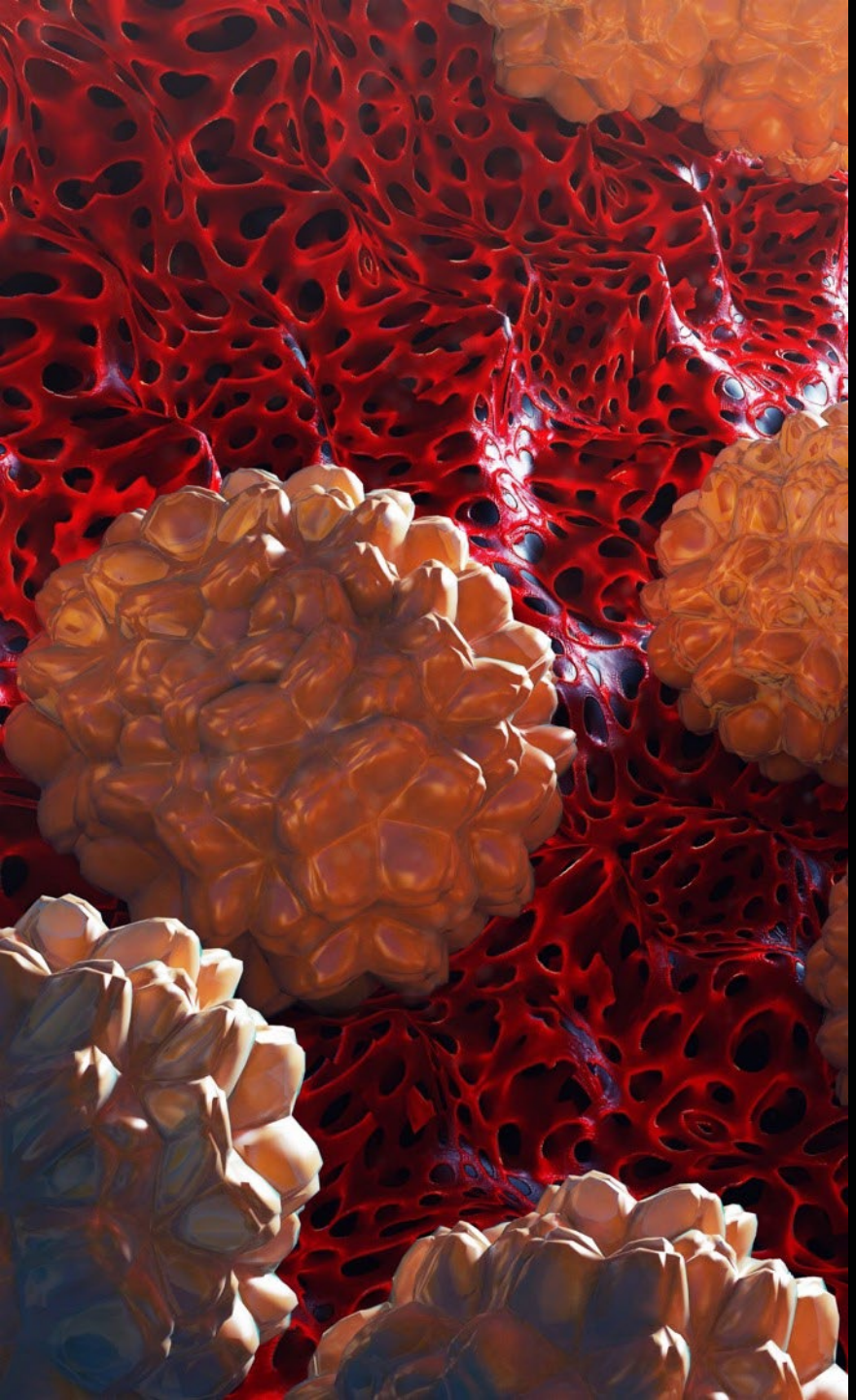
signature science LLC

© 2025 Signature Science, LLC



Pathogen Assessment by Tracking Variation in Sequences of Concern in Bacterial Genomes

Nexus AIxBio Workshop #1
13 November 2025

Gene D. Godbold
Krista L. Ternus
Joshua C. Gil

signature science LLC

SoC Accounting in Bacterial Pathogens: Requirements

- **Sequence of Concern Dataset**: 2600+ from 63 pathogenic bacterial genera
  - Identified from publications in microbial pathogenesis
  - Accessions found in UniProt
  - Annotated with terms specific for pathogenic function
    - Adherence, Invasion, Dissemination
    - Host Damage
    - Host Innate Immune Signaling Subversion
    - Counter Host Innate Immune Effectors
    - Manipulate Host Cell Biology
- **Genomes:** Complete genes or reference genomes are used, pulled from Refseq and Genbank (total = 183,128)
- **Bioinformatics/Software:** Every genome scanned for Sequences of Concern using Tblastn. The number of sequences are compiled for further analyses.

