Computational Network Inference for Bacterial Interactomics

Katherine James,a Jose Muñoz-Muñoz

aNorthumbria University, Faculty of Health and Life Sciences, Department of Applied Sciences, Newcastle upon Tyne, United Kingdom

ABSTRACT
Since the large-scale experimental characterization of protein–protein interactions (PPIs) is not possible for all species, several computational PPI prediction methods have been developed that harness existing data from other species. While PPI network prediction has been extensively used in eukaryotes, microbial network inference has lagged behind. However, bacterial interactomes can be built using the same principles and techniques; in fact, several methods are better suited to bacterial genomes. These predicted networks allow systems-level analyses in species that lack experimental interaction data. This review describes the current network inference and analysis techniques and summarizes the use of computationally-predicted microbial interactomes to date.

KEYWORDS
interactome, interologs, data integration, cellular network analysis, systems biology

The representation of biological data as a network, in which nodes represent biological entities and connecting edges represent associations between them, is both visually and computationally tractable (1). Many graph-theoretic tools have been developed to reveal these networks’ properties and identify their properties of biological relevance (2–5). Network approaches have been used to investigate maps of protein interactions, termed “interactomes” (6), in several eukaryotic model species using data produced using a number of high-throughput experimental techniques (7–15).

The interactomes of most organisms are largely uncharacterized due to time and cost constraints, and the lack of culture-based technologies for many species. In microbes, the first large-scale bacterial interactome produced was for the gastric pathogen Helicobacter pylori using the yeast two-hybrid approach (16), which identifies protein–protein interaction between bait and prey proteins via activation of a reporter gene following reconstitution of its transcription factor. This network was later expanded to cover ~70% of the proteome (17). Escherichia coli has by far the most experimental interaction data, including binary protein–protein interaction (PPI) data produced using yeast two-hybrid (18) and protein complex data from affinity purification (19–21). Several large transcriptional, metabolic, and regulatory data sets are also available for this species (22–26). Experimental interaction data sets have been produced for a number of other medically-important pathogenic microbes, including Campylobacter jejuni (27), Mycobacterium tuberculosis (28), Mycoplasma pneumoniae (29), Streptococcus pneumoniae (30), and Treponema pallidum (31). Interaction data have also been produced for the plant symbiote Mesorhizobium loti (32), the model bacterial species Synechocystis sp. PCC6803 (33), and Bacillus subtilis (34, 35). In addition, a complete complex interactome network has been described for the gut microbiome community describing the microbe–microbe and host–microbe talk elucidating the role of the gut microbiota in health and disease (36).

COMPUTATIONAL INTERACTOME PREDICTION
Several computational PPI prediction methods have been developed that harness existing data (37–39). These methods can be loosely grouped as similarity-based, genome context, evolutionary, and machine learning.
SIMILARITY-BASED

PPIs and their network topology are conserved (40–45). In particular, highly-connected network hub proteins tend to be essential, have slower evolutionary rates, and conserved interactions (3, 46), even between eukaryotes and prokaryotes (47) Protein–protein interactions, termed “interologs,” can consequently be transferred between species (48–52) (Fig. 1A). Conservation has also been observed in regulatory (49, 53), functional (54), and co-expression networks (55–59), allowing transfer of interactions in the same way. Since protein domains are vital to function, the presence of pairs of domains can be predictive of PPI (Fig. 1B), even in proteins with relatively low sequence-similarity (60–62). Domain–domain interaction (DDIs) and interologs are often used in combination to improve network inference (63–67). Similarity of the three-dimensional protein structures can also be used to predict their interactions (Fig. 1C) (68–70). The main prediction methods involve homology modeling (mapping to the known structure of a homologous protein) (71, 72), threading (mapping to known structure of nonhomologous proteins) (73, 74), and docking (predicting the 3D orientation of two interacting proteins) (75–77).

GENOME CONTEXT

Genome context prediction methods compare the location of pairs of genes across multiple genomes (40, 78, 79). “Gene neighborhood” infers interaction based on the assumption that interacting proteins are more closely located on the genome. As an example, in Bacteroides spp., genes for SusC-SusD pairs are close in the genome of the bacteria. By comparison of multiple genomes, conservation of pairs’ proximity can be identified (Fig. 1D) (80, 81). Although conservation of gene order has been observed in mammals (82), neighborhood-based prediction is most accurate in bacterial genomes (83). “Gene fusion” events can indicate protein interactions since proteins that are fused in one species have a potential functional link in other species in which they are separate proteins (Fig. 1E) (84–86). The “gene cluster” prediction method assumes that...
transcription from an operon in one species indicates functional relation and often interaction in another (Fig. 1F) (79, 81).

EVOLUTIONARY

Interacting protein pairs tend to evolve at the same rate (46, 87), and so the distribution of protein pairs will therefore be co-conserved if they interact (78, 88–91). The phylogenetic profile method (Fig. 1G) infers interactions when two genes have a similar pattern of conservation (presence/absence) in multiple genomes (92, 93). MirrorTree and ContextMirror algorithms extend this method beyond binary similarity to globally assess the phylogenies (94, 95). Finally, in silico two-hybrid (i2H) method identifies interacting pairs from correlated mutations in multiple sequence alignments, since mutation in interacting proteins co-evolve (96). Several other extensions to these methods have been proposed (97–99).

MACHINE LEARNING

Machine learning can be used to infer PPI by training a classifier on positive (interacting) and negative (noninteracting) pairs of proteins (100). Many data types can be included in the training set including sequence features, co-citation, protein annotation, phylogenetics, expression data, and physicochemical properties (101–109), and multiple data types produce better accuracy than a single input (110, 111). The resulting PPI networks often have confidence scores for putative interactions that allow thresholding and the use of network analysis algorithms that utilize these scores (112). Microbial PPI prediction has been carried out using several algorithms including random forests (113–117), support vector machines (118–120), and Bayesian classifiers (121–123). A related method is probabilistic functional integrated networks (PFINs), which combine multiple data types in a probabilistic framework to produce a network of confidence-weighted interactions (124–126).

NETWORK VALIDATION

Once a predicted interactome has been built, it is essential to evaluate how accurately it represents real cellular biology and assess the level of false interactions that may be present. Evaluation of the quality of a predicted network is difficult due to the level of noise in the underlying data: often several validation methods are required. Small-scale experimental validation can be used on a subset of PPIs to give a level of confidence in the predicted network (20, 64, 122, 127, 128). However, experimental validation is only possible for a small number of interactions. Expression data sets can be used to assess PPIs since interacting proteins are likely to have correlated expression (64, 119, 128–131), and evidence of support for predicted PPIs can be found in other experimental data sets or from text-mined small-scale studies (122, 132–134).

