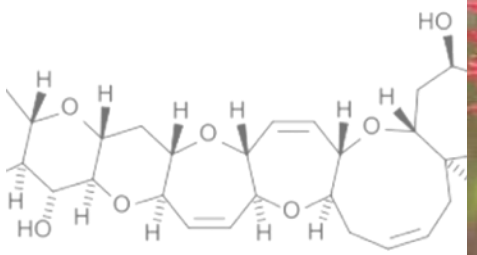
© Copyright 2023

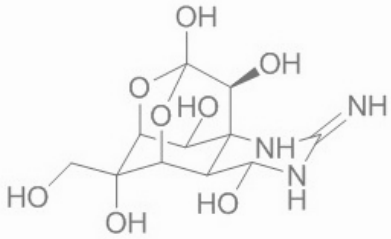
What Are “Bad” Sequences?

Gene Godbold, Signature Science

Encoded biological threats are primarily from microbes that cause infectious diseases and secondarily from metazoan and plant venom components/toxins



ciguatoxin



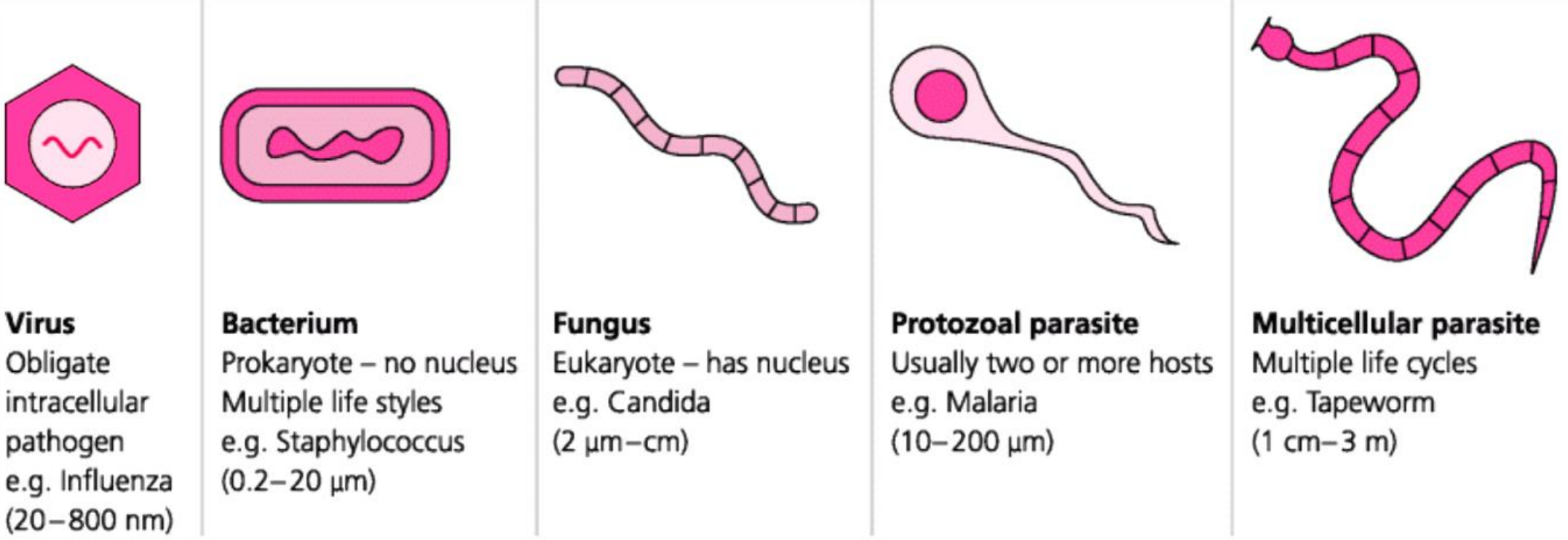
tetrodotoxin

Accumulator of batrachotoxin

RICE Aclid signature science LLC

“Bad Bugs” Are Parasites

Parasite = A smaller organism that lives on or in a larger host organism and causes harm (*disease* = *loss of homeostasis*) to the host



Virus
Obligate intracellular pathogen
e.g. Influenza
(20–800 nm)