Godbold, Proescher & Gaudet 2025 https://pmc.ncbi.nlm.nih.gov/articles/PMC11927349/

Adherence, Invasion, Dissemination

GO ID	GO Term
GO:0044409	entry into host
GO:0044417	translocation of molecules into host
GO:0141018	adhesion of symbiont to host via host extracellular matrix
GO:0141024	adhesion of symbiont to host cell surface via host membrane carbohydrate
GO:0141025	adhesion of symbiont to host cell surface via host glycoprotein
GO:0141026	adhesion of symbiont to host cell surface via host membrane cholesterol
GO:0035756	symbiont-mediated migration across host epithelium
GO:0044067	symbiont-mediated perturbation of host cell-cell junction
GO:0106259	symbiont mediated cell-to-cell migration in host
GO:0141142	symbiont-mediated migration across host tissue barrier

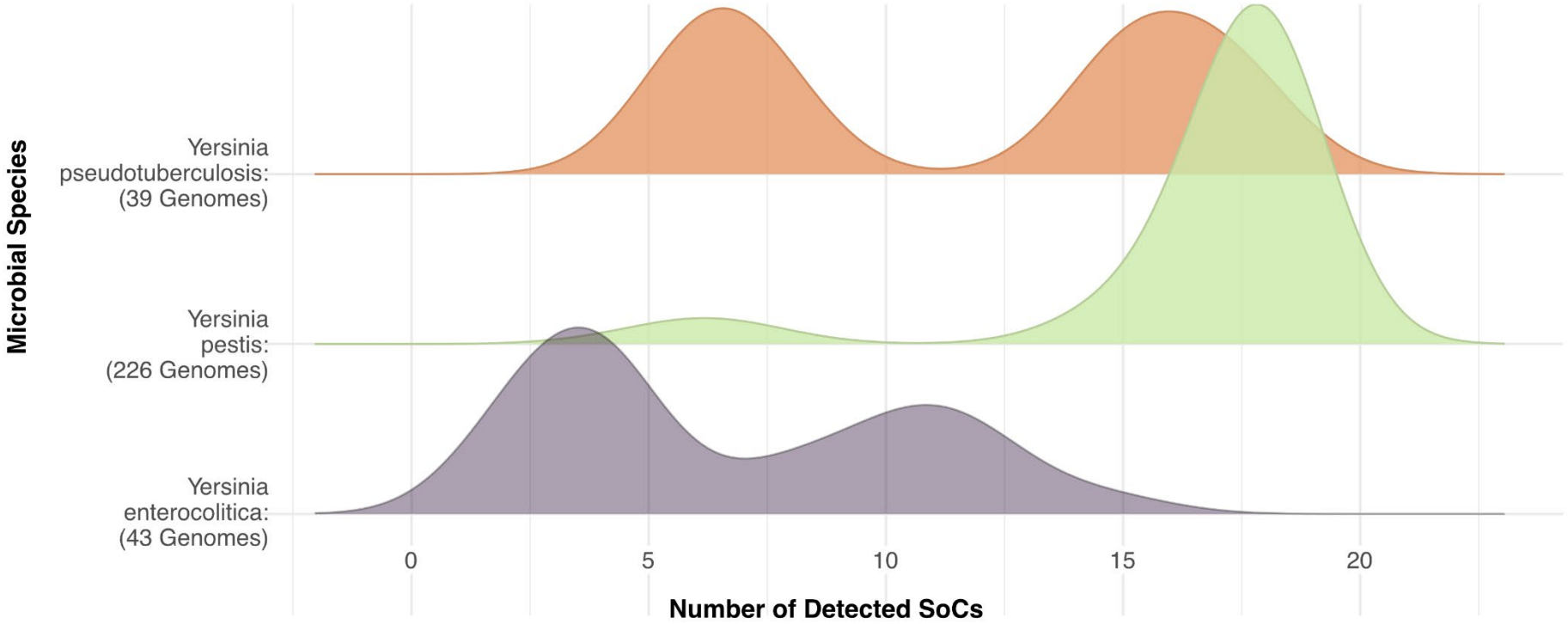
Subversion of Host Innate Immune Signaling

GO ID	GO Term
GO:0141105	symbiont-mediated suppression of host toll-like receptor signal transduction
GO:0039537	symbiont-mediated suppression of cytoplasmic pattern recognition receptor signaling pathway
GO:0140886	symbiont-mediated suppression of host interferon-mediated signaling pathway
GO:0039514	symbiont-mediated suppression of host JAK-STAT cascade
GO:0141074	symbiont-mediated suppression of host cGAS-STING signal transduction
GO:0141078	symbiont-mediated suppression of host RIG-I signaling pathway
GO:0039580	symbiont-mediated suppression of host PKR/eIFα signaling
GO:0085034	symbiont-mediated suppression of host NF-κB cascade
GO:0141070	symbiont-mediated suppression of host MAPK cascade
GO:0141072	symbiont-mediated suppression of host tumor necrosis factor signaling pathway
GO:0141135	symbiont-mediated suppression of host chemokine signal transduction pathway
GO:0141081	symbiont-mediated suppression of host inflammasome-mediated signal transduction
GO:0141083	symbiont-mediated suppression of host reactive oxygen species generation