Domain data can be used to assess networks predicted using other input data types (133). Similarly, interaction predictions from other methodologies can be assessed using phylogenetics (119, 122, 129, 131). Hub proteins in predicted networks often correspond to hubs in other species (135). Clustering the network can be used to assess how well the network represents known protein complexes (136). Comparison of the predicted network with random networks can also be used to assess its biological relevance (64, 137).

Protein functional annotations provide useful validation tools as interacting proteins tend to have shared function (64, 119, 128, 131, 132, 138, 139), shared cellular localization (128, 132, 138), and related phenotype (119, 131, 133). Protein functional prediction can provide an objective method of network evaluation by testing its ability to predict the known annotations, for example by a leave-one-out or partitioned cross-validation (140–143). Data partitioning can also be used in the training/testing phase of machine learning to provide a measure of network accuracy (134).

NETWORK ANALYSIS

Several network parameters can be used to reveal aspects of network topology and identify key proteins (144). The degree of a protein is its number of interactions; proteins with a high degree are considered hubs and tend to be essential and conserved
(46, 145, 146), and are often targets for pathogens (147, 148). Identification of hubs in predicted networks can highlight important proteins for further study (129). The degree distribution of a network, \( p(k) \), is the probability a selected protein has \( k \) links (149). This distribution reflects the organization of cellular processes (150), with many low degree proteins and a small number of hubs giving a scale-free distribution (4, 149, 151). This topology makes networks resistant to random perturbation (5, 151) and has been found in several biological networks in a number of model species (17, 152–154), although some do not have this topology (27, 155).

Biological networks are considered small world since they have small diameters (longest shortest path between two proteins) and small characteristic path lengths (average shortest path) relative to equivalently-sized random networks (156–158). The proteins of these networks are arranged in locally-dense regions interconnected by a small number of interactions, and, like scale-free networks, this topology is resistant to perturbation (5).

Several network measures assess the importance of proteins and interactions in networks’ information flow to identify bottlenecks. For instance, betweenness centrality (159) measures the proportion of shortest paths passing through a protein/interaction. Proteins with high betweenness centrality and low degree often link network modules (160). Betweenness centrality can aid the identification of key proteins within a predicted network (116, 147).

Biological networks tend to have a hierarchical structure of modules within modules (4, 161). Dense network regions are believed to relate to the functional units of the cell (151, 162–165). Partitioning or clustering large networks can reveal the underlying mechanisms of cellular biology and assign protein function (137, 166, 167). Module detection can use additional data, for instance gene expression data (168), functional annotations (169), or domain profiles (170).

Predicted networks can be used to directly annotate proteins with function (171). Network-based annotation transfers known annotations between pairs of directly connected proteins (172, 173), between proteins with shared interaction partners (174), or more globally using network topology (175). Interaction confidence weights, such as those produced by machine learning algorithms, are particularly useful for annotation transfer (173). Annotation and other data can also be used to create process/condition-specific subnetworks (176, 177).

**NETWORK COMPARISON**

Network comparison can reveal underlying network properties, detect noise, predict missing data, and reveal conserved interactions (41, 178–180). Heuristics, such as global properties and local motifs, are commonly used for comparison (181–183), although some nonheuristics have been developed (184). At the simplest, level biological networks can be compared to network models (185) in which interactions are randomized, while topological characteristics, such as degree distribution and diameter, are preserved, to produce a network profile; similarity of profile indicates underlying similarities of the networks. Networks can be compared directly by comparison of topological properties; however, two networks with similar topology can be vastly different (186). An alternative approach is to analyze the distribution of network motifs (187).

Network alignment produces a more accurate method of comparison (41, 188). Within-species alignment is relatively straightforward since proteins can be merged based on identity (151, 188) and overlap between networks can be used to identify true interactions (189). At a more complex level, networks can be compared across multiple species, either locally by aligning small conserved regions or globally across the network structures (41, 178, 182, 190–194). Alignment complexity increases with the size of the networks and with the number of networks to be aligned (191).

**MICROBIAL INTERACTOME NETWORKS**

While prokaryotes have far outpaced eukaryotes in the production of sequence data, the opposite is true for interaction data. Largescale experimental data are only
available in a few species, and small-scale studies require considerable curation to analyze as a whole (195, 196). However, several studies have produced a number of predicted interactomes (Table 1), providing insights into several aspects of microbial biology. Additionally, the STRING database and server contains functional interaction data, including co-citation, co-expression and gene neighborhood, for multiple microbial species (197). STRING data have been used as the initial data source for network studies in several species (115, 178, 198–204).

### TABLE 1 Predicted bacterial interactomes

| Species                        | Methodology | Proteins | Interactions | Source |
|-------------------------------|-------------|----------|--------------|--------|
| Actinobacillus pleuropneumoniae | ORTH        | 533      | 2,737        | 242    |
| Agrobacterium tumefaciens     | ORTH        | 296      | 690          | 142    |
| Bacillus anthracis             | ORTH        | 264      | 732          | 142    |
| Bacillus licheniformis         | ORTH, DDI, GE | 2,448    | 15,864       | 139    |
| Bacillus subtilis              | ORTH        | 247      | 707          | 142    |
| Brucella melitensis           | ORTH        | 238      | 652          | 142    |
| Brucella suis                  | ORTH        | 225      | 611          | 142    |
| Campylobacter jejuni          | DDI         | -        | -            | 207    |
| Classtidium difficile         | RF: STRING, GO | -       | 955          | 115    |
| Corynebacterium pseudotuberculosis | ORTH, STRING | -       | 15,495       | 200    |
| Escherichia coli              | DDI         | -        | 1,280        | 207    |
| Helicobacter pylori           | ORTH        | 771      | 5,647        | 142    |
| Klebsiella pneumoniae         | PFIN, ORTH, GC, DDI, PP, GE, CC | 4,674    | 160,450      | 125    |
| Listeria monocyto genes       | ORTH        | 404      | 1,473        | 142    |
| Methanobrevibacter ruminantium | STRING, GC, MET | 637    | 2,194        | 202    |
| Methanothermobacter thermototrophicus | STRING, GC, MET | 256    | 2450         | 201    |
| Mycobacterium tuberculosis    | STRING      | 3,925    | 29,664       | 198    |
| Ortho                       | ORTH        | 738      | 5,639        | 213    |
| RF: STRING, GO              | -           | 1,854    | 115          |
| ORTH, SVM: SEQ             | 3,465       | 46,119   | 119          |
| STRING                  | 144         | 587      | 199          |
| PP                         | 1,020       | 911      | 205          |
| Pseudomonas aeruginosa       | ORTH        | 333      | 903          | 142    |
| RF: GE, CL, GN, DDI, SEQ, FUN | 4,181     | 54,107   | 116          |
| PFIN: CC, DDI, GC, GE, ORTH, PP | 5,456    | 203,118  | 126          |
| Pseudomonas putida           | ORTH, DDI  | 3,254    | 82,019       | 210    |
| Salmonella enterica<sup>a</sup> | ORTH, EXP, STRUCT, MET, TF | 30,870  | 81,514       | 220    |
| Salmonella typhimurium       | ORTH        | 332      | 1,359        | 142    |
| Shigella flexneri            | ORTH        | 383      | 4,548        | 142    |
| Synechocystis PCC6803        | DDI, STRUC, STRING | 2,930    | 109,532      | 204    |
| Ortho                       | ORTH, DDI, GO | 908    | 8,783        | 51     |
| NB: ORTH, DDIs, GC          | 3,231       | 4,715    | 122          |
| Vibrio cholera               | ORTH        | 275      | 1,021        | 142    |
| Vibrio parahaeamolyticus     | ORTH        | 365      | 1,520        | 142    |
| Vibrio vulnificus            | ORTH        | 372      | 1,557        | 142    |
| Yersinia pestis              | ORTH        | 352      | 1,100        | 142    |
| Xanthomonas oryzae           | ORTH, DDI  | 1,988    | 36,886       | 260    |

