Categorizing Sequences of Concern by Function To Better Assess Mechanisms of Microbial Pathogenesis

Gene D. Godbold, Anthony D. Kappell, Danielle S. LeSassier, Todd J. Treangen, Krista L. Ternus

Signature Science, LLC, Charlottesville, Virginia, USA
Signature Science, LLC, Austin, Texas, USA
Department of Computer Science, Rice University, Houston, Texas, USA

Todd J. Treangen and Krista L. Ternus contributed equally to the supervision of the work.

ABSTRACT To identify sequences with a role in microbial pathogenesis, we assessed the adequacy of their annotation by existing controlled vocabularies and sequence databases. Our goal was to regularize descriptions of microbial pathogenesis for improved integration with bioinformatic applications. Here, we review the challenges of annotating sequences for pathogenic activity. We relate the categorization of more than 2,750 sequences of pathogenic microbes through a controlled vocabulary called Functions of Sequences of Concern (FunSoCs). These allow for an ease of description by both humans and machines. We provide a subset of 220 fully annotated sequences in the supplemental material as examples. The use of this compact (~30 terms), controlled vocabulary has potential benefits for research in microbial genomics, public health, biosecurity, biosurveillance, and the characterization of new and emerging pathogens.

KEYWORDS biodefense, bioinformatics, biothreat, controlled vocabulary, host-pathogen interactions, immune evasion, microbial pathogenesis, ontology, sequence of concern, sequence screening

WHAT MAKES “BAD BUGS” BAD?

The “worst” pathogens of humans cause severe disease in those possessing normal immunity. Pathogens of other organisms indirectly affect our species by damaging the livestock or crops on which we depend for sustenance. High-level biological phenotypes of microbes, such as pathogenicity, transmissibility, and environmental stability, are complex (1), but they are products of specific microbial sequences encoded within the parasite genomes. Pathogenicity toward one or more host organisms, transmissibility within a species of host organism or between that host organism and vectors (or natural reservoirs), and stability within a specified environment will not be retained if certain sequences are unexpressed.

Responsible parties have been concerned about engineered biothreats for years (1). The increasing technical prowess of synthetic biologists and the burgeoning business of nucleic acid providers have brought the limitations of existing guidance for assessing risk and the adequacy of screening protocols into sharp relief (2). In the past, the “bad microbe” model assessed threat based on pathogens that could pose a severe threat to public health and safety. The “bad microbe” conception has waned, with the “sequence of concern” (SoC) model taking its place (1).

In evaluating SoCs for their risk to public safety, we discovered a dichotomy in sequence annotation. UniProt has a well-curated set of eukaryotic and bacterial toxins, although the targets of those toxins are not always noted (3). Viral parasitism can also be adequately related with the use of existing gene ontology (GO) terms. However,
there are few terms for describing parasitism of hosts as practiced at the molecular level by bacterial, fungal, and protozoal pathogens. What terms there are have few annotations associated with them. Often, the only hint in UniProt that a sequence might be involved in deleterious host-affecting activities was through the tag “GO:0009405 (pathogenesis).” As of June 2021, this term was associated with over 277,000 UniProt accession numbers. Interestingly, the GO:0009405 pathogenesis term has been scheduled for obsolescence, with the final notice given in March 2021 (https://github.com/geneontology/go-annotation/issues/3452).

SoCs are not limited to organisms and toxins on the select agent lists (1). Simply listing the genes of those microbes and toxins would include tens of thousands of innocuous sequences that these parasites share with their close, but nonpathogenic and even nonsymbiotic, relatives (i.e., false positives). This also neglects sequences that cause damage or enable infection from human-disease-causing microbes not deemed serious enough for inclusion on select agent lists (i.e., false negatives). This minireview offers criteria to identify SoCs based on an analysis of more than 2,750 sequences culled from the professional literature for more than 105 bacterial species, 85 viruses, and 25 eukaryotic pathogens. We describe an approach to better characterize these sequences for bioinformatic applications.

WHAT ORGANISMS ENCODE SEQUENCES OF CONCERN?

Of the hundreds of thousands of species of bacteria, fungi, protozoa, worms, and viruses on the planet, only a small percentage have been documented to cause disease in the primate Homo sapiens. It was estimated in 2007 that ~1,400 microbes and parasites can produce disease in humans. Of these, 541 were bacterial, 325 fungal, 285 helminthic, 189 viral, and 57 protozoal (4). Further studies indicated that ~600 fungi can cause disease in humans (5), and well over 200 RNA viruses can infect humans (6), so the total number of human-disease-causing entities is greater than 1,750 and is probably closer to 2,000.

Parasites are distinguished from closely related symbionts by their expression of specific molecules that, when deployed appropriately, can cause a loss of homeostasis (i.e., disease), in a susceptible host. Particular environmental conditions can dispose a host toward greater susceptibility and a parasite toward greater disease-generating ability (7). While many sequences from human-disease-causing microbes have been examined empirically, “the majority…from the microorganisms responsible for the world’s most prevalent diseases remain poorly defined and uncharacterized” (8).

MICROBIAL PATHOGENESIS AND VIRULENCE FACTORS

Practitioners of the biological subspecialty of microbial pathogenesis, a hybrid of cellular biology, molecular biology, and microbiology, investigate the sequences by which microbes exploit host organisms. Perhaps the earliest exploration occurred 50 years ago in swine by Williams Smith and Margaret Linggood. They showed that nonpathogenic Escherichia coli could become an enterotoxigenic pathogen with the introduction of plasmids encoding F4 fimbriae and enterotoxin (9).

Testing a mechanism that directly contributes to pathogenesis makes for the most satisfying investigations. In 2007, experiments were conducted using mice of the same genetic background, while the Citrobacter rodentium bacteria used to infect the mice were varied in which set of up to seven effectors they expressed. The authors showed how the set of sequences expressed rendered the pathogen capable, less capable, or incapable of transmission to a new host and more or less proficient at causing lethal damage (10). Unfortunately, there are more than a few papers declaring a gene product a “virulence factor” after experiments show a “decrease in virulence” following deletion of the gene, though no mechanism can be inferred. In the absence of adequate controls, the gene product in question may simply be necessary to the normal functioning of the organism without necessarily affecting the host.
(i) When “virulence factors” are not sequences of concern. The “virulence factor” appellation is rife in the literature. “Factor” covers carbohydrates, lipids, proteins, and combinations thereof, as well as small RNAs. Encoded virulence factors are *prima facie* candidates for SoCs. However, molecules called virulence factors are not always a threat to a host. Bacterial siderophores are called virulence factors, but most are scavenging molecules without which the bacterium would perish in any environment where metal cofactors are rare. It makes more sense to designate these “virulence lifestyle” sequences (11), or perhaps “proliferative factors.” The less-than-discriminating use of “virulence factor” makes it difficult for investigators to discern what sequences actually harm a host (12). Not all virulence factors are SoCs.

Researchers designations of virulence factors are critical for curators to recognize them, but the less-than-thoughtful use of the nomenclature can create problems for bioinformaticians. An analysis of 2,000 purported virulence factors from over 50 bacterial pathogens found that just 620 were specific to pathogens while 1,368 were common to both pathogens and nonpathogens. The 620 pathogen-specific virulence factors were more likely to reside in pathogenicity islands and be secreted via a secretion system (13). In contrast, the 1,368 “common” virulence factors are probably not SoCs. If put into a reference database of “virulence factors,” they would be false positives. An adequate system for categorizing SoCs should recognize these differences.

(ii) Existing virulence factor data sets and the importance of manual curation of function. Many databases of virulence factors do not curate their sequences according to an established rubric that allows for the extraction of function. The Virulence Factor Database (VFDB) is limited to bacteria pathogenic for humans. The developers eschew manual curation (14). The data set associated with VFDB includes ~3,400 sequences from ~21 bacterial species. No justification is given for the presence of constituent sequences. No curations keyed to individual sequences are provided. The Pathogen-Host Interaction Database (PHI-Base) captures the genetics of pathogen-host interactions from the primary literature along with some functional details, but it principally notes changes in virulence that accompany genetic variants. The effect that these parasite sequences have on the host are of secondary importance (15). The same is true of the Victors database (16). A comparison of bacterium-related databases suggests that functional annotation of SoCs is not a significant concern (17). We think that manual curation is required to adequately annotate the consequences that SoCs have on host processes and enable further advances in computational biology.

**IDENTIFYING AND ASSESSING SEQUENCES OF CONCERN**

There is a chicken-and-egg aspect to identifying SoCs. One must have some idea of what microbial features might be threatening to know what to examine, but it is not until “enough” sequences are perused that the important aspects can be recognized categorically. By reviewing the literature, we discovered sequences that appear important to pathogenesis for parasites of humans, as well as those of animals and plants necessary to human well-being. We have documented over 2,750 of these, which we hope is a fair sample to develop a conceptualization for understanding biothreats.

Assessing sequences of concern for their danger in a bioengineering, gain-of-function (GoF) scenario required us to consider two parameters: (i) the effect on the host, including which host processes are manipulated, and (ii) how directly the sequence exerts its effects. For this minireview, we limit ourselves to reviewing functions of SoCs (FunSoCs) from microbes targeting mammals. The FunSoCs are summarized in Fig. 1 and discussed below. Included as supplemental material is a table of short definitions for the FunSoCs (Data Set S1) and a spreadsheet (Data Set S2) with 220 sequence types from 60 pathogenic species (bacterial, fungal, protozoal, viral) annotated with UniProt accession numbers, FunSoCs, and PubMed identifiers to illustrate our curation.