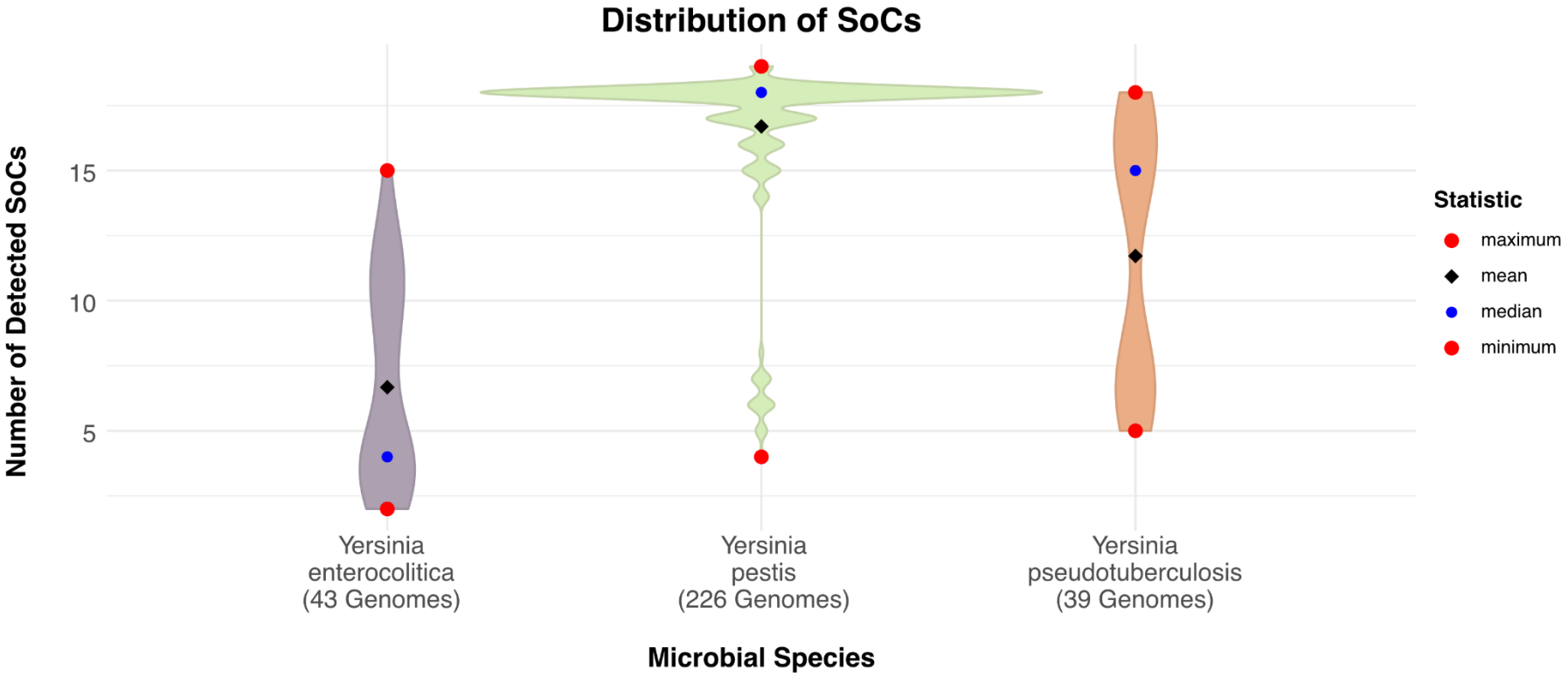
Subversion of Host Innate Immune Effectors

GO ID	GO Term
GO:0141140	symbiont-mediated suppression of host immunoglobulin-mediated immune response
GO:0141114	symbiont-mediated suppression of host complement activation
GO:0052067	symbiont-mediated perturbation of host phagocytosis
GO:0141145	symbiont-mediated suppression of host neutrophil extracellular trap formation
GO:0140133	symbiont-mediated suppression of host cytokine production
GO:0141173	symbiont-mediated suppression of host pro-inflammatory cytokine signaling
GO:0141184	symbiont-mediated activation of host anti-inflammatory cytokine signaling
GO:0039588	symbiont-mediated suppression of host antigen processing and presentation
GO:0141059	symbiont-mediated disruption of host antimicrobial peptide activity
GO:0141082	symbiont-mediated detoxification of host-generated reactive oxygen species