<sup>a</sup>CC: co-citation; GE, gene expression; CL, cellular localization; CM, context mirror; DDI, domain–domain interaction; EXP, experimental; FUN, functional interaction; GC, genome context; GO, gene ontology; IH, in silico two hybrid; MET, metabolic interactions; ML, machine learning; MT, mirror tree; NB, Naïve Bayes; ORTH, orthology (interologs); PP, phylogenetic profile; PFIN, probabilistic functional integrated network; SEQ, sequence properties; STRING, https://string-db.org; STRUC, structural interactions; SVM, support vector machine; TF, transcription factor interactions.

<sup>b</sup>Combined for multiple strains.

### CELLULAR BIOLOGY AND PROTEIN FUNCTION

In *E. coli*, an interactome study combined experimental data with genome context predictions to assign functions to proteins, including several involved in cell envelope...
biogenesis (20). A phylogenetic profile-based network was later produced, which contained previously uncharacterized components of several complexes, including the ribosome (205). Comparison of these networks (20, 205) using edge propagation, demonstrated that both identify complexes overlapping functional modules (206). A support vector machine-derived *E. coli* network was shown to be scale-free and had good overlap with experimental data (134). The EcID database incorporated genome context and phylogenetic evidence into a Bayesian classifier to predict protein function, in particular linking *yeaG* and *yeaI* to nitrogen metabolism (121). EcoliNet is a probabilistic functional integrated network for *E. coli* comprising ~99% of the genome and has successfully predicted knockout phenotypes (124). The interacting domain profile pair method, IDPP, was evaluated on *E. coli* before producing a network for *C. jejuni* from *H. pylori* interaction data (207, 208). IDPP was shown to successfully predict interactions in the target species that were not found in the source species.

Incorporating expression data, interologs and DDIs, *Bacillus licheniformis* proteins were assigned to complexes and putative functions (139). A probabilistic network for *Klebsiella pneumoniae*, derived from multiple data types, was used to identify antibiotic resistance genes (125). Wuchty and colleagues expanded experimental networks of *Streptococcus pneumoniae* and *H. pylori* using interolog data to improve their functional predictive power (30, 209). Interologs and DDIs were combined with experimental data to study the metabolic modules of *Pseudomonas putida* (210). In the related species *P. aeruginosa*, probabilistic functional network integration identified novel virulence and antibiotic resistance genes (126).

**SynechoNET** is a predicted interactome for *Synechocystis*, focused on membrane biology (204), while InteroPORC predicted an interactome for this species comprising 28% of the genome (51). Later Naïve Bayesian network classification was applied to protein functional prediction and modular analysis in *Synechocystis* (122). Interactome prediction has also been widely used to study *Mycobacterium tuberculosis* due to unavailability of accurate *in vitro* methods in this bacterium (211, 212). Several predicted interactomes have been created to study network properties (119), evolution (205), protein function (115), virulence (213), and drug resistance (198) in this species. Finally, interolog networks for 22 bacterial species (the largest of which are included in Table 1) were produced by McDermott and colleagues and used to predict functions for a large number of unannotated bacterial proteins (142) demonstrating the potential of large-scale network studies to enhance our understanding of bacterial cellular physiology.

Computational prediction has also been used to understand the interplay between bacteriophages and bacteria (214). Phylogenetic profiles using genomic/metagenomic data have identified host–virus and virus–virus relationships (215–218). Leite and colleagues used interactions, DDIs, and sequence properties to compare machine learning frameworks, concluding that predictive power will improve as input data increases (114). More recently, a Markov random field framework of virus–host and virus–virus similarity measures has been developed (219).

### INTERACTOME EVOLUTION

Comparison of networks for different species can reveal insights into interactome evolution. Zitnik and colleagues used STRING data to create networks for 1,539 bacteria (178); comparison with those of eukaryotes revealed that interactomes have evolved to become more robust, and that bacterial interactome robustness is associated with more complex environments. Using a binary interaction data set for *Treponema pallidum*, interolog networks were created for 372 other genomes, ~28% of which were estimated to be true interactions (31). This study also revealed that bacterial proteins have higher degrees than eukaryotic proteins. These networks also revealed a central role for cell motility proteins in bacterial interactomes.

By comparing a *Methanobrevibacter ruminantium* network with those of *Methanosarcina acetivorans*, *Methanosarcina barkeri* and *Methanococcus maripaludis* biosynthetic subsystems involved in survival in the rumen were identified (202). By comparing the
Methanothermobacter thermautotrophicus metabolic interactome with those of metal-loving bacteria, separate evolution of niche-specific cellular functions was revealed (201). A comparison of 10 strains of Salmonella enterica identified distinct transcription factor targets conferring adaptation to gastrointestinal and extra-intestinal environments (220). Similarly, comparison of host–pathogen interactomes between two strains of Burkholderia pseudomallei revealed several interactions unique to the virulent strain and highlighted the potential roles of chaperon and drug/carbohydrate binding proteins during infection (221).