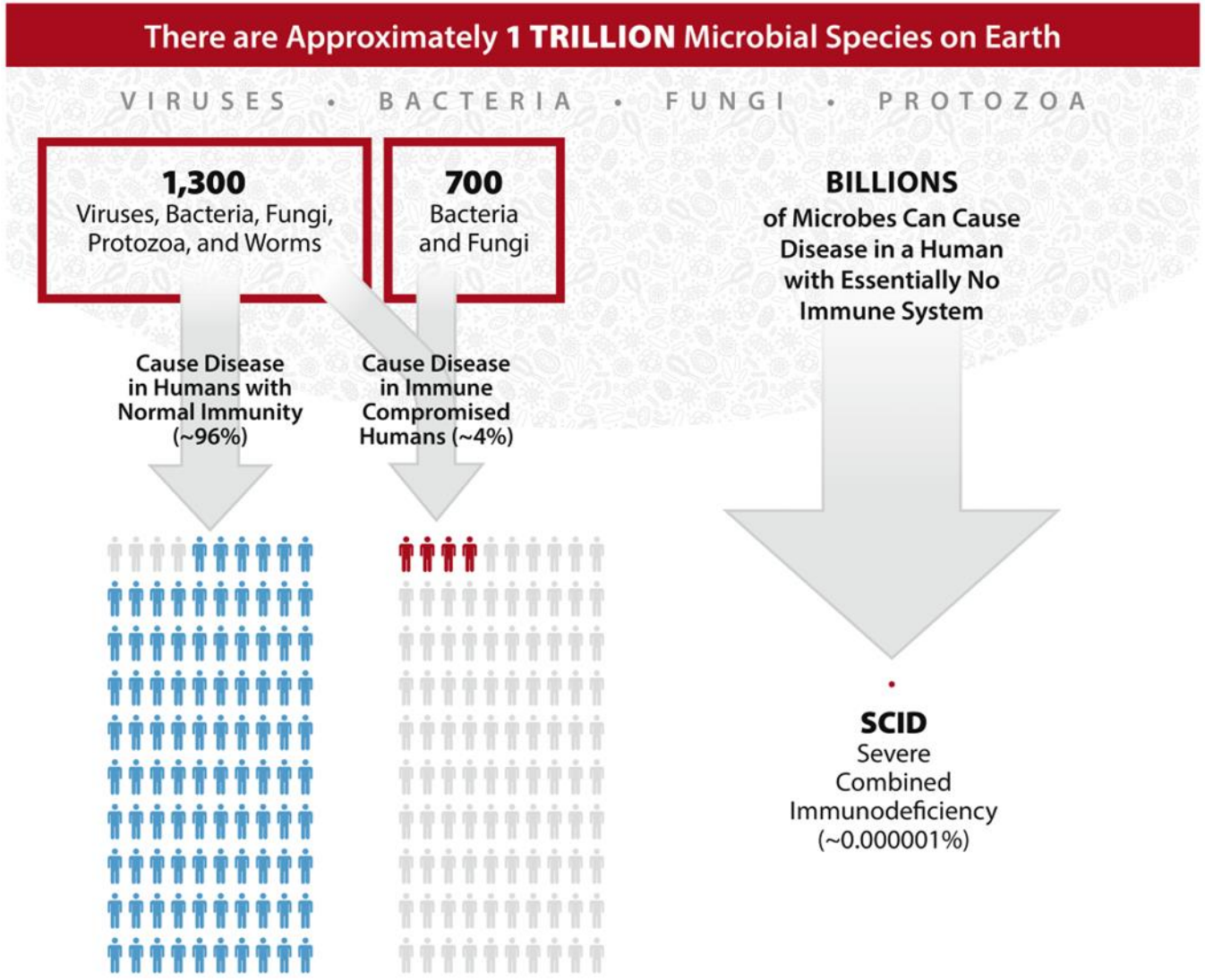
Bacterium
Prokaryote – no nucleus
Multiple life styles
e.g. Staphylococcus
(0.2–20 µm)

Fungus
Eukaryote – has nucleus
e.g. Candida
(2 µm–cm)

Protozoal parasite
Usually two or more hosts
e.g. Malaria
(10–200 µm)