(i) **What is the effect of the SoC on the host?** (a) Host damage as the sine qua non of pathogenicity. It is generally true that lethal infections are deadly because one or more organs become disabled from cumulative damage. Ascertain the proximal cause of damage can be problematic. Host damage can be the direct result of the
parasite’s action on the host, the host’s reaction to the parasite, or both. While infectious disease theorists of the 20th century once credited the pathogen with unique disease-causing ability, this is no longer tenable (18–20).

Damage is the hallmark of pathogenicity (21). Since this is the case, “toxins” might be said to occupy the preeminent place among virulence factors since they are among the most damaging of molecules deployed by pathogens. In bacteria, toxins are distinguished from damaging effectors in that the former are capable of mediating their own attachment and invasion into a cell, while effectors must be secreted (22).

The term “toxin” is notably nonspecific and amounts to little more than a verbal tag that a molecule is inimical to the life of one or more taxa. But the taxa susceptible to the toxin need to be understood. Alpha-amanitin, bicuculline, carbon monoxide, chlorine gas, ciguatoxin, cyanide, MARTX from Vibrios, ricin, and sarin have disparate modes of action and are all deadly to mammals if administered appropriately. In contrast, the toxins of toxin/antitoxin (TA) systems are not hazardous for mammals; however, they might be administered (23). Of course, toxins do not exhaust the range of damaging biological sequences. The following paragraphs attempt to categorize host damage caused by SoCs.
1. Some SoCs lead to direct killing of a cell by enzymatically compromising a vital process (like translation) or by perforating the membrane via a pore-forming protein or a destabilizing enzymatic process. This includes disabling a cell, as with Shiga toxin from *Shigella/Escherichia coli* (24), or membrane destabilization, as with candidalysin from *Candida albicans* (25, 26). The tripartite HBL enterotoxin from *Bacillus cereus* (27–29), beta toxin from *Clostridium perfringens* (30, 31), leukotoxin from *Aggregatibacter actinomycetemcomitans* (32, 33), VCC from *Vibrio cholerae* (34), VopT (35) and VopV (36) from *Vibrio parahaemolyticus*, and phospholipase A2 from *Vibrio vulnificus* (37) are other such SoCs.

2. Degrading a tissue can be accomplished by proteolysis of the extracellular matrix, loosening the attachments between cells, or liberating a cell from a tissue. The last is sometimes called the “cytopathic effect.” SoCs that accomplish this include aerolysin from *Aeromonas hydrophila* (38), InhA from *Bacillus anthracis* (39), fragilysin from *Bacteroides fragilis* (40), HtrA from *Campylobacter jejuni* (41), candidalysin from *C. albicans* (42), N51 from dengue virus (43), secreted autotransporter toxin (44), the plasmid-encoded toxin (45), cell cycle-inhibiting factor (46), cytolysin A (47), NleA, Map, and EspF (48) from *E. coli*, CagA from *Helicobacter pylori* (49, 50), Mip from *Legionella pneumophila* (51), collagenase A from *Leptospira* (52), Alp1 from *Neosartorya fumigata* (53), MIF from *Plasmodium berghei* (54), exfoliative toxins A and B from *S. aureus* (55), and VopF from *V. cholerae* (56).

3. Disabling an organ system is the severest type of damage. SoCs that accomplish this include ExoU (57) and ExIA (58) from *Pseudomonas aeruginosa*, CARDS toxin from *Mycoplasma pneumoniae* (59), epsilon toxin (60) and iota toxin (61) from *C. perfringens*, edema toxin (62) and lethal toxin (63, 64) from *B. anthracis*, cholera toxin from *V. cholerae* (65), pneumolysin from *Streptococcus pneumoniae* (66), TcdA and TcdB from *Clostridioides difficile* (67), lethal toxin from *Paeniclostridium sordelli* (68, 69), staphylococcal and streptococcal superantigens (70, 71), and NSP4 from rotavirus, a rare example of a viral toxin for mammals (72, 73).

4. SoCs in the class that specifically instigate a damaging inflammatory response appear to directly interact with host components to provoke an inflammatory reaction. These include alpha-hemolysin from *Staphylococcus aureus* (74, 75), PE11 (76) and PE_PGRS17 (77) from *Mycobacterium tuberculosis*, Loa22 from *Leptospira interrogans* (78), pertussis toxin from *Bordetella pertussis* (79, 80), SipA from *Salmonella* (81, 82), ExhC from *Staphylococcus sciuri* (83, 84), Nhaa from *Neisseria meningitidis* (85), GRA24 from *Toxoplasma gondii* (86), VvpM from *Vibrio vulnificus* (87), and nucleocapsid protein (88, 89), spike glycoprotein (90, 91), membrane protein (92), ORF3a (93), and Nsp1 (90) from SARS-CoV. Induction of inflammation can be hard to differentiate from the host reaction to microbial provocation that results in inflammasome activation (94). This class of effectors may require splitting into microbially induced versus microbially provoked host inflammation.

(b) Immune subversion as an essential condition for pathogenicity. Stanley Falkow observed that the avoidance of host defense mechanisms was a feature of disease-causing bacteria (95). Sequences that subvert innate immune pathways are also found in fungal and protozoan parasites and are a universal feature of viruses. Immune systems embody the “wisdom” of hundreds of millions of years of adaptation over which they have had to detect, deflect, and defeat micro- and macroparasites (96–98). More than 6% of all human genes have a role in immunity (99). Immune systems impose layers of molecular and cellular obstacles to thwart invaders that breach epithelial barriers. Parasites survive these host stratagems by employing molecules that mask their presence, mimic and/or misdirect host responses, or simply eliminate immune effectors. Of the SoCs that we documented, ~60% of the viral sequences and ~20% of the bacterial and eukaryotic sequences subvert host immune responses.

Deficits in immune detectors and effectors of a host can render commensal symbionts pathogenic and infections with “nuisance” organisms lethal. Subtle changes in
the sequence of a single host immune effector molecule can mean the difference between life and death during challenge with a parasite (100). The study of human immune deficiencies shows the critical importance of these components of innate immunity for defense against the specific, usually narrow, set of parasites against which they defend (101–103). Many infections run their nonlethal course according to the life cycle of the parasite when facing an average host immune response. These are sometimes called “self-limiting” infections, but a defect or deficiency in a host immune component can abolish the limitation and produce a life-threatening disease.

Of the ~2,000 parasites that can cause disease in humans, the majority are opportunistic: limited to infecting immunocompromised persons (4, 5). The “opportunity” occurs when a proto-parasite encounters an individual whose immune defenses are diminished from (i) loss of barrier function, (ii) congenital immune defects, (iii) infection with HIV, (iv) immune-suppressing pharmacotherapy, or (v) other disease states that alter the homeostasis of the host. These render the host susceptible to microbial parasites that could not successfully establish themselves otherwise. SoCs mediating immune subversion essentially make a host susceptible in the absence of a compromised immune system. Some immune-evading SoCs from Streptococcus are shown in Fig. 2.

1. Suppression of host immune signaling. SoCs that subvert the immune system by disrupting host immune signaling comprise a large set; some subdivisions are listed below.
   a. Disruption of host mitogen-activated protein kinase signaling. Some SoCs work by directly interfering with a component of the host’s mitogen-activated protein kinase signaling pathways (e.g., p38MAPK, JNK, ERK1/2) or a molecule proximal to them. For example, NleD (104) and NleL (105) of E. coli, SptP (106, 107) and SpvC (108, 109) from Salmonella, OspF from Shigella (108,
b. Inhibition of host NF-κB activation. Some SoCs affect IkappaB, RelA, p50, IKK, NEMO, or a molecular constituent proximal to them, for example, AexU from *Aeromonas hydrophila* (114), BopN from *Bordetella* (115, 116), TssM from *Burkholderia pseudomallei* (117), AvrA (118) and GtgA (119) from *Salmonella*, InlC from *Listeria* (120), Ipah1.4, Ipah2.5 (121), Ipah9.8 (122, 123), and OsPG from *Shigella* (124, 125), NleC (126, 127), NleE (128), and NleH1/2 (129, 130) from *E. coli*, MavC from *L. pneumophila* (131), and BPLF1 from Epstein-Barr virus (132).

c. Manipulation of host signaling through tumor necrosis factor (TNF) receptor-associated factor (TRAF). SoCs can manipulate sequences downstream of the TNF receptor and upstream of NF-κB, for example, TssM from *B. pseudomallei* (117), NleB from *Citrobacter rodentium* (128, 133, 134), BPLF1 (132) and LMP1/BNLF1 (135) from Epstein-Barr virus, NleB1 (128, 136) and Tir (137, 138) from *E. coli*, SseK1 from *Salmonella* (128, 139), OspI from *Shigella* (140), GRA7 (141) and GRA15 (142) from *T. gondii*, K7R from vaccinia virus (143, 144), and YopJ from *Yersinia pestis* (145).

d. Disruption of signaling from host Toll-like receptors. Disruption of signaling from Toll-like receptors can occur through alteration of the abundance of the host ligand or receptor, alteration of the ability of the ligand to bind to the receptor, or direct agonism/antagonism of the host receptor or cellular cofactors. SoCs engaging in these activities include PI-PLC from *B. anthracis* (146), envelope glycoprotein from Ebola virus (147), BGLF5 from Epstein-Barr virus (148), PE9-PE10 from *M. tuberculosis* (149), and SsI3 (150) and SsI4 (151) from *S. aureus*.

e. Disruption of host JAK-STAT signaling. Many viral proteins, including NSP2 from Chikungunya virus (152) and ORF6 from severe acute respiratory syndrome coronavirus (SARS-CoV) (153), target the JAK-STAT signaling pathway for antiviral defense.

f. Disruption of host RIG-1 signaling. Keeping RIG-1 inactive through sequestration or targeted destruction of RIG-I or proteins immediately proximal to it via ubiquitination is a function of many viral proteins. The 3C proteinase of human poliovirus cuts host RIG-1 to prevent interferon activation (154).

g. Disruption of host protein kinase R activity. The disruption of host protein kinase R activity can occur by sequestering viral double-stranded RNA (dsRNA), by manipulating the phosphorylation of host elongation factor 2-alpha, and by directly binding host PKR. E3L of vaccinia virus binds viral dsRNA to prevent it from activating of host protein kinase R and OAS (155). NS1 from influenza virus (156) and VP35 from Marburg virus (157) also attenuate antiviral signaling.

h. Inhibition of host STING activity. Both E1A from human adenovirus and E7 from papillomavirus inhibit the cGAS-STING pathway, along with many other viral proteins (158).

