Network-Based Analysis of Virulence Factors for Uncovering *Aeromonas Veronii* Pathogenesis

Hong Li
Hainan University
Xiang Ma
Hainan University
Yanqiong Tang
Hainan University
Dan Wang
Hainan University
Ziding Zhang
China Agricultural University
Zhu Liu (zhuliu@hainanu.edu.cn)
Hainan University

Research Article

Keywords: Aeromonas veronii, protein-protein interaction network, virulence factor, pathogenesis

Posted Date: April 26th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-439941/v1

Abstract

Background: *Aeromonas veronii* is a pathogen that causes serious harm to aquaculture. Virulence factors are its pathogenic basis, which could promote pathogens to colonize the host, evade host defense and so on. But because experimental verification of virulence factors is time-consuming and laborious, the number of known virulence factors is limited. In this past, most studies only focused on single virulence factor, resulting the biased interpretation for pathogenesis.

Results: In this study, a protein-protein interaction (PPI) network at genome-wide scale for *A. veronii* was first constructed. Then, virulence factors were predicted and mapped on the network. Topological characteristics of the virulence factors were analyzed. The results showed that the virulence factors had higher degree and betweenness centrality than the other proteins in the network. In particular, the virulence factors tended to interact with each other and were enriched in two network modules. One of the modules mainly consisted of histidine kinases, response regulators, diguanylate cyclases and phosphodiesterases, which played important roles in two-component regulatory systems and the synthesis and degradation of cyclic-diGMP. Furthermore, an interspecies PPI network between *A. veronii* and its host *Oreochromis niloticus* was also constructed. The structures and interacting sites of the virulence factors and host proteins were added to the interspecies PPI network. By analyzing the interspecies PPI network, we found that the virulence factors could competitively bind host proteins and some of the interacting sites of the virulence factors were shared by different host proteins. Drugs could be designed to target these sites and further prevent pathogen to interfere with host pathways.

Conclusions: Our results indicated that the virulence factors regulated the virulence of *A. veronii* by involving in signal transduction pathway and manipulate host biological processes by mimicking and competitively binding host proteins. Our results deepened the understanding for pathogenesis and had important theoretical significance for designing targeted antibacterial drugs.

Background

Aeromonas veronii is the main pathogenic bacteria of aquatic animals and widely found in freshwater and seawater[1]. It can cause ulcerative syndrome, hemorrhagic septicaemia and even mass mortality in aquatic animals such as *Oreochromis niloticus*[2], resulting in great economic losses to aquaculture industry. *A. veronii* also infects humans and has been classified into the quarantine objects of water quality and food safety in some countries[3, 4]. Virulence factors play an important role in the pathogenic process, which assist pathogens to adhere to and invade host cells, resist host immune defenses and compete for nutrients. Although virulence factors have been identified in many pathogens, the virulence factors in *A. veronii* remain elusive.

Virulence factors can be classified into three categories based on their subcellular localization, including cytosolic, membrane associated and secreted virulence factors[5]. Cytosolic virulence factors promote the rapid adaptation of pathogens to metabolic, physiological and morphological changes. Membrane

associated virulence factors contribute to the adhesion and evasion of pathogens to host cells. Comparatively, secreted virulence factors play more important roles. They can be delivered from pathogen cells into host cells or host environment[6, 7] and further directly participate in host biological processes by interacting with host proteins. Thus, discovering virulence factors, especially secreted virulence factors, is essential for understanding *A. veronii* pathogenesis and its relationship with the host.

Protein-protein interaction (PPI) networks have been demonstrated to be powerful in predicting potential virulence factors based on the “guilt-by-association” principle[8, 9]. By integrating PPI networks and known virulence factors[9], Zheng et al. identified the virulence factors of six species with high accuracy. Similarly, Cui et al. identified the virulence factors of three species by integrating PPI networks, known virulence factors and KEGG pathways[8]. In terms of network biology, PPI networks are also cornerstones to evaluate the functional importance of proteins. It has been shown that the proteins with high degree (hubs) or betweenness centrality (bottlenecks) tend to be essential proteins encoded by essential genes[10, 11]. Once the genes encoding hubs or bottlenecks are knocked out or mutated, many phenotypic traits will be affected, even resulting in death. For example, the lethality rate of yeast will increase approximately threefold after knocking out the genes encoding hubs compared with those encoding non-hubs[12]. Thus, many researchers are interested in computing and comparing the topology parameters of proteins in PPI networks. Moreover, PPI networks can be analyzed at the module level[13]. Within the PPI network, a module consists of physically or functionally related proteins, which are assembled together to perform a specific function. Further, different modules act synergetically to fulfill cellular functions. On the whole, constructing PPI networks can assist in identifying key proteins and understanding pathogenic mechanisms from a systems perspective. However, *A. veronii* PPI network at genome-wide scale is still not available.