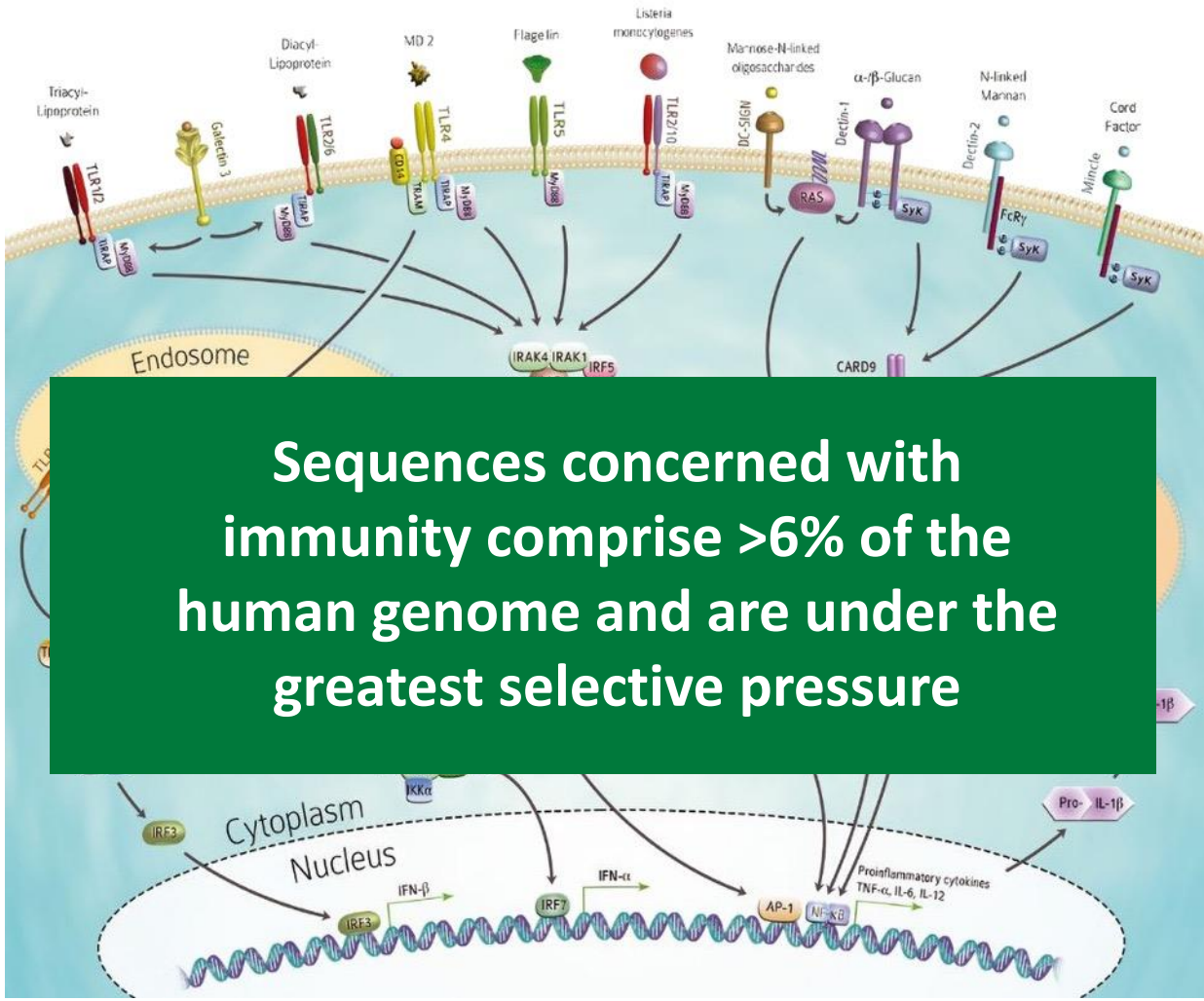
Multicellular parasite
Multiple life cycles
e.g. Tapeworm
(1 cm–3 m)

How many bad bugs are there?



How Do Hosts Resist Parasites?

- Barriers
  - Physical barriers: skin
  - Chemical barriers: stomach acid, mucus, antimicrobial peptides
  - Biological barriers: microbiomes (skin, gut, etc.)
- Immune system
  - Innate (constant, invariant)
  - Adaptive (develops, specific)
- *To be a parasite requires a "susceptible host," but what makes a host susceptible?*



Sequences concerned with immunity comprise >6% of the human genome and are under the greatest selective pressure

The outcome of any host-parasite encounter is chiefly governed by the immune defense of the host and the virulence factors of the parasite

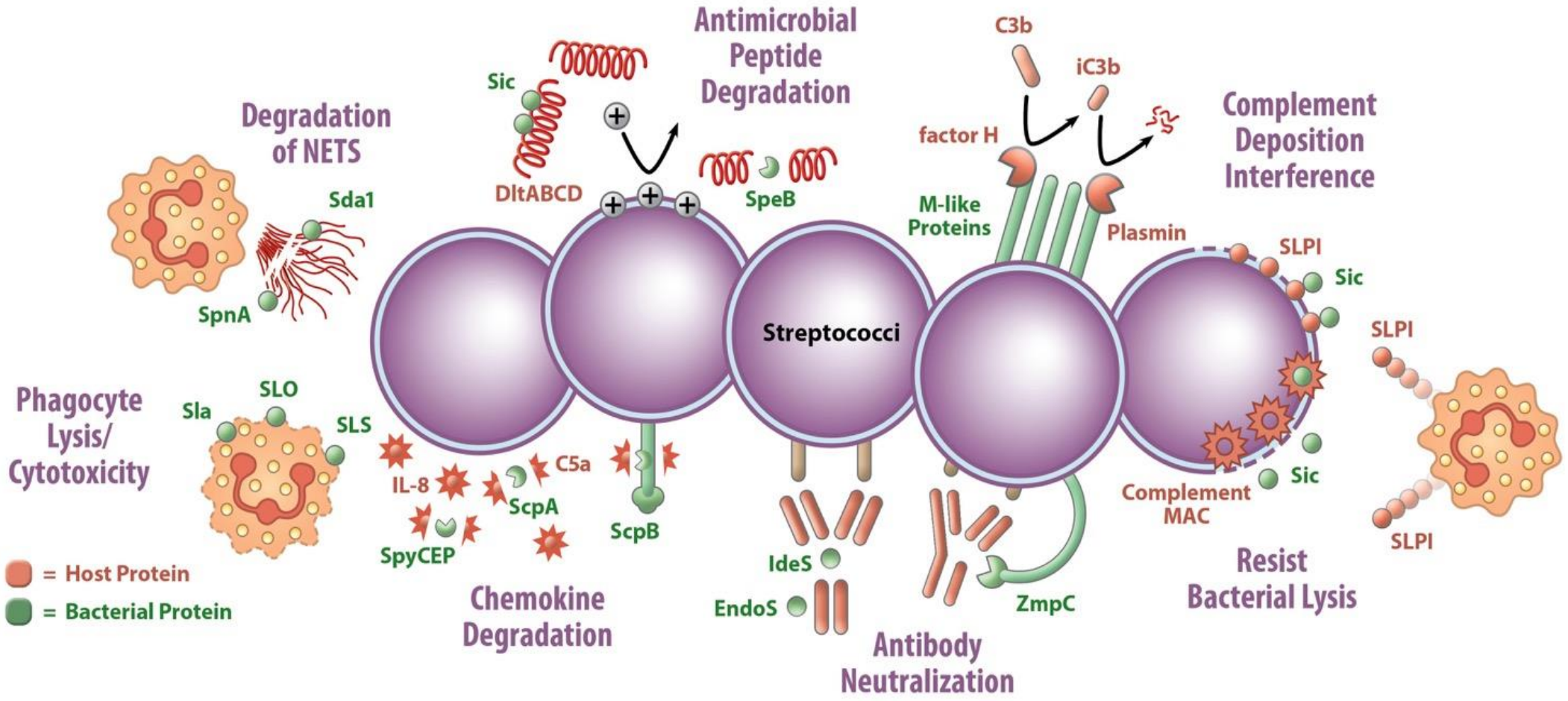
- What would happen if we lacked an immune system?
- **Hypothesis**: Sequences that subvert immunity *make* hosts susceptible.