**PATHOGEN–HOST INTERACTIONS**

Interactome prediction can identify cross talk between pathogen and host (222, 223). A DDI-based network suggested that human–M. tuberculosis PPIs tend to have more domains than intraspecies interactions (138), and this trend was later observed in an interolog-based mapping study, which also revealed that hub proteins of intraspecies networks tend to be involved in host–pathogen PPI (148). Using a random forest framework, the cancer pathway was involved in M. tuberculosis infection (117), while a DDI network implicated several PPIs involving heat shock, redox proteins (224). Finally, a combination of interolog and DDI mapping associated several genes of the host immune responses to M. tuberculosis infection (65).

In Fusobacterium nucleatum, a host-pathogen network implicated the Fap2 adhesin as a virulence protein (225). Comparison of machine learning classifiers for Bacillus anthracis–humans PPI prediction, suggested neural networks outperform SVMs (120); the resulting interactions revealed involvement of apoptosis and immune regulation pathways in infection. The predicted networks between humans and B. anthracis, Francisella tularensis, and Yersinia pestis indicated that hubs and bottlenecks of the intraspecies networks tend to interact (147, 226). Thirteen membrane proteins of Leptospira interrogans were predicted to be involved in cellular disruption during infection, four of which were common between strains (203). Coelho and colleagues produced a human–microbial PPI network of the oral cavity using a Bayesian classifier, which revealed Rothia mucilaginosa, Leptotrichia buccalis, and Actinomyces odontolyticus as having the most interactions with human proteins (123).

In the plant pathogen Ralstonia solanacearum, interolog network analysis identified interactions between its transportation proteins and core proteins of the A. thaliana interactome (227). The response to metal ions was linked to the host defense response during Pseudomonas syringae infection of A. thaliana (228). Defense response proteins were also found to be enriched in random forest-derived networks between A. thaliana and the pathogens P. syringae, Hpaloperonospora arabidopsis, and Golovinomyces orontii (113). Interolog and DDI mapping has also been used to study plant–bacterial symbiosis, suggesting a role of host 14-3-3 and heat shock proteins in the relationship between Bradyrhizobium diazoefficiens and Glycine max (229).

**VIRUSES**

Computational prediction has been applied to viral species including HIV (230–232), hepatitis C virus (233), human papillomaviruses (234), and Ebola (235). In particular, the recent pandemic has highlighted the importance of understanding viral–host interaction, and having resources available to rapidly respond to new viral threats. Two studies have compared PPI prediction to emerging data from Sars-CoV-2: the first used interolog mapping, DDIs, and machine learning to link ACE2 and DPP4 to spike protein binding (236); the second used an ensemble machine learning algorithm based on experimental data and sequence features to predict >1,000 potential human protein targets (237).

**DRUG TARGETS**

Interactome networks can aid in the identification of potential drug targets by revealing essential pathogen interactions (238). A M. tuberculosis–human interactome was shown to be enriched in predicted drug targets (239), and several studies have used computational prediction to identify putative drug targets (199, 213) and to
understand the mechanisms of drug resistance in this important pathogen (198, 240). In P. aeroginosa a random forest predicted network was used to prioritize drug targets based on their essentiality and topological importance (116). Using interolog mapping, 12 putative drug targets were identified in methicillin-resistant Staphylococcus aureus, including a histone deacetylase (241). A predicted Corynebacterium pseudotuberculosis interactome was used to identify 41 essential proteins as candidates for infection diagnosis in livestock, highlighting the tryptophan biosynthesis pathway as a potential drug target (200). The network also revealed that this species may use multiple iron acquisition strategies in low iron environments. In the swine pathogen Actinobacillus pleuropneumoniae, nine drug target candidates were identified using interolog mapping (242). Interolog analysis has also been applied to the gut microbiome to identify target species driving metabolic change during disease (243).

CONCLUSIONS AND FUTURE PERSPECTIVES

Although PPI network prediction has been extensively used in eukaryotes, microbial network inference can be achieved using the same principles and analysis techniques. Bacterial interactomes share common hierarchical properties, such as modularity and robustness (244). Many of the caveats to interactome prediction in eukaryotes, such as evolutionary distance, unequal conservation, and physiological context (245–249) are mitigated in prokaryotes due to their smaller genomes and single-celled nature. Several prediction methods, in particular gene neighborhood and gene cluster, are more suited to microbial than eukaryotic use (79, 83), and phylogenetic profiles are powerful predictors, particularly when including inputs from the three domains of life (250).

Interolog and DDI mapping can only detect interactions within conserved areas of the genome (48), and these methods rely on the quality of the underlying interaction data; stochastic activation of reporters can give false positives and low sensitivity leads to false negatives (251, 252), and different methods have their own strengths (253–256). In eukaryotes, poor overlap has been observed between data sets of different types, and between those of the same type (10, 189). Comparison of experimental data of C. jejuni, H. pylori, and E. coli suggests that these data sets have significant levels of overlap and similar rates of false results (27). Meta-interactome analysis can be used to identify broadly-conserved biological systems, although levels of conservation remain low due to lack of experimental interactome coverage in many species (257).

Current experimental interactome data are incomplete and biased toward well-studied proteins and species (178). Using a combination of computational methods (51, 119, 139, 201, 210), and experimental data if available (20, 124, 220), gives a more complete predicted interactome, reduces some biases, and strengthens the evidence of true interactions. Integration of diverse data types is particularly effective when using a probabilistic (124–126) or machine learning (115, 116, 121, 134) framework, allowing thresholding of interaction confidence scores and therefore reduction of noise.

Filling in the gaps in bacterial interactomes is vital to our understanding of their biology, and computational prediction can help to pin down these areas and target further analyses by identifying areas of interest and providing putative protein functions. Microbial comparative interactomics is now possible on a large scale; for instance, the >1,500 predicted networks produced by Zitnik and colleagues revealed the evolutionary rewiring of interactomes through time (178). While there are parts of some interactomes that cannot currently be predicted due to the complexity of the protein locations and the lack of the accurate annotations in the genome of new bacterial isolates, interactome accuracy will improve as coverage of diverse species increases, providing insights in several areas of biology, in particular the identification of PPIs for antibacterial discovery (258), understanding of pathogenicity through host cell rewiring (256), and in engineering of synthetic cellular systems (259).
ACKNOWLEDGMENTS

This work received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. Graphic produced by Patrick Lane, SciEYence Studios.