2. Resistance to phagocytosis. SoCs interfering with host phagocytosis of microparasites act through a variety of mechanisms, including inhibiting opsonization, manipulating the cytoskeletal dynamics of host phagocytes, and antagonizing phagocyte receptors. These SoCs include AexU from *A. hydrophila* (159), BadA from *Bartonella henselae* (160), AC toxin (161) and BteA (162) from *Bordetella*, OsPB from *Borelia burgdorferi* (163), Hgt1p from *Candida albicans* (164), App1 from *Cryptococcus* (165, 166), GeE from *Enterococcus faecium* (167), EspJ (168) and Ptc (169) from *E. coli*, RodA from *Neosartorya fumigata* (170), ExoS (171) and ExoT (172, 173) from *P. aeruginosa*, aureolysin (174), CHIPS (175), Efb (176, 177), Sbi (178), SCIN (179), and Spa (179, 180) from *S. aureus*, BibA (181), M protein (182), ScpA (183), and Sic (70) from...
Streptococcus, RtxA from *V. vulnificus* (184), VopQ from *V. parahaemolyticus* (185), and PsA (186), YopE (187), YopH, YopO/YpkA, and YopT (188–190) from *Yersinia*.

3. **Resistance to complement-mediated killing.** Host complement effectors can be directly proteolyzed, as by Vag8 of *B. pertussis* (191), or inactivated indirectly, as by CipA from *Acinetobacter baumannii*, which recruits host plasminogen to the bacterial surface (192). BcIA of *B. anthracis* mediates serum resistance by recruiting factor H, a host complement control protein, to the bacterial surface (193).

4. **Resistance to antimicrobial peptides.** Host antimicrobial proteins are cationic peptides that interact with the negatively charged bacterial membrane. They can be destroyed by bacterial proteases, including OmpA from *Klebsiella* (194), ClpX from *B. anthracis* (195), CPAF from *Chlamydia* (196), staphylokinase from *S. aureus* (179), SepA from *Staphylococcus epidermidis* (197), DRS (198), SspA, SspB (199), SpeB, and Sic from *Streptococcus* (200, 201), and OmpU from *V. cholerae* (202).

5. **Resistance to oxidative killing.** Host oxidases can be neutralized by bacterial effectors, including superoxide dismutase from *B. anthracis* (203), SodC from *Coxiella burnetii* (204), KatN from *E. coli* (205), SodC from *Francisella tularensis* (206), and SOK from *S. aureus* (207). The generation of reactive oxygen (or nitrogen) species can also be countered by upstream legerdemain, as with EtPE from *Ehrlichia chaffeensis* (208), Ndk (209), PPE2 (210), PES, PE5, PE15 (211), and PE_PGRS62 (212) from *M. tuberculosis*, SopB from *Salmonella* (213), VopL from *V. parahaemolyticus* (214), and YopH from *Y. pestis* (188, 215).

6. **Countering immunoglobulin.** Parasite effectors can sequester, destroy, or neutralize immunoglobulins by other means, as exemplified by BatB from *Bordetella* (216), IgA1P from *Haemophilus influenzae* (217), IbpA from *Histophilus somni* (218, 219), Sbi (178), Ssl7 (220), Spa (180), and staphylokinase (221) from *S. aureus*, IdeS (222), EndoS (223), SibA (224, 225), and ZmpC (226) from *Streptococcus*, and InvD from *Y. pseudotuberculosis* (227).

7. **Defeat of cytokines.** Pertussis toxin from *B. pertussis*, Lpd from *P. aeruginosa*, CHIPS, Eap, FPRL1 inhibitory protein, and Ssl5 (228) from *S. aureus*, PrpL (229) and SpyCEP (230) from *Streptococcus*, BARF1 from human herpesvirus 4 (HHV-4) (231, 232), and a plethora of orthopoxviral receptors/binding proteins can form associations with host TNF, interleukins, chemokines, and interferons to dysregulate host immune signaling (233–243).

8. **Inhibition of antigen presentation.** Pertussis toxin from *Bordetella* (80), EsxG, EsxH (244), Vpu from HIV-1 (245), ORF66 from HHV-3 (246), BILF1 (247, 248), BNLF2a (249–251), and BZLF1 from HHV-4 (252), E1A and E3 from human adenovirus (253, 254), LpqH (255, 256), LprA (257), LprG (258), and PPE38 (259) from *M. tuberculosis*, SteD from *Salmonella* (260), and IpaH4.5 from *Shigella* (261) inhibit host antigen presentation by various mechanisms.

9. **Resistance to other host immune effectors.** As they are expiring, host neutrophils process their nuclear DNA to create neutrophil extracellular traps (NETs) capable of trapping and killing microparasites. Bacterial nucleases that can counter these include Nuc from *Neisseria gonorrhoeae* (262) and EndA (263), Sda1 (264), and SpnA (265) from *Streptococcus*.

10. **Immunomodulation.** Immunomodulation occurs when a parasite protein directly affects aspects of the host immune system in a fashion that does not suggest an obvious advantage for the parasite relative to the host. Sequences include Hcp from *A. hydrophila* (266), AnKX (267) and LegC4 (268, 269) from *Legionella*, EspC (270), Psts1 (256), PE9, and PE10 (149) from *M. tuberculosis*, and SspH2 from *Salmonella* (271).

(c) **Adherence to the host cell.** To affect the host, symbionts need to either secrete toxins that act while the microbe is at a distance from the host cell or contact host cells or tissues directly. This requires specific adhesin molecules that anchor them, however durably, to the host. Toxins also require adhesins to recognize target cells. Adherence can be to specific host protein receptors, to carbohydrate moieties of glycoproteins or
glycolipids, to membrane cholesterol, and/or to components of the host extracellular matrix. Such proteins are abundant, and host attachment is often just one of their functions (272, 273).

(d) Dissemination in the host. Dissemination factors enable the breaching of host barriers. A breach can happen by proteolytic digestion of tissues or the release of junctional adhesins to allow parasite passage. SoCs that degrade tissue can also be dissemination factors. Examples include ExoS and ExoU from P. aeruginosa (274), InhA from B. anthracis (39, 275), and staphylococcal exfoliative toxins (50, 55, 276).

(e) Host cell invasion. A microsymbiont can "enter" a host cell easily when the host cell is a professional phagocyte, but this happens under conditions unfavorable for symbiont survival. Invasins mediate microbial entry into a range of host cells, including nonphagocytic ones, in ways that allow the parasite a greater probability of reproductive success. Bacterial toxins also possess invasive subunits that enable their entry into host cells; this distinguishes them from effectors, which require a secretion system (22).

(f) Movement in host cell. Movement within a host cell allows a parasite to circumvent host barriers and avoid programmed defenses. Some intracellular bacteria, as well as vaccinia virions, hijack host actin polymerization to propel themselves into adjacent cells. They thus avoid exposure to the hazards of the extracellular milieu (277).

(g) Niche creation in host cells. Some cellular microbial symbionts manipulate host cell processes to create intracellular niches, where they are protected from host destruction and in which they replicate. This has been investigated most thoroughly in Brucella, Chlamydia, Coxiella, Ehrlichia, Legionella, Listeria, Mycobacteria, and Salmonella. SoCs from these bacteria are generally secreted and subvert the normal endosomal and cytoskeletal dynamics of the host cell. Sorting out the mechanisms for these effectors—there are hundreds just in Legionella—is exceedingly complicated, as many are redundant (278).

(ii) How directly does the sequence exert its effect? When considering the ease with which the disease-causing capacity of a pathogen might be enhanced by sequence addition/gain-of-function (GoF), it is important to consider how directly the SoC acts on the host. SoCs that act independently without the need for extra (i.e., secondary or tertiary) sequences would affect virulence more parsimoniously. There are at least four levels of SoC involvement in pathogenesis.