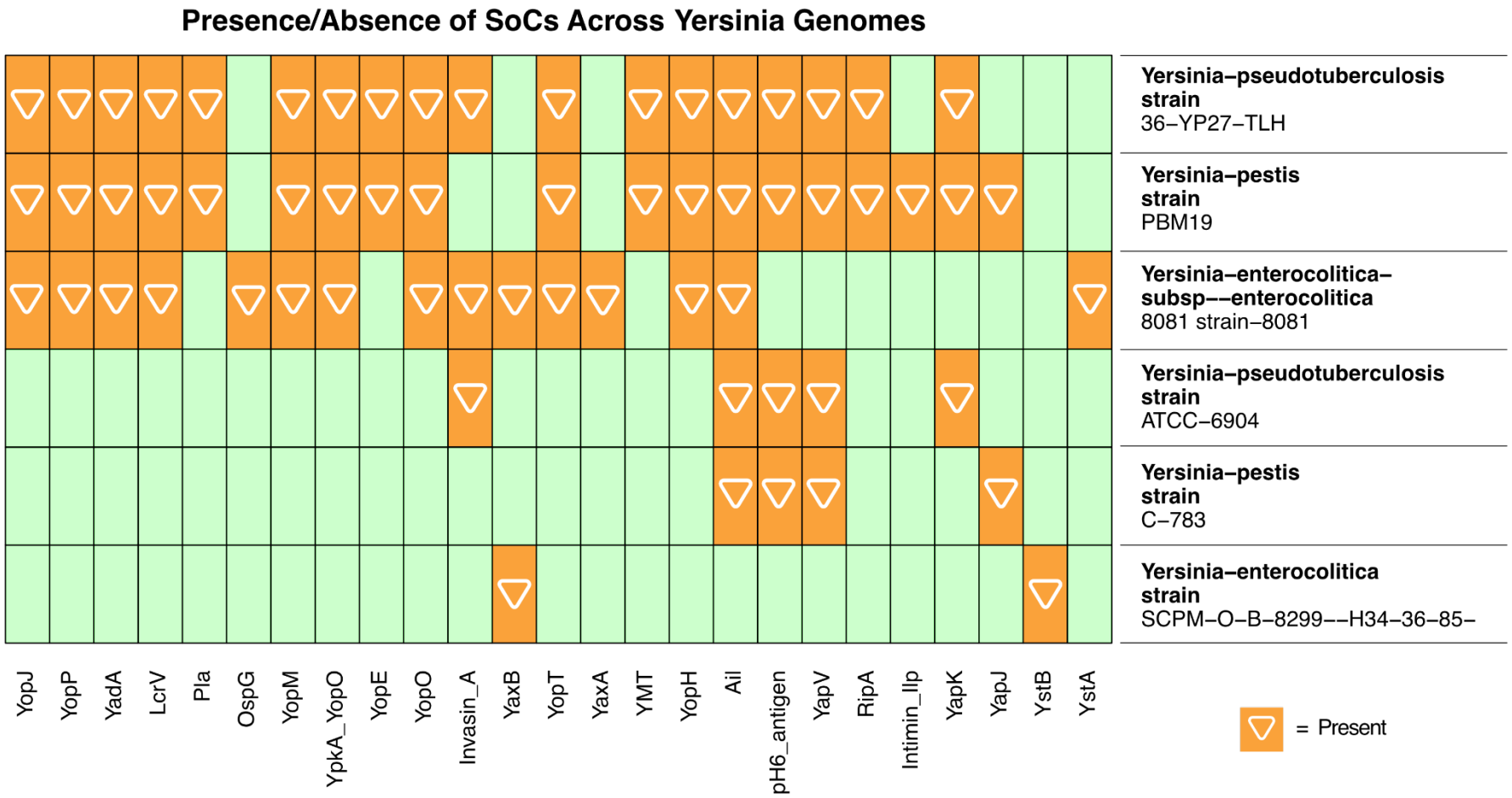
Distribution of SoCs in Pathogenic *Yersiniae*



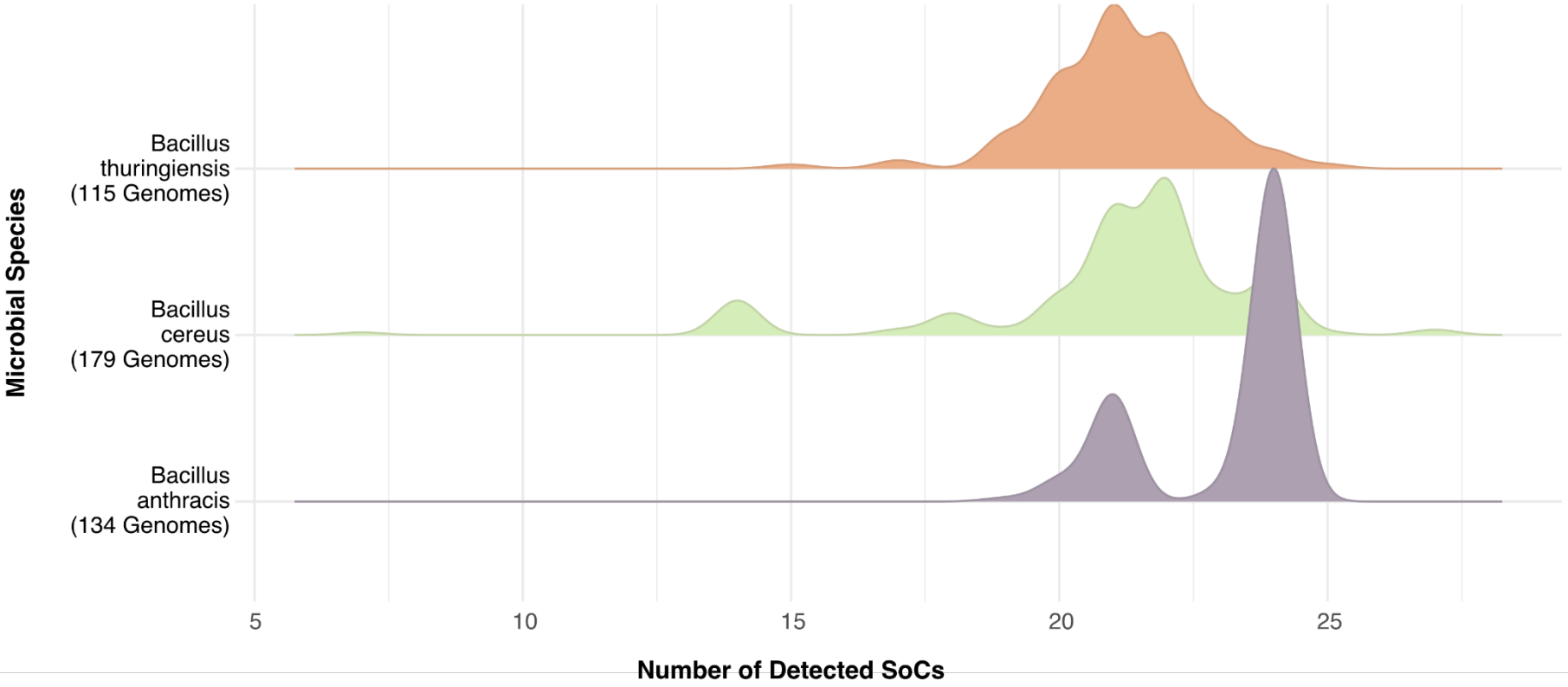
Distribution of SoCs in Pathogenic *Yersiniae*