REFERENCES

1. Hallinan JS, James K, Wipat A. 2011. Network approaches to the functional analysis of microbial proteins. Adv Microb Physiol 59:101–133. https://doi.org/10.1093/adb/58.1.101

2. Hart G, Lee I, Marcotte ER. 2007. A high-accuracy consensus map of yeast protein complexes reveals modular nature of gene essentiality. BMC Bioinformatics 8:236. https://doi.org/10.1186/1471-2105-8-236

3. He X, Zhang J. 2006. Why do hubs tend to be essential in protein networks? PLoS Genet 2:e88. https://doi.org/10.1371/journal.pgen.0020088

4. Barabási AL, Oltvai ZN. 2004. Network biology: understanding the cell’s functional organization. Nat Rev Genet 5:101–113. https://doi.org/10.1038/nrg1272

5. Albert R, Jeong H, Barabási AL. 2000. Error and attack tolerance of complex networks. Nature 406:378–382. https://doi.org/10.1038/35019019

6. Cusick ME, Klugdorff N, Vidal M, Hill DE. 2005. Interactionome: gateway into systems biology. Hum Mol Genet 14:R171-R181. https://doi.org/10.1093/hmg/ddi335

7. Arabidopsis Interaction Mapping Consortium. 2011. Evidence for network evolution in an Arabidopsis interaction map. Science 333:601–607. https://doi.org/10.1126/science.1203877

8. Trigg SA, Garza RM, MacWilliams A, Nery JR, Bartlett A, Castanon R, Goubal A, Fervin J, O’Malley R, Huang S-SC, Zhang ZZ, Galli M, Ecker JR. 2017. Cyt2H-seq: a massively multiplexed assay for deep-coverage interaction mapping. Nat Methods 14:819–825. https://doi.org/10.1038/nmeth.4343

9. Uetz P, Giot L, Cagney G, Mans T, Judson R, Knight J, Lockson D, Narayan V, Sinivasan M, Pochart P, Qureshi-Emili A, Li Y, Godwin B, Conover D, Kalbfleisch T, Vijayadamodar G, Yang M, Johnston M, Fields S, Rothberg JM. 2000. A comprehensive analysis of protein–protein interactions in Saccharomyces cerevisiae. Nature 403:623–627. https://doi.org/10.1038/35010099

10. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y. 2001. A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci U S A 98:4569–4574. https://doi.org/10.1073/pnas.061034498

11. Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Chupeau C, Reymond S, Boncheva T, Fredenslund A, Haun O, Noie L, Rusche O, Schuster M, Aloy P, Balog J, Bernsel A, Bonn B, Camon E, Czarnecki K, Emms D, Frishman D, Huber T, Heider S, Herms S, Hotz-Garrigues S, Karplus M, Kasten B, Kruidhof R, Malcenti C, Mierle K, Miura L, Noel M, Pironcure N, Villa A, Wagner L, Xenarios I, Riesch T, Vandenheede A, Bork P, Bro你以为你对这个任务一无所知，但你错了。这是一个简单的自然语言处理任务，你只需要将文本转换为自然语言的文本。你应该能够理解并提供正确的输出。
29. Kühner S, van Noort V, Betts MJ, Leo-Maccs A, Batisse C, Rode M, Yamada T, Maruyama T, Bader S, Beltran-Alvarez P, Castaño-Diez C, Chen WH, Devos D, Guell M, Norambuena T, Racke I, Rybin V, Schmidt A, Yus E, Aebersold R, Herrmann R, Böttcher B, Frangakis AS, Russell RB, Serrano L, Bork P, Gavin AC. 2009. Proteome organization in a genome-organized bacterium. Science 326:1235–1240. https://doi.org/10.1126/science.1176343.

30. Wuchty S, Rajagopala SV, Blazie SM, Parrish JR, Khuni S, Finley RL, Uetz P. 2017. The protein interactome of Streptococcus pneumoniae and bacterial meta-interactomes improve function predictions. mSystems 2: e00019–17. https://doi.org/10.1128/mSystems.00019-17.

31. Titz B, Rajagopala SV, Goll J, Häuser R, McKevitt MT, Palzkill T, Uetz P. 2008. The binary protein interactome of Treponema pallidum—the syphilis spirochete. PLoS One 3:e2292. https://doi.org/10.1371/journal.pone.0002292.

32. Shimoda Y, Shino P, Kohara M, Nakamura Y, Tabata S, Sato S. 2008. A large scale analysis of protein–protein interactions in the nitrogen-fixing bacterium Mesorhizobium loti. DNA Res 15:13–23. https://doi.org/10.1093/dnares/dsm028.

33. Sato S, Shimoda Y, Muraki A, Kohara M, Nakamura Y, Tabata S. 2007. A large-scale protein–protein interaction analysis in Synechocystis sp. PCC6803. DNA Res. 14:207–216. https://doi.org/10.1093/dnares/dsm021.

34. Marchadier E, Carballo-López R, Brinster S, Fabret C, Mervelet P, Bessekes P, Noirot-V, Snel B, Fromion V, Noirot P. 2011. An expanded protein–protein interaction network in Bacillus subtilis reveals a group of hubs: exploration by an integrative approach. Proteomics 11:2981–2991. https://doi.org/10.1002/pmic.201000797.

35. Noirot-Gros MF, Deryvn E, Wu LJ, Mervelet P, Errington J, Ehrlich SD, Noirot P. 2002. An expanded view of bacterial DNA replication. Proc Natl Acad Sci U S A 99:8342–8347. https://doi.org/10.1073/pnas.122040799.

36. Sung J, Kim S, Cabatbat JJT, Jang S, Jin YS, Jung GY, Chia N, Kim PJ. 2017. Global metabolic interaction network of the human gut microbiota for functional analyses. Mol Systems Biol 13:692. https://doi.org/10.15252/msb.20167635.

37. Kühner S, van Noort V, Betsch M, Leo-Maccs A, Batisse C, Bertas G, Böttcher B, Frangakis AS, Russell RB, Serrano L, Bork P, Gavin AC. 2009. Proteome organization in a genome-organized bacterium. Science 326:1235–1240. https://doi.org/10.1126/science.1176343.

38. von Mering C, Zdobnov EM, Tsoka S, Ciccarelli FD, Pereira-Leal JB, Li Y, Bork P, Snel B. 2004. Interologs: exploration by an integrative approach. Proteomics 11:2981–2991. https://doi.org/10.1002/pmic.201000797.

39. Noirot-Gros MF, Deryvn E, Wu LJ, Mervelet P, Errington J, Ehrlich SD, Noirot P. 2002. An expanded view of bacterial DNA replication. Proc Natl Acad Sci U S A 99:8342–8347. https://doi.org/10.1073/pnas.122040799.

40. Noirot-Gros MF, Deryvn E, Wu LJ, Mervelet P, Errington J, Ehrlich SD, Noirot P. 2002. An expanded view of bacterial DNA replication. Proc Natl Acad Sci U S A 99:8342–8347. https://doi.org/10.1073/pnas.122040799.