1. Type 1 sequences that directly interact with host molecules to contribute to disease are the most concerning. The SoCs described above (i.e., damage, immune evasion, adherence, invasion, movement, dissemination, niche creation) act directly to produce a specific deleterious effect.
2. Type 2 sequences make or modify molecules that affect the host. These include toxin synthases, enzymes that make capsules rendering bacteria resistant to phagocytosis, and "passive immune evasion" enzymes which alter microbial molecules to protect the possessor from host recognition and/or immune effectors. Examples of the latter include AlmG, a peripheral membrane aminoacyl transferase from V. cholerae that modifies lipopolysaccharide to resist host cationic antimicrobial peptides (279), and Cbu0678 from C. burnetii, which changes the O antigen of lipopolysaccharide (LPS) to decrease immune recognition (280).
3. Type 3 sequences are secretion system components that transport directly acting SoCs to the correct location for function. These include chaperones for the effector proteins.
4. Type 4 sequences are transcription factors regulating the expression of sequences that produce effects directly. While they can be very important for the virulence of a microbe and greatly influence how pathogenic a specific microorganism can be, they might be replaced in a GoF scenario by similar factors.

(iii) What host cellular process is affected? We found it helpful to annotate SoCs with the host processes that they modulate, as these can often be discerned before the biochemical mechanisms are discovered. No fewer than nine aspects of eukaryotic host
cell biology are targeted by parasite proteins for manipulation: (1) transcription, (2) translation, (3) the cell cycle, (4) apoptosis, (5) ubiquitination, (6) small GTPase dynamics, (7) cytoskeleton dynamics, (8) endomembrane dynamics, and (9) autophagy/xenophagy. Viruses tend to manipulate the first five processes, while bacteria, particularly intracellular parasites, affect the final six, with overlap at apoptosis and ubiquitination.

DISCUSSION

Gauging the risks of an emerging pathogen strain or one created through microbial engineering (accidental or otherwise) requires a good comprehension of the pathogenic possibilities of SoCs from natural parasites of humans and livestock. An assessment of existing controlled vocabularies revealed a gap for sequences from nonviral parasites. We documented the role played in disease of over 2,750 parasite proteins from thousands of papers. These were annotated with the FunSoC schema, which categorizes their host-affecting features. The 220 sequences mentioned in this text are provided with full annotations in Data Set S2 in the supplemental material, with definitions provided in Data Set S1.

FunSoCs are tidy enough for human comprehension. For a given SoC, they provide a quick assessment for ~30 host-affecting functions. However, they are insufficiently granular for capturing the molecular details necessary for a comprehensive appreciation of function. We think that these details are better understood with a new adjunct to GO, Pathogen Gene Ontology (PathGO). This resource is being developed by a group at the Johns Hopkins University Applied Physics Laboratory and consists of ~180 terms (https://github.com/jhuapl-bio/pathogenesis-gene-ontology). These are being rooted in biological process and molecular function terms of the Gene Ontology resource (281, 282). We have been suggesting terms and contributing annotations during development. Data Set S2 features a preview of PathGO terms in column F, along with the relevant PubMed ID accession numbers as citations. PathGO will be described in a future publication.

(i) The utility of gain-of-function experiments in microbial pathogenesis. Sometimes eliminating a bacterial sequence suspected of involvement in pathogenicity has no effect. *Legionella pneumophila* exhibits so much functional redundancy in its effectors that the loss of one or two sequences of a certain type may not affect the phenotype (283). Investigators of bacterial adhesion face a similar situation when the suspected adhesin originates in a microbe with multiple ways of associating with a target cell. Researchers circumvent this by studying the adhesin in the background of a specially selected “nonadherent” bacterium (284–289). Experiments in which a sequence “adds” virulence to commensals or avirulent microbes is more interpretable than attempts to ascertain virulence by subtraction from a pathogenic background. The former involves a GoF for the avirulent microbe.

Only a few efforts to make bad bugs worse intentionally have been described (290). However, there are hundreds of publications relating the expression of one or more sequences from an infectious parasite in a heterologous organism. Two dozen of these are noted in column E of Data Set S2. Altered organisms typically display a new property consistent with the suspected pathogenic function of the sequence in the original organism. These GoF experiments are illuminating but can also be problematic (291, 292). The role that a sequence plays in the pathogenicity of a microbe can depend on other proteins and/or the timing of its expression. Simply expressing the sequence in another microbe, even a similar one, is no guarantee that it will perform similarly. The question can be settled only empirically within the limits of the model. The most dramatic example of a GoF experiment with biothreat implications is the notorious mouse interleukin-4 (IL-4) expression in Ectromelia virus that was astoundingly lethal in even vaccinated animals (293). An intriguing bacterial example involves the secreted protease SpyCEP of group A *Streptococcus*. When the nontoxic SpyCEP was expressed in the nonpathogenic bacterium *Lactococcus lactis*, it rendered the cheese-making micute capable of infection in a mouse leg wound model. The SpyCEP protease
degrades the chemokine interleukin-8, which host neutrophils use to coordinate their defense, “sniffing out” bacteria within infected tissues. Interruption of this coordination produced a systemic disease that had lethal consequences for the host within 24 h of inoculation (230).

(ii) Recognized criteria for sequences of concern improve biosecurity. For those worried about either the accidental engineering of pathogens via synthetic biology or the production of bioweapons with enhanced efficacy, a concerning sequence is one that, when transferred to a different microbe, increases the ability of that microbe to damage a susceptible host, increasing the pathological consequences of infection. But, as the cases of SpyCEP and murine IL-4 demonstrate, the disease-causing properties of microbes have interesting dependencies that cannot be understood in the absence of experiments. We think that the criterion of enhanced pathogenicity upon expression in a heterologous nonpathogen is a good starting place for identifying SoCs, but most will not be discovered through such GoF experiments. Our annotation project has demonstrated that there are thousands of microbial sequences that can reasonably be assumed to enhance the pathogenic ability of a heterologous microbe if transferred. In such cases, the disease-causing properties of these sequences are described in the context of the original pathogenic organism and not in a heterologous, nonpathogenic microbe. We assume that these sequences may retain their properties if transferred to a similar microbe. At the very least, it does not seem responsible to assume that they would be innocuous. Documenting these sequences enables them to be recognized via bioinformatics and thus improves biosecurity for those involved in the manufacture of synthetic nucleic acids (2).

Toxins and microbial effectors that damage the human host are of greatest concern. Among these, SoCs that provoke organ failure have the most severe consequences. Next in importance are sequences that subvert host immunity. Noting the host cellular process(es) with which a SoC interacts and how directly it affects host molecules allows a better understanding of its role in microbial pathogenesis. Formalizing these criteria improve recognition of SoCs from the literature, provide the means for distinguishing them by function, and permit the reporting of these functions in bioinformatic applications. We think that the FunSoC vocabulary and data sets annotated with it can be a resource for computational epidemiology, microbial genomics and forensics, DNA synthesis screening, human disease modeling, and biosecurity assessment.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, XLSX file, 0.02 MB.
SUPPLEMENTAL FILE 2, XLSX file, 0.3 MB.

ACKNOWLEDGMENTS

G.D.G. acknowledges Sara R. Nitcher for fruitful discussions on virulence factor curation as well as Barbara J. Mann and Gregory D. Koblenz for comments on the manuscript. He has profited from the extended reflections of L. A. Pirofski and A. Casadevall on what makes bad bugs bad while recognizing that they might object to that formulation. We thank Jim Gibson of Signature Science for his work on the figures. We gratefully acknowledge Jody Proescher, Ron Jacak, and Briana Vecchio-Pagan of the Johns Hopkins University Applied Physics Laboratory for their labor on PathGO, the Pathogen Gene Ontology.

We were either fully or 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 U.S. Government.
REFERENCES

1. National Research Council. 2010. Sequence-based classification of select agents: a brighter line. National Research Council, Washington, DC.

2. Diggans J, Leprout E. 2019. Next steps for access to safe, secure DNA synthesis. Front Bioeng Biotechnol 7:86. https://doi.org/10.3389/fbioe.2019.00086.

3. Jungo F, Bouguerel-L, Xenarios I, Poux S. 2012. The UniProtKB/Swiss-Prot Tox-Prot program: a central hub of integrated venom protein data. Toxicon 60:551–557. https://doi.org/10.1016/j.toxicon.2012.03.010.

4. Woolhouse M, Gaunt E. 2007. Ecological origins of novel human pathogens. Crit Rev Microbiol 33:231–242. https://doi.org/10.1080/10408410701647560.

5. Brown GD, Denning DW, Levitz SM. 2012. Tackling human fungal infections. Science 336:6467. https://doi.org/10.1126/science.1222236.

6. Woolhouse MEJ, Griegley L. 2018. Epidemiological characteristics of human-infective RNA viruses. Sci Data 5:180017. https://doi.org/10.1038/sdata.2018.17.

7. Méthot P-O, Alizon S. 2014. What is a pathogen? Toward a process view of host-parasite interactions. Virulence 5:775–785. https://doi.org/10.4161/21505594.2014.960726.

8. Bouzid M, Hunter PR, Chalmers RM, Tyler KM. 2013. Cryptosporidium pathogenicity and virulence. Clin Microbiol Rev 26:115–134. https://doi.org/10.1128/CMR.00076-12.

9. Smith HW, Linggood MA. 1971. Observations on the pathogenic properties of the K88, Hly and Ent plasmids of porcine diarrhoea. J Med Microbiol 4:467–472. https://doi.org/10.1099/jmm.0.00076.

10. Wickham ME, Brown NF, Boyle EC, Coombes BK, Finlay BB. 2007. Virus exclusion is positively selected by transmission success between mammalian hosts. Curr Biol 17:783–788. https://doi.org/10.1016/j.cub.2007.03.067.