RICE Aclid signature science LLC

Immune Subversion: Direct

1. Suppress Host Cell Immune Signaling
2. Subverting antigen presentation
3. Countering immune effectors

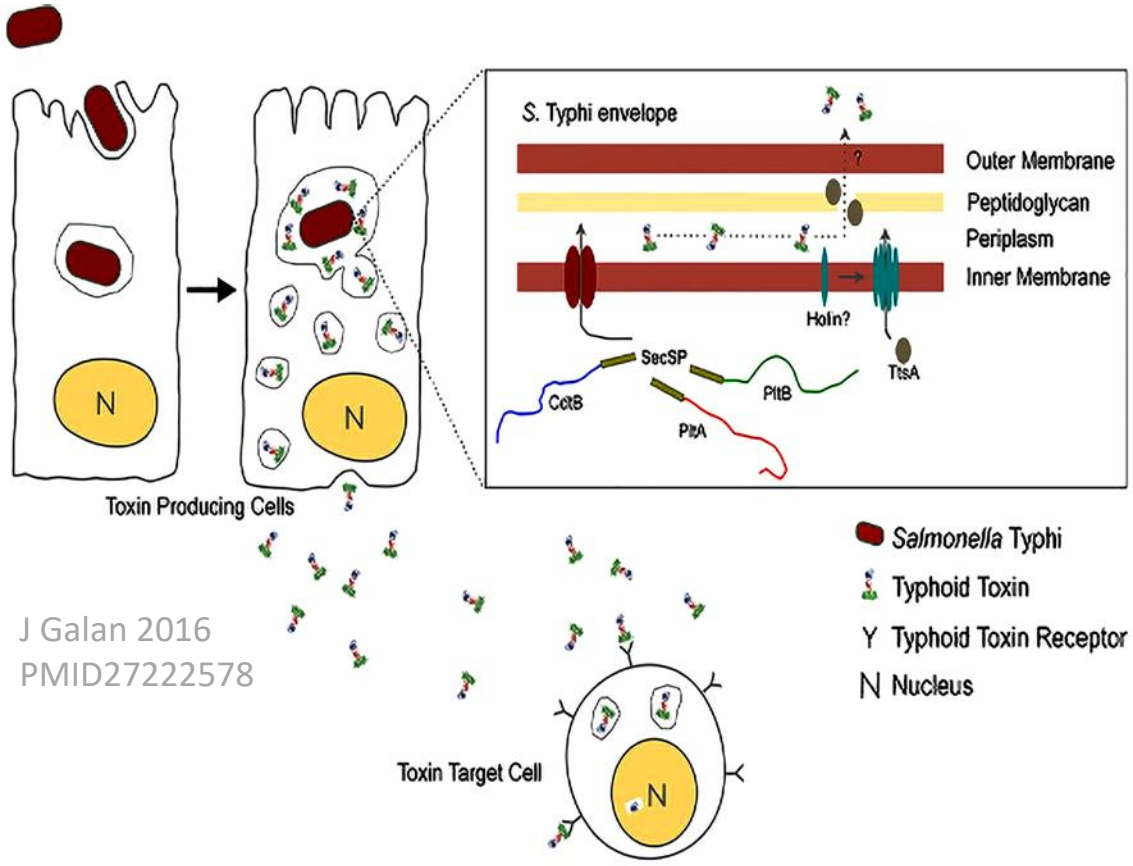
Sequences Streptococci Use to Directly Subvert Immunity



http://journals.asm.org/doi/10.1128/IAI.00334-21

How We Curate - Evaluating Sequences for Biothreat

1. Where does it act? **Parasite** or **Host**?
2. Host **Extracellular** or **Intracellular**?
3. How does it get to where it acts? Secretion? Adherence + invasion?)
4. What is the targeted molecule?
5. Which host activity or pathway does it affect?
6. What is the effect on the host?