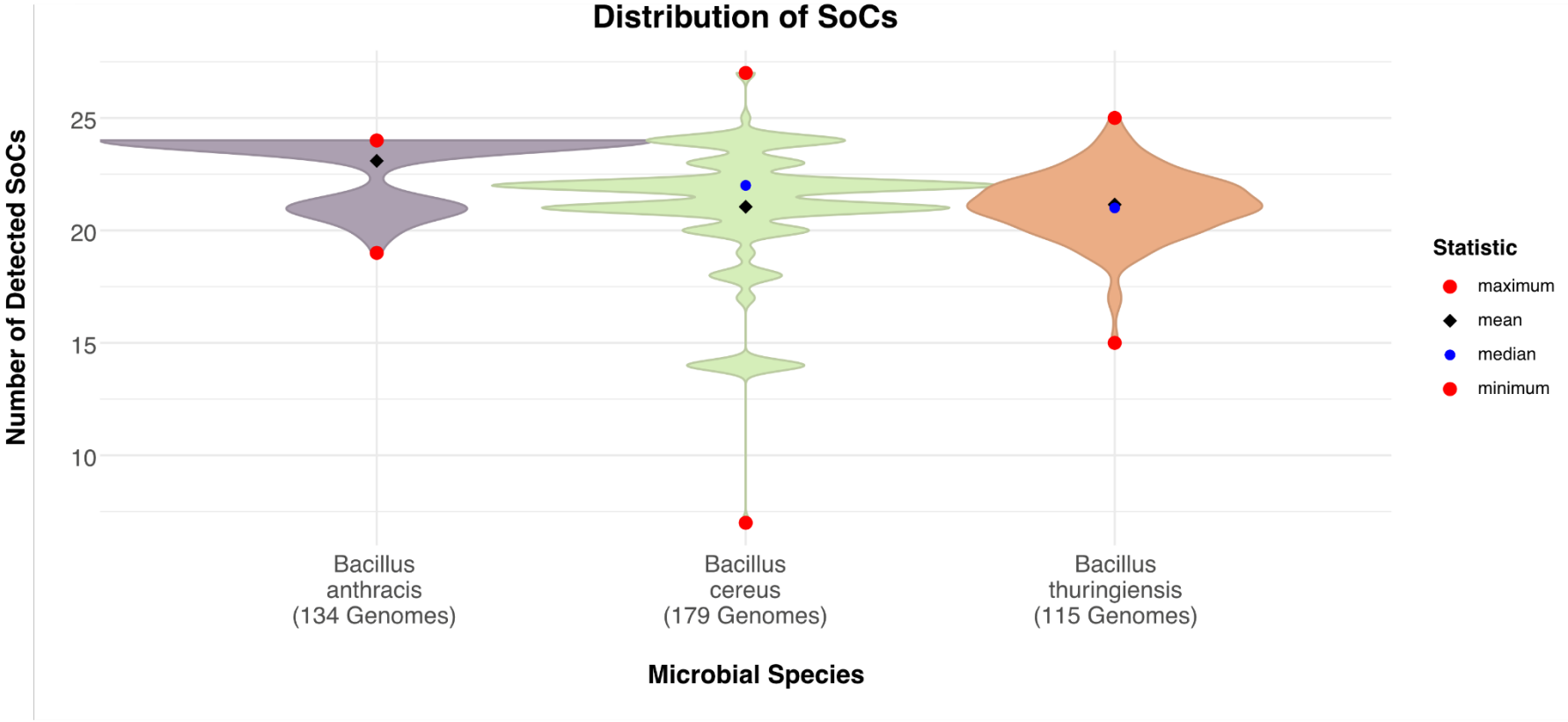
SoCs in Pathogenic *Yersiniae* (Max/Min)