11. Wassenaar TM, Gaaster W. 2001. Bacterial virulence: can we draw the line? FEMS Microbiol Lett 201:1–7. https://doi.org/10.1111/j.1574-6968.2001.tb0724x.x.

12. Segura M, Fittipaldi N, Calzca G, Gottschalk M. 2017. Critical Streptococcus suis virulence factors: are they all really critical? Trends Microbiol 25:585–599. https://doi.org/10.1016/j.tim.2017.02.005.

13. Niu C, Yu D, Wang Y, Ren H, Jin Y, Zhou W, Li B, Cheng Y, Yue J, Gao Z, Liang L. 2013. Common and pathogen-specific virulence factors are different in function and structure. Virulence 4:473–482. https://doi.org/10.4161/viru.25730.

14. Liu B, Zheng D, Jin Q, Chen L, Yang J. 2019. VFBDB 2019: a comparative web-based knowledge base of virulence factors in human and animal agents: a brighter line. National Research Council, Washington, DC.

15. Afouras N, Martinez MC, Pedraza-Hernández SE, Leipziger J, Praetorius HA, Skals M. 2012. Leukotoxin from Clostridium perfringens selective channels in lipid bilayers. Infect Immun 80:5546–5551. https://doi.org/10.1128/IAI.0022034510363682.

16. Kosack KE. 2020. PHI-base: the pathogen-host interactions database.

17. Wickham ME, Brown NF, Boyle EC, Coombes BK, Finlay BB. 2007. Virus exclusion is positively selected by transmission success between mammalian hosts. Curr Biol 17:783–788. https://doi.org/10.1016/j.cub.2007.03.067.

18. Casadevall A, Pirofski L. 2003. The damage-response framework of microbial pathogenesis. Nat Rev Microbiol 1:17–24. https://doi.org/10.1038/nrmicro732.

19. Casadevall A, Pirofski L. 2004. Microbiology: ditch the term pathogen. Nature 516:165–166. https://doi.org/10.1038/516165a.

20. Pirofski L, Casadevall A. 2015. What is infectiveness and how is it involved in infection and immunity? BMC Immunol 16:13. https://doi.org/10.1186/s12865-015-0076-1.

21. Pirofski L-A, Casadevall A. 2018. The damage-response framework as a tool for the physician-scientist to understand the pathogenesis of infectious diseases. J Infect Dis 218:57–511. https://doi.org/10.1093/infdis/jiy083.

22. Galán JE. 2009. Common themes in the design and function of bacterial effectors. Cell Host Microbe 5:571–579. https://doi.org/10.1016/j.chom.2009.04.008.

23. Song S, Wood TK. 2018. Post-segregational killing and phage inhibition are not mediated by cell death through toxin/antitoxin systems. Front Microbiol 9:814. https://doi.org/10.3389/fmicb.2018.00814.
Minireview Infection and Immunity

Gastrointestinal Liver Physiol 279:G1129–G1134. https://doi.org/10.1152/ajpgi.2000.279.6.G1129

41. Boehm M, Hoy B, Rohde M, Tegtmeyer N, Baik KT, Oyarzabal OA, Brandedt L, Wesseler S, Backert S. 2012. Rapid paracellular transmigration of Campylobacter jejuni across polarized epithelial cells without affecting TER: role of proteolytic-active HtrA cleaving E-cadherin but not fibronectin. Gut Pathog 4:3. https://doi.org/10.1186/1757-4943-4.3

42. Allert S, Förster TM, Svensson C-M, Richardson JP, Pawlik T, Hebecker B, Rupprecht S, Jurachmitz M, Schaller M, Blagojevic M, Morchhaußner J, Finde GT, Jacobsen ID, Naglik JR, Kasper L, Mogensen JS, Hube B. 2018. Candida albicans-induced epithelial damage mediates translocation through intestinal barriers. mBio 9:e00195-18. https://doi.org/10.1128/mbio.00195-18

43. Chen H-R, Lai Y-C, Yeh T-M. 2018. Dengue virus non-structural protein 1: a pathogenic factor, therapeutic target, and vaccine candidate. J Biomed Sci 25:58. https://doi.org/10.1186/s12929-018-0462-0

44. Guignot J, Chaplais C, Cocomnier-Poitier M-H, Servin AL. 2007. The secreted autotransporter toxin, Sat, functions as a virulence factor in Afa/Dr diffusely adhering Escherichia coli by promoting lesions in tight junction of polarized epithelial cells. Cell Microbiol 9:204–221. https://doi.org/10.10111/1462-5822.2006.00782.x

45. Villaseca JM, Navarro-García F, Mendoza-Hernández G, Nataro JP, Cravioto A, Eslava C. 2000. Pet toxin from entericaggregative Escherichia coli produces cellular damage associated with fodrin disruption. Infect Immun 68:5920–5927. https://doi.org/10.1128/IAI.68.12.5920-5927.2000

54. Maslansky O, Ledger TN, Boursy M, Ohara M, Tu X, Goffaux F, Mainj R, Rosenlone I, Sugai M, De Rycke J, Oswald E. 2003. Enteropathogenic and enterohemorrhagic Escherichia coli deliver a novel effector called Cif, which blocks cell cycle G2/M transition. Mol Microbiol 50:1553–1567. https://doi.org/10.1046/j.1365-2958.2003.03821.x

55. Oscarsson J, Mizuno Y, Li L, Xai WH, Wieslander A, Uhlin BE. 1999. Molecular analysis of the cytolytic protein ClyA (SheA) from Escherichia coli. Mol Microbiol 32:1226–1238. https://doi.org/10.1046/j.1365-2958.1999.01435.x

56. Thanabalasuir A, Koutoursis A, Welenl M, Mice M, Hecht G, Gruenheid S. 2010. The bacterial virulence factor NleA is required for the disruption of intestinal tight junctions by enteropathogenic Escherichia coli. Cell Microbiol 12:31–41. https://doi.org/10.1111/j.1462-5822.2009.01376.x

57. Appa VR, Vogelmull R, Covacci A, Topkins LS, Nelson WJ, Falkow S. Disruption of the epithelial apical-junctional complex by Helicobacter pylori. Science 300:1430–1434. https://doi.org/10.1126/science.1081919

58. Sousa S, Leucit M, Cossart P. 2005. Microbial strategies to target, cross or disrupt epithelia. Curr Opin Cell Biol 17:489–498. https://doi.org/10.1016/j.cceb.2005.08.013

59. Wagner C, Khan AS, Kamphausen T, Schmausser B, Unal C, Lorenz U, Finkel T, Leppla SH. 2013. Key tissue targets responsible for anthrax-toxin-induced lethality. Nature 501:63–68. https://doi.org/10.1038/nature12510

60. Savva CG, Clark AR, Naylor CE, Popoff MR, Moss DS, Basak AK, Tilby RW, Bokori-Boni M. 2019. The pore structure of Clostridium perfringens epsilon toxin. Nat Commun 10:2641. https://doi.org/10.1038/s41467-019-10645-8

61. Redondo LM, Redondo EA, Daillo GC, Leiva CL, Diaz-Carrasco JM, Bruzzoze OA, Cangelosi A, Goegohegan P, Fernandez-Mayekea M. 2017. Effects of Clostridium perfringens iota toxin in the small intestine of mice. Anaerobe 48:83–88. https://doi.org/10.1016/j.anaero.2017.07.007

62. Firoved AM, Miller GC, Moayeri M, Kakkar R, Shen Y, Wiggins JF, McNally EM, Tang W-J, Leppa SP. 2005. Bacillus anthracis edema toxin causes extensive tissue lesions and rapid lethality in mice. Am J Pathol 167:1309–1320. https://doi.org/10.1016/S0002-9440(06)61218-7

63. Liu S, Zhang Y, Moayeri M, Liu J, Crown D, Fattah RJ, Wein AN, Yu Z-X, Finkel T, Leppa SH. 2013. Key tissue targets responsible for anthrax-toxin-induced lethality. Nature 501:63–68. https://doi.org/10.1038/nature12510

64. Hutt JA, Lovchik JA, Drysdale M, Sherwood RL, Brasel T, Lipscomb MF, Lyons CR. 2014. Lethal factor, but not edema factor, is required to cause fatal anthrax in cynomolgus macaques after pulmonary spore challenge. Am J Pathol 184:3205–3216. https://doi.org/10.1016/j.ajpath.2014.08.008

65. Sanchez J, Holmgren J. 2005. Virulence factors, pathogenesis and vaccine protection in choleragenoid and ETEC diarrhoea. Curr Opin Immunol 17:388–398. https://doi.org/10.1016/j.coi.2005.06.007

66. Alhamdi Y, Neill DR, Abrams ST, Malak HA, Yahya R, Barrett-Jolley R, Wang G, Kadioglu A, Toh C-H. 2015. Circulating pneumolysin is a potent inducer of cardiac injury during pneumococcal infection. PLoS Pathog 11:e1004836. https://doi.org/10.1371/journal.ppat.1004836

67. Janaki G, Giesman T, Aktories K. 2007. Rho-glucosylating Clostridium difficile toxins A and B: new insights into structure and function. Glycobiochem 17:15–22. https://doi.org/10.1093/glycob/cwm004

68. Geny B, Khan H, Fitting C, Zarantonelli L, Mazuet C, Cayet N, Sazantov M, Prevost M-C, Cavallion J-M, Huere M, Popoff MR. 2007. Clostridium sordelli lethal toxin kills mice by inducing a major increase in lung vascular permeability. Am J Pathol 170:1003–1017. https://doi.org/10.2353/ajpath.2007.06053c

69. Hao Y, Sen T, Opp JS, Young VB, Thiele T, Srinivas G, Huang SK, Aronoff DM. 2010. Lethal toxin is a critical determinant of rapid mortality in rodent models of Clostridium sordelli endometritis. Anaerobe 16:155–160. https://doi.org/10.1016/j.anaero.2009.06.002

70. Bisno AL, Brito MO, Collins CM. 2003. Molecular basis of group A streptococcal virulence. Lancet Infect Dis 3:191–200. https://doi.org/10.1016/S1473-3099(03)00576-0

71. Tarkowski A, Collins LV, Jonsson I-M, Eriksson K, Sakiniene E, Verdrengh M. 2002. Analysis of pulmonary in...
92. Wang Y, Liu L. 2016. The membrane protein of severe acute respiratory syndrome coronavirus functions as a novel cytosolic pathogen-associated molecular pattern to promote beta interferon induction via a Toll-like receptor-related TRAF3-independent mechanism. mBio 7:e01872-15.