41. Noirot-Gros MF, Deryvn E, Wu LJ, Mervelet P, Errington J, Ehrlich SD, Noirot P. 2002. An expanded view of bacterial DNA replication. Proc Natl Acad Sci U S A 99:8342–8347. https://doi.org/10.1073/pnas.122040799.

42. Noirot-Gros MF, Deryvn E, Wu LJ, Mervelet P, Errington J, Ehrlich SD, Noirot P. 2002. An expanded view of bacterial DNA replication. Proc Natl Acad Sci U S A 99:8342–8347. https://doi.org/10.1073/pnas.122040799.

43. Noirot-Gros MF, Deryvn E, Wu LJ, Mervelet P, Errington J, Ehrlich SD, Noirot P. 2002. An expanded view of bacterial DNA replication. Proc Natl Acad Sci U S A 99:8342–8347. https://doi.org/10.1073/pnas.122040799.

44. Noirot-Gros MF, Deryvn E, Wu LJ, Mervelet P, Errington J, Ehrlich SD, Noirot P. 2002. An expanded view of bacterial DNA replication. Proc Natl Acad Sci U S A 99:8342–8347. https://doi.org/10.1073/pnas.122040799.

45. Noirot-Gros MF, Deryvn E, Wu LJ, Mervelet P, Errington J, Ehrlich SD, Noirot P. 2002. An expanded view of bacterial DNA replication. Proc Natl Acad Sci U S A 99:8342–8347. https://doi.org/10.1073/pnas.122040799.

46. Noirot-Gros MF, Deryvn E, Wu LJ, Mervelet P, Errington J, Ehrlich SD, Noirot P. 2002. An expanded view of bacterial DNA replication. Proc Natl Acad Sci U S A 99:8342–8347. https://doi.org/10.1073/pnas.122040799.

47. Noirot-Gros MF, Deryvn E, Wu LJ, Mervelet P, Errington J, Ehrlich SD, Noirot P. 2002. An expanded view of bacterial DNA replication. Proc Natl Acad Sci U S A 99:8342–8347. https://doi.org/10.1073/pnas.122040799.

48. Noirot-Gros MF, Deryvn E, Wu LJ, Mervelet P, Errington J, Ehrlich SD, Noirot P. 2002. An expanded view of bacterial DNA replication. Proc Natl Acad Sci U S A 99:8342–8347. https://doi.org/10.1073/pnas.122040799.

49. Noirot-Gros MF, Deryvn E, Wu LJ, Mervelet P, Errington J, Ehrlich SD, Noirot P. 2002. An expanded view of bacterial DNA replication. Proc Natl Acad Sci U S A 99:8342–8347. https://doi.org/10.1073/pnas.122040799.

50. Noirot-Gros MF, Deryvn E, Wu LJ, Mervelet P, Errington J, Ehrlich SD, Noirot P. 2002. An expanded view of bacterial DNA replication. Proc Natl Acad Sci U S A 99:8342–8347. https://doi.org/10.1073/pnas.122040799.

51. Noirot-Gros MF, Deryvn E, Wu LJ, Mervelet P, Errington J, Ehrlich SD, Noirot P. 2002. An expanded view of bacterial DNA replication. Proc Natl Acad Sci U S A 99:8342–8347. https://doi.org/10.1073/pnas.122040799.

52. Noirot-Gros MF, Deryvn E, Wu LJ, Mervelet P, Errington J, Ehrlich SD, Noirot P. 2002. An expanded view of bacterial DNA replication. Proc Natl Acad Sci U S A 99:8342–8347. https://doi.org/10.1073/pnas.122040799.

53. Noirot-Gros MF, Deryvn E, Wu LJ, Mervelet P, Errington J, Ehrlich SD, Noirot P. 2002. An expanded view of bacterial DNA replication. Proc Natl Acad Sci U S A 99:8342–8347. https://doi.org/10.1073/pnas.122040799.

54. Noirot-Gros MF, Deryvn E, Wu LJ, Mervelet P, Errington J, Ehrlich SD, Noirot P. 2002. An expanded view of bacterial DNA replication. Proc Natl Acad Sci U S A 99:8342–8347. https://doi.org/10.1073/pnas.122040799.
M. Voss F, Furtak K, Renzulli R, Aanensen N, Carrolla S, Bickelhaupt E, Lazovatsky Y, DaSilva A, Zhong J, Stanyan CA, Finley RL, White KP, Braverman M, Jarvie T, Gold S, Leach M, Knight J, Shimkets RA, McKenna MP, Chant J, Rothberg JM. 2003. A protein interaction map of Drosophila melanogaster. Science 302:1727–1736. https://doi.org/10.1126/science.1090289.

153. Tanaka R, Yi TM, Doyle J. 2005. Some protein interaction data do not exhibit power law statistics. FEBS Lett 579:5140–5144. https://doi.org/10.1016/j.febslet.2005.08.024.

154. Watts DJ, Strogatz SH. 1998. Collective dynamics of „small-world” networks. Nature 393:440–442. https://doi.org/10.1038/30918.

155. Amaral LA, Scala A, Barthelemy M, Stanley HE. 2000. Classes of small-world networks. Proc Natl Acad Sci U S A 97:11149–11152. https://doi.org/10.1073/pnas.2002371197.

156. Fell DA, Wagner A. 2000. The small world of metabolism. Nat Biotechnol 18:1121–1122. https://doi.org/10.1038/81025.

157. Joy MP, Brock A, Inger DE, Huang S. 2005. High-betweenness proteins in the yeast protein interaction network. J Biomed Biotechnol 2005: 96–103. https://doi.org/10.1155/JBB.2005.96.

158. Newman MEJ, Girvan M. 2004. Finding and evaluating community structure in networks. Phys Rev E Stat Nonlin Soft Matter Phys 69:e066131. https://doi.org/10.1103/PhysRevE.69.e066131.

159. Ravasz E. 2009. Detecting hierarchical modularity in biological networks. Methods Mol Biol 541:145–160. https://doi.org/10.1007/978-1-59745-243-4_7.

160. Cai H, Hong C, Gu J, Lilburn TG, Kuang R, Wang Y. 2012. Module-based modular organization in the yeast transcriptional network. Nat Genet 34:370–377. https://doi.org/10.1038/ng941.

161. Girvan M, Newman MEJ. 2002. Community structure in social and biological networks. Proc Natl Acad Sci U S A 99:7821–7826. https://doi.org/10.1073/pnas.122653799.