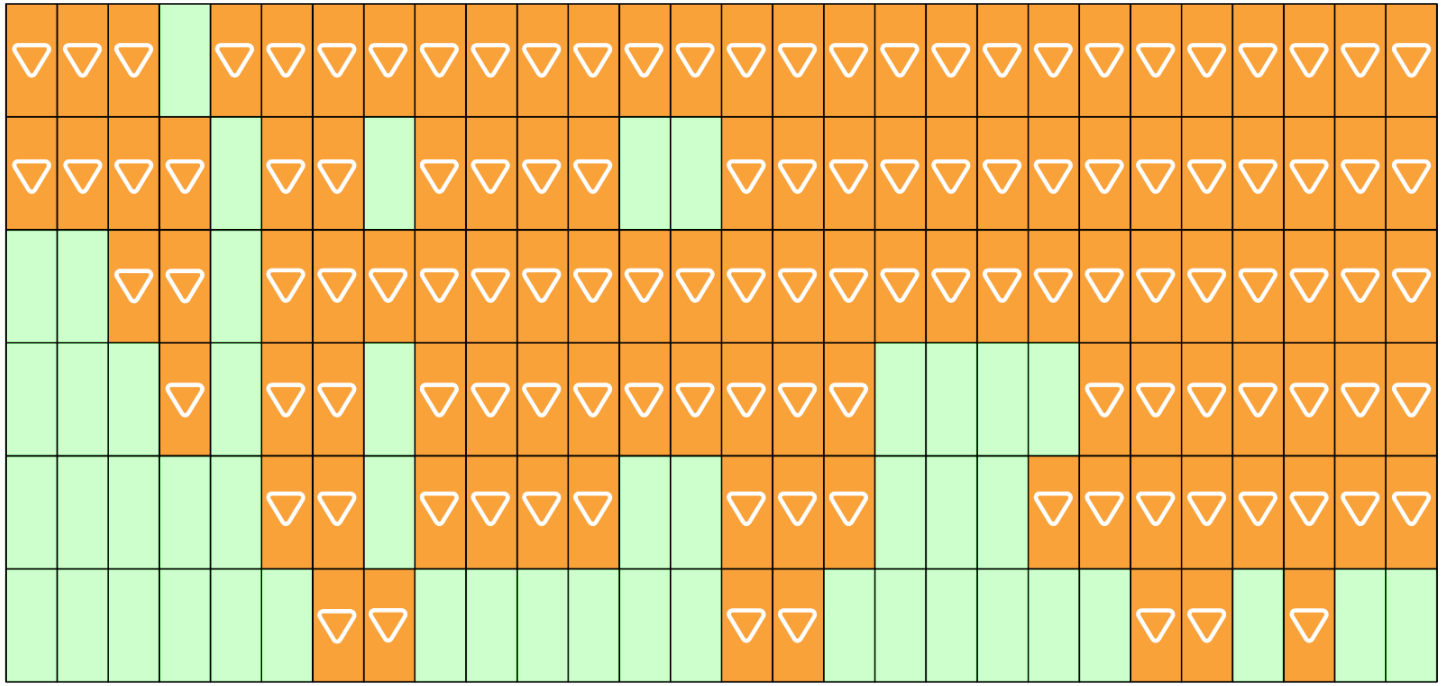
SoCs in pathogenic *Bacilli*



SoCs in pathogenic *Bacilli*



SoCs in pathogenic *Bacilli* (max/min)



Lethal_factor
Edema_factor
Hemolysin_II
PI_PLC
Certhrax_toxin
Sphingomyelinase
Cereolysin_O
HBL_enterotoxin_B
NheA
InhA
NheB
NheC
Cytotoxin_K2
Cytotoxin_K2_HlgB
Alveolysin
Thiol_cytolysin
Anthrolysin_O
BslA_Slgp1
BslA
Protective_antigen
AdsA
SodC
SodA1
ClpX
SodA2
MPRF
Cereolysin_A
CerA_PLC