93. Li J, Vinh DC, Casanova J-L, Puel A. 2017. Inborn errors of immunity underlying fungal diseases in otherwise healthy individuals. Curr Opin Immunol 40:46–57. https://doi.org/10.1016/j.coi.2017.01.016.

94. Schnappauf O, Chae JJ, Kastner DL, Aksentijevich I. 2019. The pyrin domain protein, GRA24, modulates the early immune response to infection by inhibiting Raf activation. Cell Microbiol 5:267–275. https://doi.org/10.1111/j.1462-202X.2007.00734.x.

95. Guiney DG. 2005. The role of host cell death in Salmonella infections. Curr Microbiol 49:131–150. https://doi.org/10.1007/s00284-005-0731-2.

96. Lovley DR. 2008. Microbial catabolism of complex organic compounds. Science 322:1318–1322. https://doi.org/10.1126/science.1166197.

97. Kasamatsu J. 2013. Evolution of innate and adaptive immune systems in jawless vertebrates. Microbiol Immunol 57:1–12. https://doi.org/10.1111/j.1348-0421.2012.00500.x.

98. Viljakainen L. 2015. Evolutionary genetics of insect innate immunity. Brief Funct Genomics 14:407–412. https://doi.org/10.1093/bfgp/elv002.

99. Abbas AR, Baldwin D, Ma Y, Ouyang W, Gurney A, Martin F, Fong S, van Lookeren Campagne M, Godowski P, Williams PM, Chan AC, Clark HF. 2005. Immune response in silico (IRIS): immune-specific genes identified from a compilation of microarray expression data. Genes Immun 6:319–331. https://doi.org/10.1038/sj.gene.6364173.

100. Andoniou CE, Sutton VR, Wikstrom ME, Fleming P, Thyt KYT, Matthews AY, Kaiserman D, Schuster IS, Coudert JD, Eldi P, Chaudhri G, Karupiah G, Bird PI, Tranap J, Delegi-estopi MA. 2014. A natural genetic variant of granzyme B confers lethality to a common viral infection. PLoS Pathog 10:e1004526. https://doi.org/10.1371/journal.ppat.1004526.

101. Casanova J-L. 2015. Human genetic basis of interindividual variability in the course of infection. Proc Natl Acad Sci U S A 112:E7118–E7127. https://doi.org/10.1073/pnas.1521644112.

102. Casanova J-L. 2015. Severe infectious diseases of childhood as monogenic inborn errors of immunity. Proc Natl Acad Sci U S A 112:E7127–E7137. https://doi.org/10.1073/pnas.1521651112.

103. Li J, Vinh DC, Casanova J-L, Puel A. 2017. Inborn errors of immunity underlying fungal diseases in otherwise healthy individuals. Curr Opin Immunol 40:46–57. https://doi.org/10.1016/j.coi.2017.01.016.

104. Baruch K, Gur-Arie L, Nadler C, Koby S, Yerushalmi G, Ben-Neria Y, Yogev O, Shaulian E, Gutman C, Zarihav R, Rosenshine I. 2011. Metallo-protease type III effectors that specifically cleave JNK and NF-κB. EMBO J 30:221–231. https://doi.org/10.1038/emboj.2010.297.

105. Sheng X, You Q, Zhu H, Chang Z, Li Q, Wang H, Wang C, Wang H, Hui L, Du C, Xie Y, Zeng R, Lin A, Shi D, Ruan K, Yan J, Gao GF, Shao F, Hu R. 2017. Bacterial effector NleL promotes enterohemorrhagic E. coli-induced attaching and effacing lesions by ubiquitylating and inactivating JNK. PLoS Pathog 13:e1006534. https://doi.org/10.1371/journal.ppat.1006534.

106. Lin SL, Le TX, Cowen DS. 2003. SpP, a Salmonella typhimurium type III secreted protein, inhibits the mitogen-activated protein kinase pathway by inhibiting Raf activation. Cell Microbiol 5:267–275. https://doi.org/10.1046/j.1462-202X.2003.00274.x.

107. Guiney DG. 2005. The role of host cell death in Salmonella infections. Curr Microbiol 49:131–150. https://doi.org/10.1007/s00284-005-0731-2.

108. Li X, Xu H, Zhou Y, Zhang J, Long C, Li S, Chen S, Zhou J-M, Shao F. 2007. The phosphorelay ElsA activates interleukin-6 expression through cellular transcription factor NF-kappaB. Virology 365:324–335. https://doi.org/10.1016/j.virol.2007.04.009.

109. Law AHY, Lee DCW, Cheung BKW, Yim HCH, Lau ASY. 2007. Role for non-structural protein 1 of severe acute respiratory syndrome coronavirus in chemokine dysregulation. J Virol 81:416–422. https://doi.org/10.1128/JVI.02336-05.

110. dosch SF, Mahajan SD, Collins AR. 2009. SARS coronavirus spike protein-induced innate immune response occurs via activation of the NF-kappaB pathway in human monocyte macrophages in vitro. Virus Res 142:19–27. https://doi.org/10.1016/j.virusres.2009.01.005.

111. Wang Y, Liu L. 2016. The membrane protein of severe acute respiratory syndrome coronavirus functions as a novel cytokotic pathogen-associated molecular pattern to promote beta interferon induction via a Toll-like-receptor-activated TRAF3-independent mechanism. mBio 7:e01872-15. https://doi.org/10.1128/mBio.01872-15.

112. Siu K-L, Yuen K-S, Castano-Rodriguez C, Ye Z-W, Yeung M-L, Fung S-Y, Yuan S, Chan C-P, Yuen K-Y, Enjuanes L, Jin D-Y. 2019. Severe acute respiratory syndrome coronavirus ORF3a protein activates the NLRP3 inflammasome by promoting TRAF3-dependent ubiquitination of ASC. FASEB J 33:8865–8877. https://doi.org/10.1096/fj.201802418R.

113. Mohamed MR, McFadden G. 2009. NF-κB inhibitors: strategies from poxviruses. Curr Opin Microbiol 12:315–3132. https://doi.org/10.1016/j.cub.2009.11.010.

114. Melillo A, Conti L, Liotta CL. 2012. Mutations in the NF-κB responsive element of the Survivin promoter abolish the ability of NF-κB to promote Survivin expression. J Immunol 189:2887–2898. https://doi.org/10.4049/jimmunol.1103476.

115. Rueda F, de la Torre F, Sánchez-Criado MA, Martin F, Martin F, Yanez-Mtnez A, Fuentes J, Martín F, Yanez-Mtnez A, Fuentes J. 2012. The NF-κB inhibitors A20 and RIP140 inhibit the respiratory burst of human granulocytes. J Biol Chem 287:3935–3945. https://doi.org/10.1074/jbc.M111.282028.

116. Zanetti E, de la Torre F, Sánchez-Criado MA, Martin F, Martin F, Yanez-Mtnez A, Fuentes J, Martín F, Yanez-Mtnez A, Fuentes J. 2013. The NF-κB inhibitors A20 and RIP140 inhibit the respiratory burst of human granulocytes. J Biol Chem 287:3935–3945. https://doi.org/10.1074/jbc.M111.282028.
150. Yokoyama R, Itoh S, Kamoshida G, Takii T, Fujii S, Tsuji T, Onozaki K. 2012. Staphylococcus superantigen-like protein 3 binds to the Toll-like receptor 2 extracellular domain and induces cytokine production induced by Staphylococcus aureus, cell wall component, or lipopeptides in murine macrophages. Infect Immun 80:2816–2825. https://doi.org/10.1128/IAI.00399-12.

151. Bardocz BW, Vos R, Bouman T, Aerts PC, Bestebroer J, Huizinga EG. 2012. Evasion of Toll-like receptor 2 activation by staphylococcal superantigen-like protein 3. J Mol Microbiol Biotechnol 21:1120–1125. https://doi.org/10.1159/000341966.

152. Fros JJ, Liu WJ, Prow NA, Geertsema C, Ligtenberg M, Vanlandingham D. 2018. The role of Borrelia burgdorferi secreted from the host cell response. J Exp Med 200:1267–1276. https://doi.org/10.1084/jem.20172534.

153. Barral PM, Sarkar D, Fisher PB, Racaniello VR. 2009. RIG-I is cleaved during picornavirus infection. Biochemistry 43:6100–6104. https://doi.org/10.1021/bi901477x.