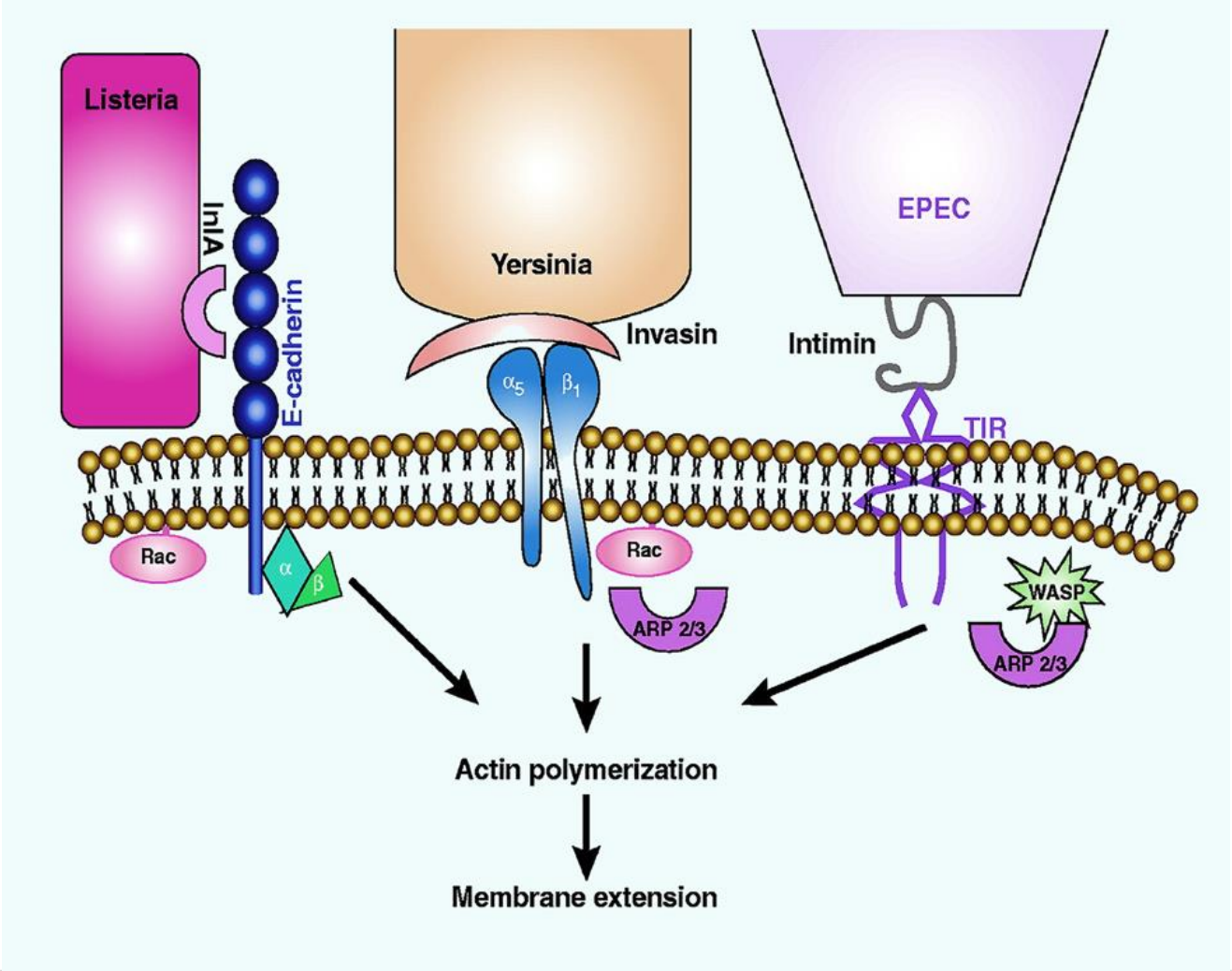


J Galan 2016
PMID27222578

Typhoid toxin targets host epithelial cells and leukocytes where it (1) degrades DNA and (2) ADP-ribosylates unknown host proteins