Bacillus–cereus strain 03BB87

Bacillus–anthracis–str–– Ames–Ancestor– strain– Ames–Ancestor––A2084

Bacillus–thuringiensis strain–CTC

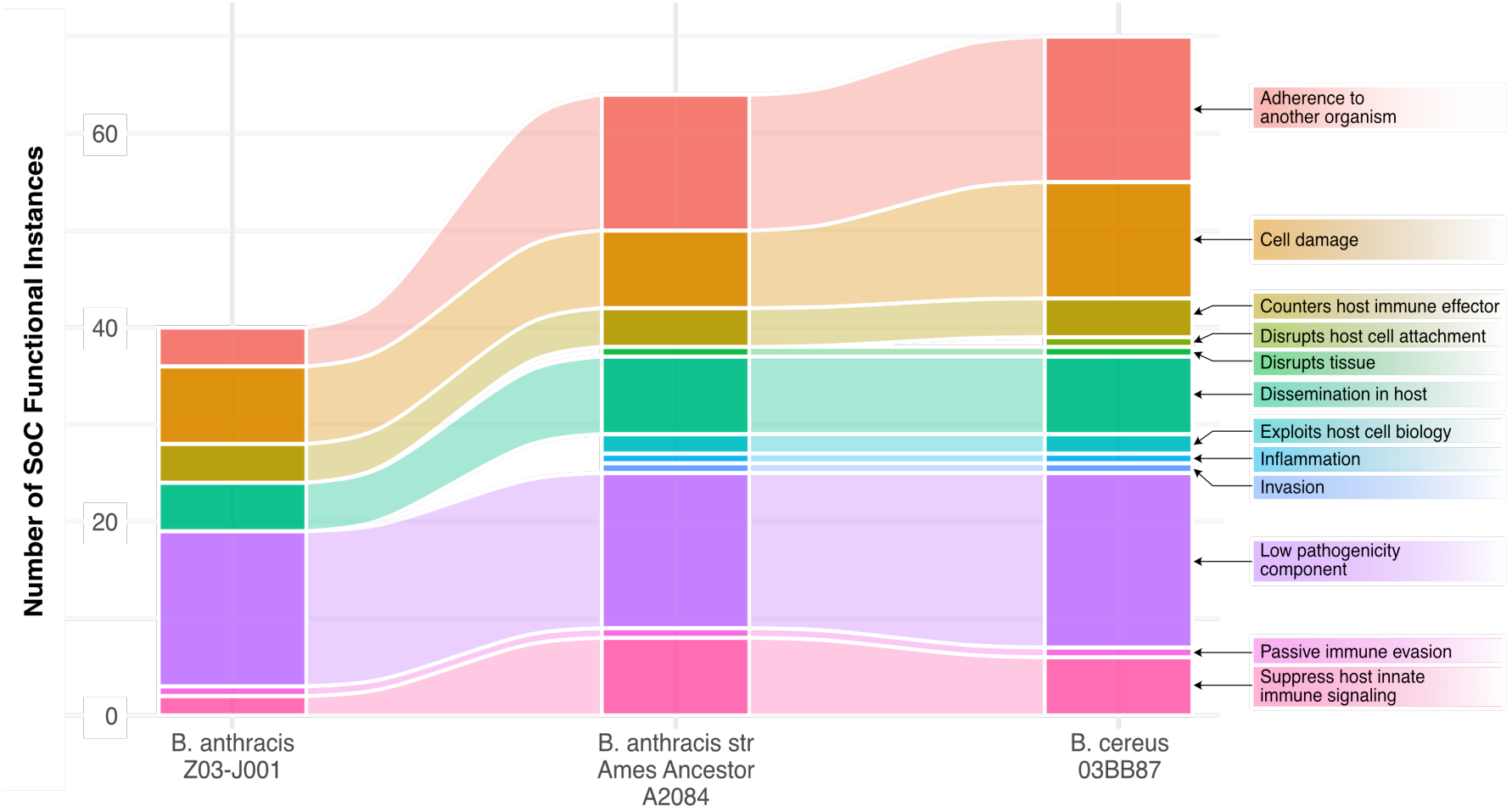
Bacillus–anthracis strain–Z03–J001

Bacillus–thuringiensis strain–Bt–Gxmzu777–1

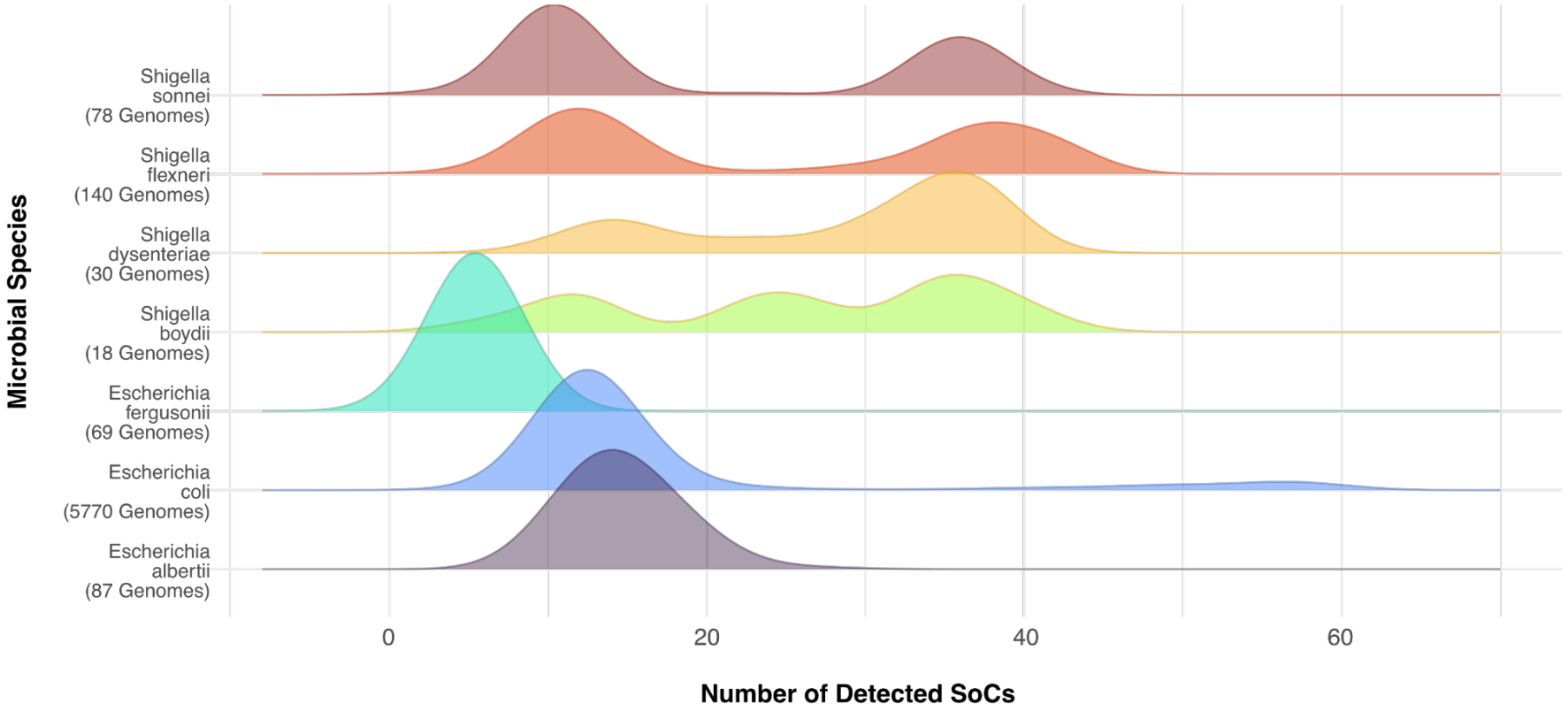