154. Hammel M, Sfyroera G, Ricklin D, Magotti P, Lambris JD, Geisbrecht BV. 2007. A structural basis for complement inhibition by Staphylococcus aureus. Nat Immunol 8:643–653. https://doi.org/10.1038/ni0707-643.

155. Lee LYL, Höök M, Haviland D, Wetsel RA, Yonter EO, Syribeys P, Vernachio RA, Gouda H, Torigoe H, Saito A, Sato M, Arata Y, Shimada I. 1992. Three-dimensional solution structure of the B domain of staphylococcal protein A: comparisons of the solution and crystal structures. Biochemistry 31:9665–9672. https://doi.org/10.1021/bi00155a020.

156. Hartiala P, Hytönen J, Suhonen J, Leppäranta O, Tuominen-Gustafsson E, Chatterjee S, Caramalho R, Haas H, Lass-Förl C, Shaughnessy J, Ram S, Gow NAR, Orth-Höller D, Würzner M, Speth C, Rambach G, Binder U, Chernet I. 2012. Strep-Affanex inhibits opsonophagocytosis. Cell Microbiol 8:1281–1288. https://doi.org/10.1111/j.1462-2920.2006.00725.x.

157. Marchès O, Covarelli V, Dahan S, Cougoule C, Bhatta P, Frankel G, Caron A, Morace G, Luberto C, Tomlinson S, Del Poeta M. 2009. App1: an antiphagocytic protein. J Clin Invest 112:1080–1094. https://doi.org/10.1172/JCI18309.
229. Bradshaw JL, Caballero AR, Bierdeman MA, Adams KV, Pipkins HR, Tang A, Oggioni MR, Memmi G, Maggi T, Chiavolini D, Iannelli F, Pozzi G. 2003. EndoS, an immunomodulatory proteinase encoded by an orthopoxvirus IFN-gamma-binding protein. Proc Natl Acad Sci U S A 100:3734. https://doi.org/10.1073/pnas.103734.

230. Oggioni MR, Memmi G, Maggi T, Chiavolini D, Iannelli F, Pozzi G. 2003. EndoS, an immunomodulatory proteinase encoded by an orthopoxvirus IFN-gamma-binding protein. Proc Natl Acad Sci U S A 100:3734. https://doi.org/10.1073/pnas.103734.

231. Gehring AJ, Dobos KM, Belisle JT, Harding CV, Boom WH. 2004. PPE38 protein of Mycobacterium tuberculosis: Structure of CrmE, a virus-encoded tumour necrosis factor receptor. Biochemistry (Mosc) 74:1356–1362. https://doi.org/10.1134/S0006297900120098.

232. Alcami A, Koszinowski UH. 2000. Viral mechanisms of immune evasion. Annu Rev Immunol 21:377–408. https://doi.org/10.1146/annurev.immunol.21.3.377.

233. Ohashi M, Fogg MH, Orlova N, Quink C, Wang F. 2012. An Epstein-Barr virus-encoded BILF1 receptor for signaling and major histocompatibility complex class I downregulation. mBio 3:e00170-18. https://doi.org/10.1128/mBio.00170-18.

234. Douglas JL, Gustin JK, Viswanathan K, Mansouri M, Moses AV, Früh K. 2010. The great escape: viral strategies to counter BST-2/tetherin. PLoS Pathog 6:e1000913. https://doi.org/10.1371/journal.ppat.1000913.

235. Alcami A, Koszinowski UH. 2000. Viral mechanisms of immune evasion. Annu Rev Immunol 21:377–408. https://doi.org/10.1146/annurev.immunol.21.3.377.

236. Alcami A, Koszinowski UH. 2000. Viral mechanisms of immune evasion. Annu Rev Immunol 21:377–408. https://doi.org/10.1146/annurev.immunol.21.3.377.

237. Gileva IP, Nepomnyashchikh TS, Yazikov IA, Shchelkunov SN. 2009. Recombinant TNF-binding protein from viral varicella-zoster virus. Microbes and Infection 11:e1003095. https://doi.org/10.1016/j.micinf.2011.05.006.

238. Graham SC, Bahar MW, Abrescia NG, Smith GL, Stuart DJ, Grimes JM. 2007. Structure of CrmE, a virus-encoded tumour necrosis factor receptor. J Mol Biol 372:660–671. https://doi.org/10.1016/j.jmb.2007.06.082.

239. Haga IR, Bowie AG. 2005. Evasion of innate immunity by vaccinia virus. Parasitology 130(suppl):S11–S25. https://doi.org/10.1017/S0031182005000812.

240. Naicker AA, Walters R, Biedermann MA, Adams KV, Pipkins HR, Tang A, Oggioni MR, Memmi G, Maggi T, Chiavolini D, Iannelli F, Pozzi G. 2003. EndoS, an immunomodulatory proteinase encoded by an orthopoxvirus IFN-gamma-binding protein. Proc Natl Acad Sci U S A 100:3734. https://doi.org/10.1073/pnas.103734.

241. Monizani, J. A., ten Engelsen, A.-C., and A. P. H. 2011. Glysosaminoglycans mediate retention of the poxvirus type I interferon binding protein at the cell surface to locally block interferon antiviral responses. FASEB J 25:1960–1971. https://doi.org/10.1096/fj.10-171788.

242. Bremer A, Hahn BS, Hanski J, Lepisti E, Qvist J, Riemann H, Stumpe M, Wolder H, Yoo S, A. R. P. H. 2009. The interferon system and vaccinia virus evasion mechanisms. J Interferon Cytokine Res 29:581–598. https://doi.org/10.1089/jir.2009.0073.

243. Seet ST, Johnston JB, Bremer A, Barrett AW, Everett H, Cameron C, Sugiyama C. 2009. Evasion of innate and adaptive immunity. Annu Rev Immunol 21:377–423. https://doi.org/10.1146/annurev.immunol.21.120601.141049.

244. Haga IR, Bowie AG. 2005. Evasion of innate immunity by vaccinia virus. Para- sitology 130(suppl):S11–S25. https://doi.org/10.1017/S0031182005000812.

245. Bachmann MF, Moss RB, Herzenberg LA, Hsu P, Ljunggren HE, Herzenberg LA. 1990. APC-de-...
Nishifuji K, Fudaba Y, Yamaguchi T, Iwasaki T, Sugai M, Amagai M. 2005.

Shaver CM, Hauser AR. 2004. Relative contributions of

Bhavsar AP, Brown NF, Stoepel J, Wiermer M, Martin DDO, Hsu KJ, Imami

Guo Q, Bi J, Li M, Ge W, Li M, Ge W, Xu Y, Fan W, Wang H, Zhang X. 2019. ESX secre-

Hasegawa T, Minami M, Okamoto A, Tatsuno I, Isaka M, Ohta M. 2010.

Otsubo R, Mimuro H, Ashida H, Hamazaki J, Murata S, Sasakawa C. 2019.

Beiter K, Wartha F, Allbiger B, Normark S, Zychlinsky A, Henriques-Normark B. 2006. An endonuclease allows Streptococcus pneumoniae to escape from neutrophil extracellular traps. Curr Biol 16:401–407. https://doi.org/10.1016/j.cub.2006.01.056.

Liu M, Zhu H, Li J, Garcia CC, Feng W, Kirpotina LN, Hilmer J, Tavares LP, Layton AW, Quinn MT, Bothner B, Teixeira MM, Lei B. 2012. Group A Streptococcus secreted esterase hydrolyzes platelet-activating factor to promote neutrophil extracellular trap formation. Mol Microbiol 84:1254–1264. https://doi.org/10.1111/j.1365-2958.2012.07694.x.

Hasegawa T, Minami M, Okamoto A, Tatsuno I, Isaka M, Ohta M. 2010. Characterization of a virulence-associated and cell-wall-located DNase of Streptococcus pyogenes. Microbiology (Reading) 156:184–190. https://doi.org/10.1099/mic.0.031955-0.

Suarez G, Sierra JC, Kirtley ML, Chopra AK. 2010. Role of Hcp, a type 6 secretion system effector, of Aeromonas hydrophila in modulating activation of host immune cells. Microbiology (Reading) 156:3676–3688. https://doi.org/10.1099/mic.0.041277-0.

Yu X, Noll RR, Romero Dueñas BP, Allgood SC, Barker K, Caplan JL, Machner MP, LaBaer J, Qiu J, Neunuebel MR. 2018. Legionella effector AnkK interacts with host nuclear protein PLEKH1. BMC Microbiol 18:51. https://doi.org/10.1186/s12868-017-1417-7.

Shames SR, Liu L, Havey JC, Schofield WB, Goodman AL, Roy CR. 2017. Multiple Legionella pneumonia effector virulence phenotypes revealed through high-throughput analysis of targeted mutant libraries. Proc Natl Acad Sci U S A 114:E10446–E10454. https://doi.org/10.1073/pnas.1708553114.

Ngwaga T, Hydock AJ, Ganesan S, Shames SR. 2019. Potentiation of cyto-

Guo Q, Bi J, Li M, Ge W, Xu Y, Fan W, Wang H, Zhang X. 2019. ESX secre-

Shaver CM, Hauser AR. 2004. Relative contributions of Pseudomonas aer-

Tu Y, Noll RR, Romero Dueñas BP, Allgood SC, Barker K, Caplan JL, Machner MP, LaBaer J, Qiu J, Neunuebel MR. 2018. Legionella effector AnkK interacts with host nuclear protein PLEKH1. BMC Microbiol 18:51. https://doi.org/10.1186/s12868-017-1417-7.