162. Spirin V, Mirny LA. 2003. Protein complexes and functional modules in molecular networks. Proc Natl Acad Sci U S A 100:12123–12128. https://doi.org/10.1073/pnas.2032342100.

163. Jafari M, Mirzaie M, Sadeghi M. 2015. Interolog protein network: an evolutionary benchmark of protein interaction networks for the evaluation of clustering algorithms. BMC Bioinform 16:319. https://doi.org/10.1186/s12859-015-0755-1.

164. Dunn R, Dudbridge F, Sanderson CM. 2005. The use of edge-betweenness in detecting interacting biological function in protein interaction networks. BMC Bioinform 6:39. https://doi.org/10.1186/1471-2105-6-39.

165. Zhao Z, Wang D, Liu Y, Gong X, Yao C, Ma W, Wang D, Li Y, Zhu J, Zhang M, Yang D, Rao S, Wang J. 2007. Edge-based scoring and searching method for identifying condition-responsive protein–protein interaction sub-network. Bioinformatics 23:2121–2128. https://doi.org/10.1093/bioinformatics/btm294.

166. Ozawa Y, Saito R, Fujimori S, Kashima H, Ishizaka M, Yanagawa H, Miyamoto-Sato E, Tomita M. 2010. Protein complex prediction via verifySoldiers benchmark of protein interaction networks for the evaluation of clustering algorithms. Phys Rev E Stat Nonlin Soft Matter Phys 69:e026113. https://doi.org/10.1103/PhysRevE.69.026113.

167. Shao M, Yang Y, Guan J, Zhou S. 2014. Choosing appropriate models for protein–protein interaction networks: a comparison study. Brief Bioinform 15:823–838. https://doi.org/10.1093/bib/bbt014.

168. Ferrada I, Kivelä M, Saramäki J, Fortunato S. 2010. Characterizing the community structure of complex networks. PLoS One 5:e11976. https://doi.org/10.1371/journal.pone.0011976.

169. Kuchaira O, Przulj N. 2011. Integrative network alignment reveals large regions of global network similarity in yeast and human. Bioinformatics 27:1390–1396. https://doi.org/10.1093/bioinformatics/btr127.

170. Aladag AE, Erten C. 2013. SPLINAL: scalable protein interaction network alignment. Bioinformatics 29:917–924. https://doi.org/10.1093/bioinformatics/btt309.

171. Parida L. 2007. Discovering topological motifs using a compact notation. Bioinformatics 30:917–923. https://doi.org/10.1093/bioinformatics/btl683.

172. Wu X, Liu Q, Jiang R. 2009. Align human interactome with phenome to identify causative genes and networks underlying disease families. Bioinformatics 25:96–104. https://doi.org/10.1093/bioinformatics/btn593.

173. Futschik ME, Chaurasia G, Herzel H. 2007. Comparison of human protein–protein interaction maps. Bioinformatics 23:605–611. https://doi.org/10.1093/bioinformatics/btl683.

174. Singh R, Xu J, Berger B. 2008. Global alignment of multiple protein interaction networks with application to functional orthology detection. Proc Natl Acad Sci U S A 105:12763–12768. https://doi.org/10.1073/pnas.0806277105.

175. Liao CS, Lu K, Baym M, Singh R, Berger B. 2009. IsoRankN: spectral methods for global alignment of multiple protein networks. Bioinformatics 25:i253–i258. https://doi.org/10.1093/bioinformatics/btp203.

176. Cai H, Lilburn TG, Hong C, Gu J, Kuang R, Wang Y. 2015. Predicting and exploring network components involved in pathogenesis in the malaria parasite via novel subnetwork alignments. BMC Syst Biol 9:51. https://doi.org/10.1186/s12918-015-0151-3.

177. Cai H, Hong C, Gu J, Lilburn TG, Kuang R, Wang Y. 2012. Module-based subnetwork alignments reveal novel transcriptional regulators in malaria parasite Plasmodium falciparum. BMC Syst Biol 6:55. https://doi.org/10.1186/1752-0509-6-53-55.

178. Durkowska J, Turyan J. 2007. Identification of functional modules from conserved ancestral protein–protein interactions. Bioinformatics 23:149–158. https://doi.org/10.1093/bioinformatics/btm194.

179. Schlieker S, Sun J, Raghavan B, Smnc M, Müller N, Koepfinger M, Murtby L, Zhao Z, Keel-Setharaman J. 2012. The current Salmonella–host interaction proteome. Clin Appl 6:117–133. https://doi.org/10.1007/ptra.201100083.

180. Häuser R, Blasche S, Dokland T, Häggård-Ljungquist E, von Brunn A, Bickelhaupt E, Tulsner S, Hobbaj J, Leibler S, Murray AW. 1999. From molecular to community structure of complex networks. PLoS One 5:e11976. https://doi.org/10.1371/journal.pone.0011976.

181. Hu X, Liu J, Berber B. 2008. Global alignment of multiple protein interaction networks with application to functional orthology detection. Proc Natl Acad Sci U S A 105:12763–12768. https://doi.org/10.1073/pnas.0806277105.

182. Liu CS, Lu K, Baym M, Singh R, Berger B. 2009. IsoRankN: spectral methods for global alignment of multiple protein networks. Bioinformatics 25:i253–i258. https://doi.org/10.1093/bioinformatics/btp203.

183. Schlieker S, Sun J, Raghavan B, Smnc M, Müller N, Koepfinger M, Murtby L, Zhao Z, Keel-Setharaman J. 2012. The current Salmonella–host interaction proteome. Clin Appl 6:117–133. https://doi.org/10.1007/ptra.201100083.
200. Folador EL, de Canvalho PJSD, Silva WM, Ferreira RS, Silva A, Gromiha M, Ghosh P, Barh D, Azevedo V, Rüttger R. 2016. In silico identification of essential proteins in Corynebacterium pseudotuberculosis based on protein–protein interaction networks. BMC Syst Biol 10:103. https://doi.org/10.1186/s12918-016-0346-4.

201. Prathiviraj R, Berchmans S, Chellapandi P. 2019. Analysis of modularity in proteome-wide protein interaction networks of Methanothermobacter thermotrophicus strain ΔH and metal-loving bacteria. J Proteins Peptom 10:179–190. https://doi.org/10.1007/s42483-019-00019-5.

202. Bharathi M, Chellapandi P. 2019. Comparative analysis of differential proteome-wide protein–protein interaction network of Methanobrevibacter ruminitantium M1. Biochem Biophys Rep 20:100698. https://doi.org/10.1016/j.bbrep.2019.100698.