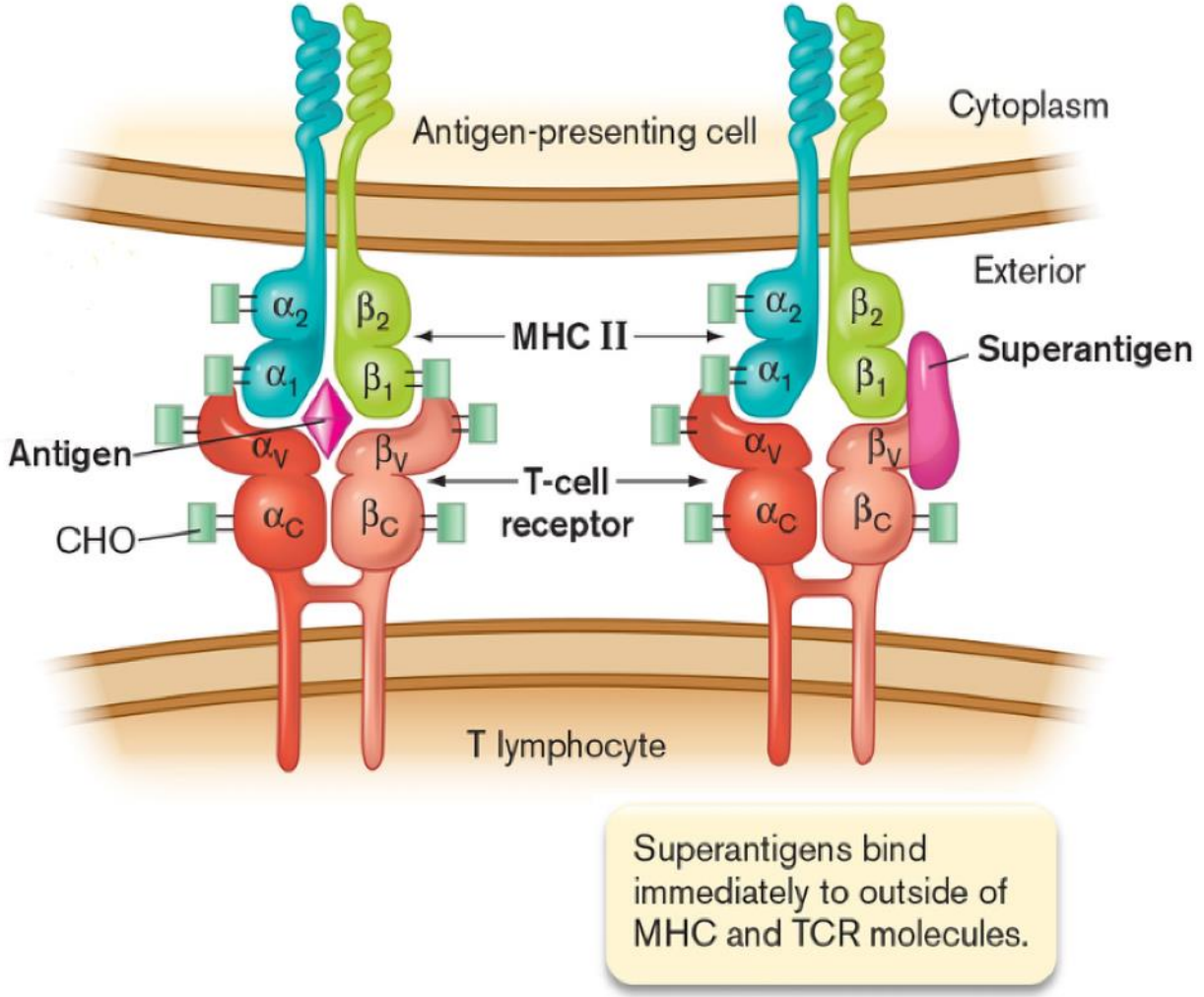
Effects of Sequences of Concern on the Host