Schroeder GN. 2017. The toolbox for uncovering the functions of Legionella Dot/Icm type Iv secretion system effectors: current state and future directions. Front Cell Infect Microbiol 7:201. https://doi.org/10.3389/fcimb.2017.00201.

Serruto D, Adu-Bobie J, Scarselli M, Veggi D, Pizza M, Rappuoli R, Aricò B. 2003. Neisseria meningitidis Aap, a new adhesin with autotaxin-dependent serine protease activity. Mol Microbiol 48:323–334. https://doi.org/10.1046/j.

Sitkiewicz I, Nagiec MJ, Sumby P, Butler SD, Cywes-Bentley C, Musser JM. 2016. The Bs1-like protein rOmpB of spotted fever group rickettsiae functions in the rickettsial invasion of Vero cells. Microbes Infect 18:619–6205. https://doi.org/10.1016/j.micinf.2014.11.024.

Schroeder GN. 2017. The toolbox for uncovering the functions of Legionella Dot/Icm type Iv secretion system effectors: current state and future directions. Front Cell Infect Microbiol 7:201. https://doi.org/10.3389/fcimb.2017.00201.

Serruto D, Adu-Bobie J, Scarselli M, Veggi D, Pizza M, Rappuoli R, Aricò B. 2003. Neisseria meningitidis Aap, a new adhesin with autotaxin-dependent serine protease activity. Mol Microbiol 48:323–334. https://doi.org/10.1046/j.

Shaver CM, Hauser AR. 2004. Relative contributions of Pseudomonas aer-

Tu Y, Noll RR, Romero Dueñas BP, Allgood SC, Barker K, Caplan JL, Machner MP, LaBaer J, Qiu J, Neunuebel MR. 2018. Legionella effector AnkK interacts with host nuclear protein PLEKH1. BMC Microbiol 18:51. https://doi.org/10.1186/s12868-017-1417-7.

Shavkarova TP, Drost J, Wierman T, Prud'Homme J, Cooney EP, Natarajan R, Anderson B, Ramamoorthy A, Schumacher LT. 2018. Multiple Legionella pneumonia effector virulence phenotypes revealed through high-throughput analysis of targeted mutant libraries. Proc Natl Acad Sci U S A 114:E10446–E10454. https://doi.org/10.1073/pnas.1708553114.

Ngwaga T, Hydock AJ, Ganesan S, Shames SR. 2019. Potentiation of cyto-

Guo Q, Bi J, Li M, Ge W, Xu Y, Fan W, Wang H, Zhang X. 2019. ESX secre-

Shaver CM, Hauser AR. 2004. Relative contributions of Pseudomonas aer-

Tu Y, Noll RR, Romero Dueñas BP, Allgood SC, Barker K, Caplan JL, Machner MP, LaBaer J, Qiu J, Neunuebel MR. 2018. Legionella effector AnkK interacts with host nuclear protein PLEKH1. BMC Microbiol 18:51. https://doi.org/10.1186/s12868-017-1417-7.

Schroeder GN. 2017. The toolbox for uncovering the functions of Legionella Dot/Icm type Iv secretion system effectors: current state and future directions. Front Cell Infect Microbiol 7:201. https://doi.org/10.3389/fcimb.2017.00201.

Serruto D, Adu-Bobie J, Scarselli M, Veggi D, Pizza M, Rappuoli R, Aricò B. 2003. Neisseria meningitidis Aap, a new adhesin with autotaxin-dependent serine protease activity. Mol Microbiol 48:323–334. https://doi.org/10.1046/j.

Shavkarova TP, Drost J, Wierman T, Prud'Homme J, Cooney EP, Natarajan R, Anderson B, Ramamoorthy A, Schumacher LT. 2018. Multiple Legionella pneumonia effector virulence phenotypes revealed through high-throughput analysis of targeted mutant libraries. Proc Natl Acad Sci U S A 114:E10446–E10454. https://doi.org/10.1073/pnas.1708553114.

Ngwaga T, Hydock AJ, Ganesan S, Shames SR. 2019. Potentiation of cyto-

Guo Q, Bi J, Li M, Ge W, Xu Y, Fan W, Wang H, Zhang X. 2019. ESX secre-

Shaver CM, Hauser AR. 2004. Relative contributions of Pseudomonas aer-

Tu Y, Noll RR, Romero Dueñas BP, Allgood SC, Barker K, Caplan JL, Machner MP, LaBaer J, Qiu J, Neunuebel MR. 2018. Legionella effector AnkK interacts with host nuclear protein PLEKH1. BMC Microbiol 18:51. https://doi.org/10.1186/s12868-017-1417-7.

Schroeder GN. 2017. The toolbox for uncovering the functions of Legionella Dot/Icm type Iv secretion system effectors: current state and future directions. Front Cell Infect Microbiol 7:201. https://doi.org/10.3389/fcimb.2017.00201.

Serruto D, Adu-Bobie J, Scarselli M, Veggi D, Pizza M, Rappuoli R, Aricò B. 2003. Neisseria meningitidis Aap, a new adhesin with autotaxin-dependent serine protease activity. Mol Microbiol 48:323–334. https://doi.org/10.1046/j.

Shavkarova TP, Drost J, Wierman T, Prud'Homme J, Cooney EP, Natarajan R, Anderson B, Ramamoorthy A, Schumacher LT. 2018. Multiple Legionella pneumonia effector virulence phenotypes revealed through high-throughput analysis of targeted mutant libraries. Proc Natl Acad Sci U S A 114:E10446–E10454. https://doi.org/10.1073/pnas.1708553114.
Gene D. Godbold began working as a government contractor in chemical and biological warfare defense in 2002 after his postdoctoral studies in infectious diseases at the University of Virginia. He has been discerning and annotating sequences of concern (SoCs) for data sets since 2005. His professional enthusiasms include (i) extracting sequence function from the literature in a (semi)automated fashion and (ii) using annotated genome sequences with epidemiological parameters and human immune modeling to anticipate the consequences of "novel" infectious agents. He is currently annotating SARS-CoV-2 proteins and variants for a U.S. CDC-funded project for which Todd Treangen is the principal investigator. He has lived in central Virginia with his family for all but 2 of the last 25 years. He and his wife, Kristen, have eight interesting children, three of whom are still at home. His professional page is https://www.linkedin.com/in/genegodbold/.

Anthony D. Kappell is a molecular biologist at Signature Science, LLC, supporting government and commercial clients with his expertise in molecular biology, microbiology, statistical analysis, and bioinformatics. His interests include (i) the structure and function of microbial communities in diverse ecosystems, including natural and engineered water systems, (ii) the anthropological impact of antibiotics and other stressors on microbial communities, and (iii) the harnessing of natural and synthetic microbial communities for human benefit. His doctoral work at the University of Texas at Arlington involved the elucidation of a two-component regulatory system for high-light-inducible proteins from cyanobacteria. He continued his work in the regulation of dark and light reactions in cyanobacteria at Oklahoma State University under the guidance of Dr. Robert Burnap. During postdoctoral studies at Marquette University, he examined the selection of antibiotic resistance in microbial communities in natural and engineered water systems.

Danielle S. LeSassier is a molecular biologist at Signature Science, LLC. She has served with multiple government-funded project teams, delivering solutions for synthetic biology, public health, and forensics challenges. She led testing of bioengineered bacteria for gas production on the DARPA-funded Bioreporters Project. She spearheaded development of artificial fingerprint samples for the IARPA Proteos Program. Her expertise supported the development of a laboratory model of viral evolution for SARS-CoV-2 on the IARPA RAVEN Project. Dr. LeSassier completed her doctoral degree at the University at Albany as part of a collaborative program between the School of Public Health and the New York State Department of Health Wadsworth Center. Her dissertation focused on the role of inteins in mycobacteria. This has led to novel insights into stress-based regulatory mechanisms of inteins and expanded the understanding of the evolution of these mobile genetic elements.

Todd J. Treangen, Ph.D., is an Assistant Professor in the Department of Computer Science at Rice University, located in Houston, TX, USA. He received his Ph.D. in computer science in 2008 from the Polytechnic University of Catalonia (Barcelona, Spain), where he focused on microbial comparative genomics. His research group focuses on large-scale computational problems specific to computational genomics and bioinformatics, with a focus on developing efficient methods and software tools targeted toward biothreat screening, infectious disease monitoring, and microbiome analysis.

Krista L. Ternus earned her Ph.D. in genetics from Texas A&M University in 2009. She completed a postdoctoral fellowship in genomics and bioinformatics at the University of Minnesota before joining Signature Science, LLC, in 2012. In collaboration with Dr. Todd Treangen, Dr. Ternus formed and led a multi-institutional team from 2017 to 2022 to develop, modify, optimize, and integrate multiple computationally predictive modules into software under the IARPA Functional Genomic and Computational Assessment of Threats (Fun GCAT) Program. The aim of the IARPA Fun GCAT Program was to detect and define pathogenic and toxic sequences by function, rather than defining sequences solely by taxonomic species of origin. This involved custom curations of sequences using pathogenesis ontologies and expansion of those labels with machine-learning techniques. The resulting software can process short or highly engineered sequences, and it has applications for both biosecurity and fundamental research questions connecting nucleotide sequences to function.