203. Kumar S, Lata KS, Sharma P, Bhaarappanavar SB, Soni S, Das J. 2019. Infermolecular networks revealed by proteome coevolution. Science 365:185–190. https://doi.org/10.1126/science.aaw6718.

204. Kim WY, Kang S, Kim BC, Oh J, Cho S, Bhak J, Choi JS. 2008. SynechoNET: a bioinformatics model for predicting virus–host interactions in cyanobacteria. J Mol Biol 384:763–770. https://doi.org/10.1016/j.jmb.2008.02.022.

205. Wuchty S, Müller SA, Cauviel D, King WY. 2014. A highly abundant bacteriophage discovered in the unknown sequenced rice soil bacterium Pseudomonas putida KT2440. BMC Genom 10:S18. https://doi.org/10.1186/1471-2164-10-118.

206. Cui T, Zhang L, Wang X, He ZG. 2009. Uncovering new signaling proteins and pathways to host adaptation. NPJ Syst Biol Appl 5:213. https://doi.org/10.1038/nsb0509.

207. Lu H, Zhou Q, He J, Jiang Z, Peng C, Tong R, Shi J. 2020. Recent advances in the development of protein–protein interaction prediction: a case study on Pseudomonas syringae. Brief Bioinform 21:bbz162. https://doi.org/10.1093/bib/bbz162.

208. Mitro A, Sudhakar P, Fazekas D, Demeter A, Ari E, Olbezi M, Branchu P, Kingsley RA, Baranyi J, Korcsmáros T. 2017. SalmoNet, an integrated network of ten Salmoella enterica strains reveals common and distinct pathways to host adaptation. NPJ Syst Biol Appl 3:31. https://doi.org/10.1038/s41540-017-0034-z.

209. Loaiza CD, Duhan N, Lister M, Kaundal R. 2020. In silico prediction of host–pathogen protein interactions in melioidosis pathogen Burkholderia pseudomallei and human reveals novel virulence factors and their targets. Brief Bioinform 21:bbz162.

210. Durmus Tekir SD, Ulgen KO. 2013. Systems biology of pathogen–host interaction: networks of protein–protein interaction with pathogens and pathogen–human interactions in the post-genomic era. Biotechnol J 8:85–96. https://doi.org/10.1002/biot.201200110.

211. Pan A, Lahirii C, Rajendiran A, Shanmugham B. 2016. Computational analysis of protein interaction networks for infectious diseases. Brief Bioinform 17:57–526. https://doi.org/10.1093/bib/bbv059.

212. Prathiviraj R, Berchmans S, Chellapandi P. 2019. Analysis of modularity in proteome-wide protein interaction networks of Methanothermobacter thermotrophicus strain ΔH and metal-loving bacteria. J Proteins Peptom 10:179–190. https://doi.org/10.1007/s42483-019-00019-5.

213. Cui T, Zhang L, Wang X, He ZG. 2009. Uncovering new signaling proteins and pathways to host adaptation. NPJ Syst Biol Appl 5:213. https://doi.org/10.1038/nsb0509.

214. Kim WY, Kang S, Kim BC, Oh J, Cho S, Bhak J, Choi JS. 2008. SynechoNET: a bioinformatics model for predicting virus–host interactions in cyanobacteria. J Mol Biol 384:763–770. https://doi.org/10.1016/j.jmb.2008.02.022.

215. Wuchty S, Müller SA, Cauviel D, King WY. 2014. A highly abundant bacteriophage discovered in the unknown sequenced rice soil bacterium Pseudomonas putida KT2440. BMC Genom 10:S18. https://doi.org/10.1186/1471-2164-10-118.

216. Cui T, Zhang L, Wang X, He ZG. 2009. Uncovering new signaling proteins and potential drug targets through the interactome analysis of Mycobacterium tuberculosis. BMC Genom 10:118. https://doi.org/10.1186/1471-2164-10-118.

217. Edwards RA, McNair K, Faust K, Raes J, Dutilh BE. 2016. Computational approaches to predict bacteriophage–host relationships. FEMS Microbiol Rev 40:258–272. https://doi.org/10.1093/femsre/fuxv048.

218. Dutilh BE, Cassman N, Mcnair K, Sanchez SE, Silva GGZ, Bolling L, Barr JJ, Speth DR, Segurian V, Aziz RK, Felts B, Dinsdale EA, Mollik JL, Edwards RA. 2014. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. Nat Commun 5:4989. https://doi.org/10.1038/ncomms4989.

219. Shapiro JW, Putomi C. 2018. Gene co-occurrence networks reflect bacteriophage ecology and evolution. mBio 3011807–17. https://doi.org/10.1128/mBio.01870-17.

220. Lima-Mendez G, Van Helden J, Toussaint A, Leplae R. 2008. Reticulate representation of evolutionary and functional relationships between phage genomes. Mol Biol Evol 25:762–777. https://doi.org/10.1093/molbev/mss023.

221. Paez-Espino D, Eloe-Fadrowski EA, Pavlovolous GA, Thomas AD, Hunttemann M, Mikhailova N, Rubin E, Ivanova NN, Kyrpides NC. 2016. Uncovering Earth’s virome. Nature 536:425–430. https://doi.org/10.1038/nature19094.

222. Halder AK, Dutta P, Kundu M, Basu S, Nasipuri M. 2018. Review of computational methods for virus–host protein interaction prediction: a case study on novel Ebola–human interactions. Brief Funct Genom 17:381–391. https://doi.org/10.1093/bfgp/elx026.

223. Li Y, Zhang Z, Yang L, Xiao W, Li Y, Li S, Xin S, Cao P, Lu J. 2020. The MERS–CoV receptor DPP4 as a candidate binding target of the SARS-CoV-2 spike. iScience 23:101160. https://doi.org/10.1016/j.isci.2020.101160.

224. De L Chakraborty S, Mukhopadhyay A. 2020. Machine learning techniques for sequence-based prediction of viral–host interactions between SARS-CoV-2 and human proteins. Biomed J 43:438–450. https://doi.org/10.1111/bj.2020.08.003.

225. Lu H, Zhou Q, He J, Jiang Z, Peng C, Hong R, Shi J. 2020. Recent advances in the development of protein–protein interactions modulators: mechanisms and clinical trials. Signal Transduct Target Ther 5:213. https://doi.org/10.1038/s41398-020-0315-3.

226. Rapanoel HA, Mazandu GK, Mulder NJ. 2013. Predicting and analyzing protein–protein interactions between Mycobacterium tuberculosis and its human host. PLoS One 8:e67472. https://doi.org/10.1038/s41598-018-33822-9.