1. **Adherence**
2. **Invasion**
3. Dissemination
4. Niche-creation
5. Damage

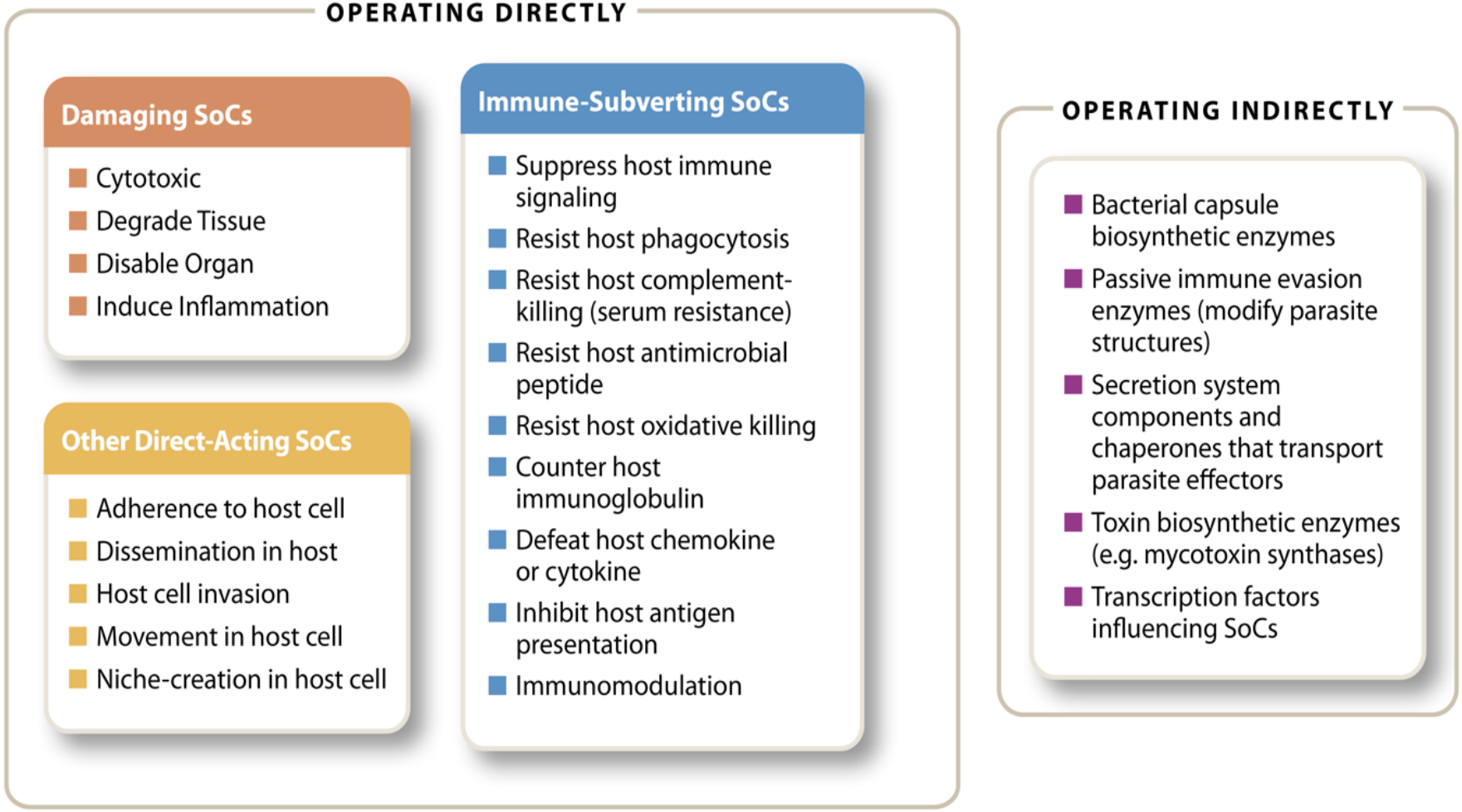


Types of Host Damage Inflicted by SoCs

- Cytotoxicity
- Cell membrane disruption
- Tissue permeabilization
- ECM destruction
- Release cell from substrate
- Disable organ
- Induce inflammation

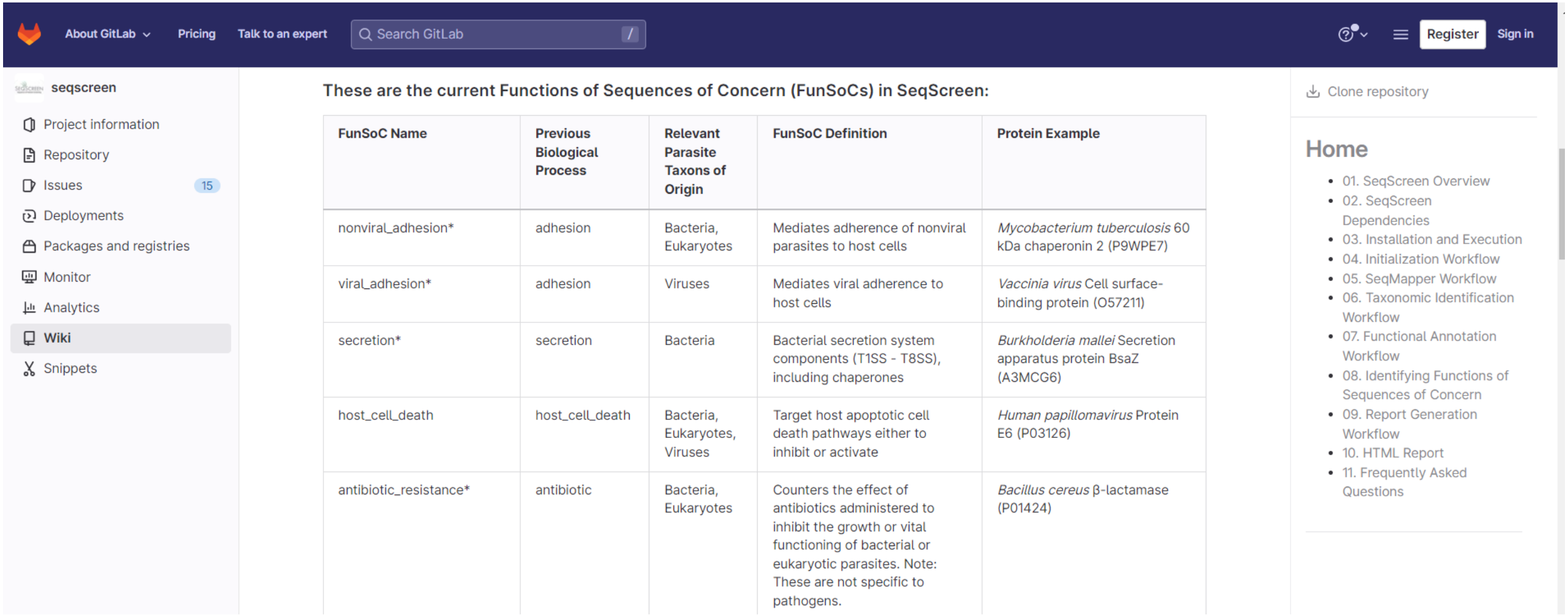


Functions of Sequences of Concern (FunSoCs)



