Improved Annotation of Sequences of Concern to Better Understand Biorisk in Engineered Microbes

Correspondence: ggodbold@signaturescience.com

Gene D. Godbold (SIGNATURE SCIENCE, LLC) • **Todd J. Treangen** (RICE UNIVERSITY, Computer Science Department) • **Krista L. Ternus** (SIGNATURE SCIENCE, LLC)

Problem Statement

As synthetic DNA sequences for microbial engineering become easier to obtain, the risk of an accidental or deliberate generation of a microbial threat increases. Effective screening requires recognition of potentially hazardous sequences.

Working Toward a Solution

- We developed open-source software, SeqScreen, to screen sequences for those with hazardous functions: sequences of concern (SoCs)
- SeqScreen requires that sequences that cause problems be recognized and classified.
- How are hazardous sequences best understood?

Microbial Pathogens for Humans[1-3]

How many microbial pathogens for humans are there? Probably fewer than 1600:

- ~600 fungi,
- ~600 bacteria,
- 250 RNA viruses
- 50 DNA viruses
- 60 protozoa

How Do SoCs Contribute to Disease?

Damaging SoCs

- Cytotoxic
- Degrade Tissue
- Disable Organ
- Induce Inflammation

Other Direct-Acting SoCs

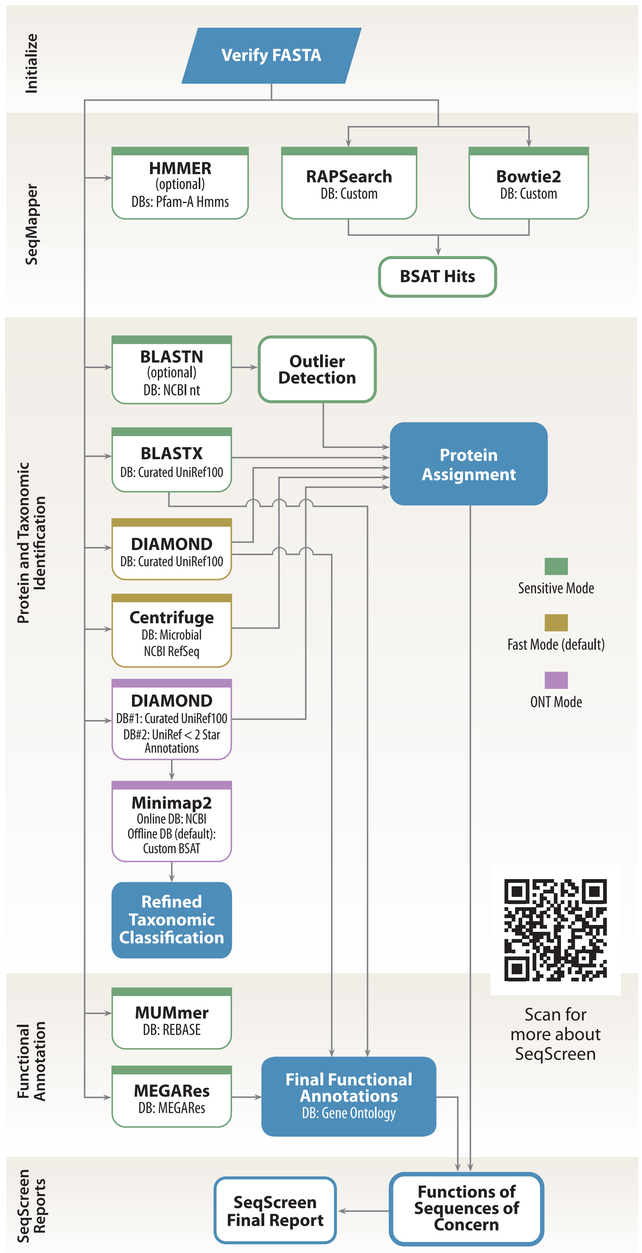
- Adherence to host cell
- Dissemination in host
- Host cell invasion
- Movement in host cell
- Niche-creationin host cell

Immune-Subverting SoCs

- Suppress host immune signaling by disrupting:
  - JAK-STAT
  - MAPK
  - NFκB
  - RIG-I
  - Protein kinase R
  - STING
  - TNF receptor-associated factors
  - Toll-like receptor binding/activation
- Resist host phagocytosis
- Resist host complement-killing (serum resistance)
- Resist host antimicrobial peptide
- Resist host oxidative killing
- Counter host immunoglobulin
- Defeat host chemokine or cytokine
- Inhibit host antigen presentation
- Immunomodulation

We have annotated functions of sequences of concern [FunSoCs] from 105 bacterial, 58 viral, 7 protozoal, and 4 fungal species pathogenic for humans.

SeqScreen Workflows



A Dataset of SoCs Annotated by Function Could Bring Clarity to Dual-Use Research of Concern (DURC) for Researchers, Regulators, and Funding Agencies

Function of Sequence of Concern (FunSoC)	Case 1 – SoC transferred to other pathogen or Case 2 – SoC altered for enhancement of original pathogen	Case 3 – SoC transferred to nonpathogen
Damaging	• Could enhance the harmful consequences of the agent	• Might enable the nonpathogen to have harmful consequences
Immune-Subverting	• Enhances the harmful consequences of the agent • Disrupts immunity or the effectiveness of an immunization against the agent • Alters the host range or tropism of the agent • Enhances the susceptibility of a host population to the agent	• Might enable the nonpathogen to have harmful consequences • Might enable the nonpathogen to infect novel hosts • Might enhance the susceptibility of a host population to the agent
Attachment Protein/ Adhesin	• Alters the host range or tropism of the agent • Enhances the susceptibility of a host population to the agent	• Probably none
Fusion Protein/Invasin	• Alters the host range or tropism of the agent • Enhances the susceptibility of a host population to the agent	• Probably none
Dissemination	• Enhances the harmful consequences of the agent • Increases the transmissibility or the ability to disseminate the agent • Enhances the susceptibility of a host population to the agent	• Probably none

Acknowledgements & References

This work was partially supported by the Fun GCAT program from the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via the Army Research Office (ARO) under Federal Award No. W911NF-17-2-0089. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, ARO, or the US Government.

References

1. Woolhouse M, Gaunt E. (2007) Ecological origins of novel human pathogens. Crit Rev Microbiol 33:231–242. 10.1080/10408410701647560.
2. Brown GD, Denning DW, Levitz SM. (2012) Tackling human fungal infections. Science 336:647. 10.1126/science.1222236.
3. Woolhouse MEJ, Brierley L. (2018) Epidemiological characteristics of human-infective RNA viruses. Sci Data 5:180017. 10.1038/sdata.2018.17.
4. Godbold, GD, Kappell, AD, LeSassier, DS, Treangen, TJ, and Ternus, KL (2022) Categorizing Sequences of Concern by Function To Better Assess Mechanisms of Microbial Pathogenesis. Infect Immun 90, e0033421. doi: 10.1128/IAI.00334-21.
5. Balaji, A, Kille, B, Kappell, AD, Godbold GD, Diep, M, Elworth RAL, Qian, Z, Albin, D, Nasko, DJ, Shah, N, Pop, M, Segarra S, Ternus, KL, Treangen TJ (2022) SeqScreen: accurate and sensitive functional screening of pathogenic sequences via ensemble learning. Genome Biol 23: 133. doi: 10.1186/s13059-022-02695-x.

Work in Progress

We are involved in reforming the Gene Ontology (GO) terms relating to host-symbiont interactions via the Multiorganism Working Group. This will provide a greater comprehension of these processes for both humans and machines.