Bacillus–cereus strain SIN1–2

▽ = Present

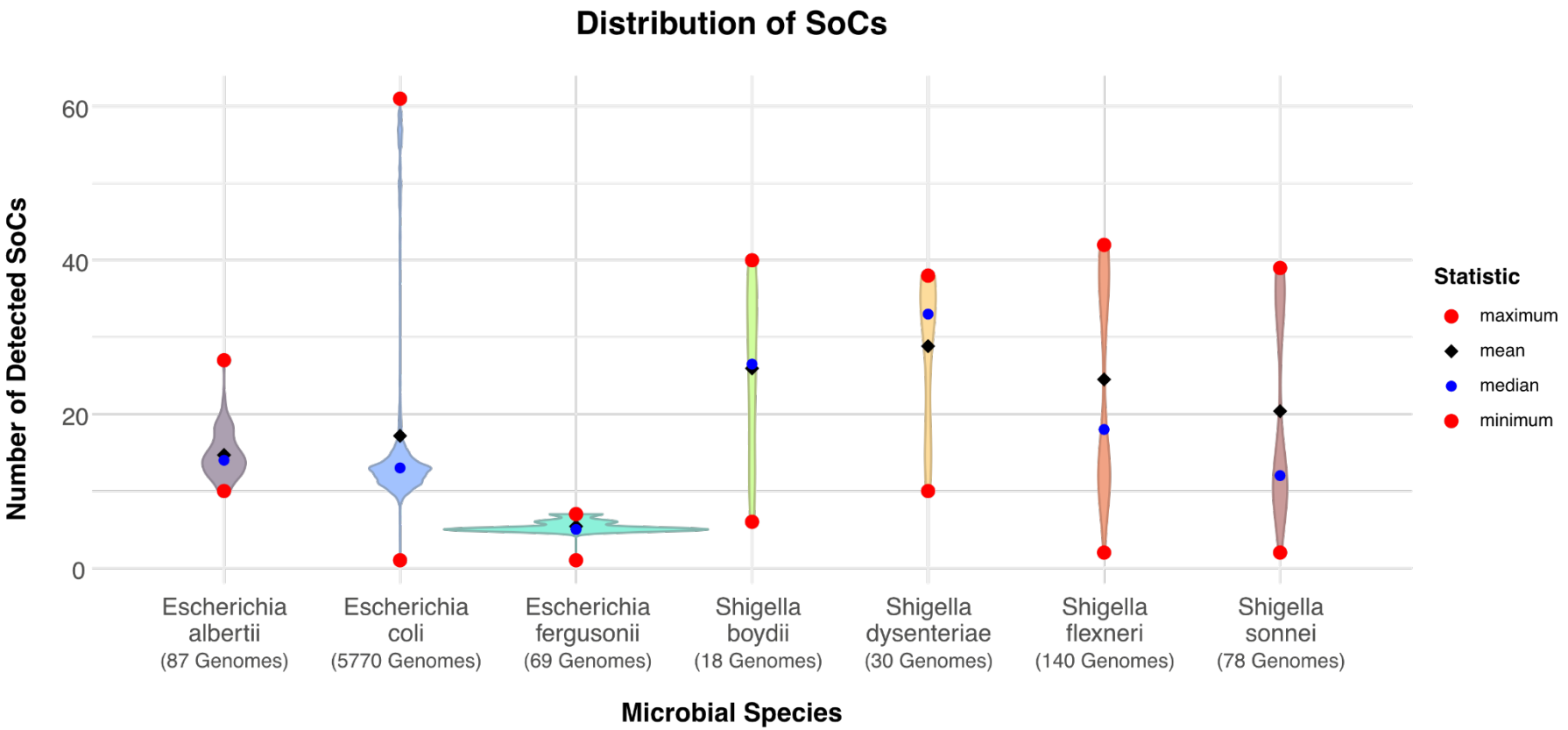
Pathogenic Functions Across 3 *Bacilli* Isolates



Distribution of SoCs in *Escherichia* and *Shigellae*



Distribution of SoCs in *Escherichia* and *Shigellae*



signature science LLC

© 2025 Signature Science, LLC