the ISO/IEC 27000 series of standards for cybersecurity and information security. The system utilizes a cloud-based architecture hosted on Amazon Web Services (AWS). The Amazon GovCloud offering was chosen to ensure compliance with the most stringent security standards and to provide the flexibility to support a wide variety of clients.



Client Data Handling

Client data is stored within secure AWS-managed databases and is encrypted using service unique keys. Client data is partitioned using individualized access controls to prevent client data exposure. Multiple event and behavior monitors are in place to ensure only authorized access to client data occurs and all access is logged to provide a record of data access.

Access Protections

A two-stage process is used to restrict access to the web service. The first stage limits access to known client organizations and the second stage individually identifies users or service connections.

Network and Communication Protections

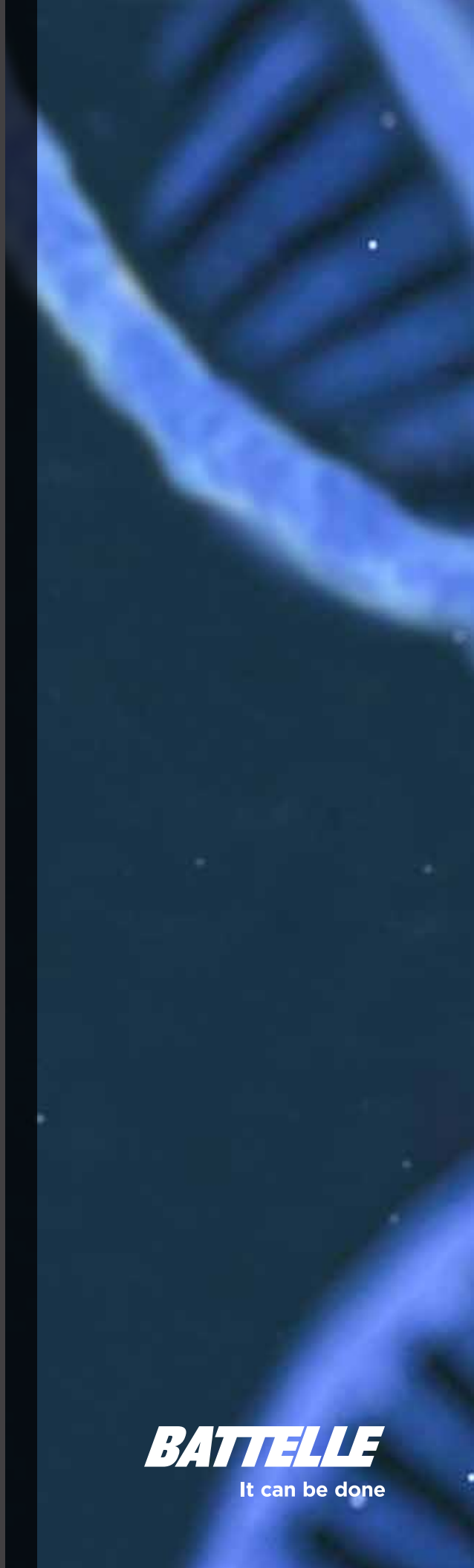
All internal and external communications are encrypted using the robust cryptographic services provided by AWS GovCloud that are validated against the United States National Institute of Standards and Technology Federal Information Processing Standard 140-2. Individual client user access is protected using encrypted communications and automated access to the service interface is protected by an encrypted Virtual Private Network (VPN) connection.

Who is Battelle?

We are the world's largest independent, nonprofit research and development organization, operating at the forefront of scientific discovery. We apply cutting-edge methods to some of today's most complex biosecurity challenges. At Battelle, you'll find:

- **Expertise:** We bring together leading experts in applied genomics, immunology, toxicology, bioinformatics, chemical and biological warfare defense, advanced analytics, cybersecurity and a host of related disciplines to develop multidisciplinary approaches to biosecurity problems.
- **Experience:** We have spent more than a decade building our proprietary database of virulence factors and other sequences of concern. We bring a deep understanding of the factors that make genetic sequences potential biosecurity threats.
- **Objectivity:** Our research is grounded in solid, objective science and proven methods. We continually reinvest in original research that benefits our clients and society.

Ready for a new approach to biosecurity risk reduction? Contact us for a demo.



Every day, the people of Battelle apply science and technology to solving what matters most. At major technology centers and national laboratories around the world, Battelle conducts research and development, designs and manufactures products, and delivers critical services for government and commercial customers. Headquartered in Columbus, Ohio since its founding in 1929, Battelle serves the national security, health and life sciences, and energy and environmental industries. For more information, visit www.battelle.org.

800.201.2011 | **solutions@battelle.org** | **www.battelle.org**

BATTELLE
It can be done

ID 634 04/18