Godbold *et al.*, 2022: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9119117/

32 Machine Learning-Predicted FunSoCs in SeqScreen



FunSoC Name	Previous Biological Process	Relevant Parasite Taxons of Origin	FunSoC Definition	Protein Example
nonviral_adhesion*	adhesion	Bacteria, Eukaryotes	Mediates adherence of nonviral parasites to host cells	*Mycobacterium tuberculosis* 60 kDa chaperonin 2 (P9WPE7)
viral_adhesion*	adhesion	Viruses	Mediates viral adherence to host cells	*Vaccinia virus* Cell surface-binding protein (O57211)
secretion*	secretion	Bacteria	Bacterial secretion system components (T1SS - T8SS), including chaperones	*Burkholderia mallei* Secretion apparatus protein BsaZ (A3MCG6)
host_cell_death	host_cell_death	Bacteria, Eukaryotes, Viruses	Target host apoptotic cell death pathways either to inhibit or activate	*Human papillomavirus* Protein E6 (P03126)
antibiotic_resistance*	antibiotic	Bacteria, Eukaryotes	Counters the effect of antibiotics administered to inhibit the growth or vital functioning of bacterial or eukaryotic parasites. Note: These are not specific to pathogens.	*Bacillus cereus* β-lactamase (P01424)

https://gitlab.com/treangenlab/seqscreen/-/wikis/08.-Identifying-Functions-of-Sequences-of-Concern

FunSoCs Reported in SeqScreen

- Appears in output as “0” for absence or “1” for presence

Query	Length	Organism	Gene Name	GO Terms	bsat	disable organ	cytotoxicity	degrade ecm	induce inflammation
BBa_K811000	915	Bifidobacterium longum subsp. infantis	Hemolysin E, chromosomal	1. membrane 2. extracellular region 3. host cell plasma membrane show more	NA	0	1	0	0

Pathogenesis Gene Ontology (PathGO)

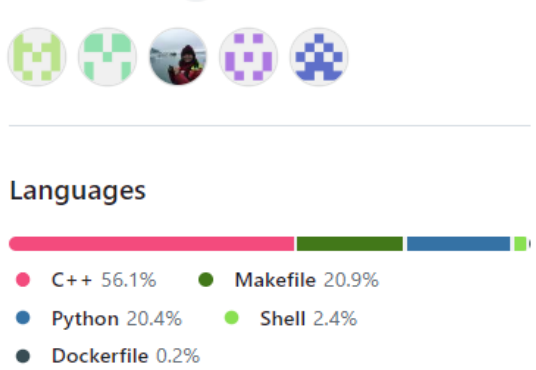
README.md

PathGO: the Pathogenesis Gene Ontology

Authors	Ron Jacak (rjacak)
	Jody Proescher (jproesch)
	Gene Godbold
	Amanda Ernlund (awe220)
	Tina Zudock (kzudock)

Description

This ontology contains terms to describe the function of genes and gene products involved in pathogenesis. Genes that encode toxins, virulence factors, and antimicrobial resistance have functions that are not represented well by the terms contained within the Gene Ontology or any of the other OBO Foundry ontologies. This ontology will collect and maintain common terms and descriptions of concepts related to pathogenesis, and could help provide more intuitive descriptions of function for these types of sequences.



https://github.com/jhuapl-bio/pathogenesis-gene-ontology

Format of PathGO Terms Reported by SeqScreen

Example of PathGO term assignment:
PATHGO:0000326 (modulates transcription in another organism)[PMID30832682]

toxin_synthase	viral_adhesion	virulence_regulator	size	organism	gene_name	uniprot	uniprot evalue	PathGO
0	0	0	501	Salmonella enterica subsp. enterica serovar Typhimurium str. LT2	Secreted effector protein SptP	P74873	2.86E-108	PATHGO:0000278 (stimulates small GTPa
0	0	0	315	Salmonella enterica subsp. enterica serovar Typhimurium str. LT2	Guanine nucleotide exchange factor sopE2	Q7CQD4	1.04E-66	PATHGO:0000255 (mediates guanine nuc
0	0	0	1011	Escherichia coli	Hygromycin-B 4-O-kinase	P00557	1.85E-216	-
0	0	0	1008	Acinetobacter baumannii	Beta-lactamase (Fragment)	A0A1P8NC	1.45E-204	-
0	0	0	630	Human herpesvirus 4 strain B95-8	Epstein-Barr nuclear antigen 1	P03211	2.90E-118	PATHGO:0000326 (modulates transcripti

seqscreen_output_directory\report_generation\case_study_3_seqscreen_report_pathgo.tsv

© Copyright 2023