Several high-throughput experimental methods, such as yeast two-hybrid screening and tandem-affinity purification coupled with mass spectrometry, have been developed to identify large-scale PPIs. But due to expensive, time consuming and labor intensive defects of experimental methods, only the PPI networks of several model organisms have been reported, such as *Arabidopsis thaliana*[14], *Saccharomyces cerevisiae*[15], *Caenorhabditis elegans*[16], *Drosophila melanogaster*[17], *Escherichia coli*[18] and *Homo sapiens*[19]. To complement the experimental methods, a plethora of computational methods have been developed. Among them, the interolog and domain-based methods are widely used. The interolog method is mainly based on the conservation of PPIs in different organisms[20]. Two proteins are predicted to interact in an organism if they have interacting homologs in another organism. While the domain-based method refers to that two proteins are more likely to interact if they contain interacting domains[21]. The PPI networks of many pathogens, such as *Ustilaginoidea virens*[22] and *Phomopsis longicolla*[23], have been successfully constructed based on these two PPI inference methods. In addition, these two methods have also been successfully applied to predict host-pathogen interspecies PPIs[22, 24].

In this study, we focused on the aquatic pathogen *A. veronii* and predicted the *A. veronii* PPI network. Potential virulence factors were predicted and mapped onto the network. The importance of the virulence factors were first evaluated based on network topology properties. Further, two modules enriched by the

virulence factors were identified, which played important roles in the process of *A. veronii* infection. To explore molecular mechanisms of pathogenicity, the interspecies PPI network between *A. veronii* and its host *O. niloticus* was also constructed. Three-dimensional structures and interacting sites were added to the interspecies PPI network, which provided more interaction details and could enhance the understanding of host-pathogen interactions. Finally, key residues of the virulence factors that frequently participated in the interactions with different host proteins were identified, which could accelerate the development of new antibacterial agents.

Methods

Construction of *A. veronii* PPI network

The interolog method was first used to infer the interactions between *A. veronii* proteins. Six organisms with large-scale experimental PPIs were selected as model organisms, including *A. thaliana*, *S. cerevisiae*, *C. elegans*, *D. melanogaster*, *E. coli* and *H. sapiens*. Protein sequences of the six model organisms were downloaded from the UniProt[44] database, and experimentally verified PPIs were collected from the BioGrid[45], IntAct[46], DIP[47] and MINT[48] databases. Additional PPIs of *A. thaliana* and *H. sapiens* were obtained from the TAIR[49] and HPRD[50] databases, respectively. Inparanoid Version 4.1[51] was used to identify the orthologs between *A. veronii* and the six model organisms. A stringent threshold (inparalog score = 1.0) was set. Furthermore, the orthologs were analogized to predict *A. veronii* PPIs based on experimentally verified PPIs of the six model organisms.

The domain-based method was also used to infer *A. veronii* PPIs. Experimentally verified domain-domain interactions as templates were collected from the 3did[52] and iPfam[53] databases. Potential domains of *A. veronii* proteins were identified by PfamScan[54] ($e \leq 1.00\times10^{-3}$). Three strict standards were adopted to improve the prediction accuracy of *A. veronii* PPIs[22]. To start with, the protein domains with length coverage < 80% were filtered. Then, the total length of all domains in a protein was required to cover ≥ 40% of the protein. Finally, two proteins were defined as a PPI only if each domain in one protein interacted with each domain in the other protein. As a result, the *A. veronii* PPI network was constructed based on the *A. veronii* PPIs predicted by the interolog and domain-based methods.

Assessment of *A. veronii* PPI network

Generally, two interacting proteins tend to have similar Gene Ontology (GO) annotations, similar gene expression patterns and the same subcellular localization. To assess the reliability of the predicted *A. veronii* PPI network, 1000 random networks were generated by randomly rewiring edges of the *A. veronii* PPI network, while preserving the degree distribution. Semantic similarities of GO terms of interacting proteins in the *A. veronii* PPI network and random networks were calculated by the R package GOSemSim[55], including biological process, molecular function and cellular component terms. Gene expression data of wide type as well as *argR*, *avrA*, *hfq*, *smpB* and tmRNA mutation in *A. veronii* from our previous studies (Supplementary Table S1) were used to evaluate the similarity of gene expression

patterns of interacting proteins, which was quantified by absolute Pearson correlation coefficient (PCC). Subcellular localization of each protein was predicted by pLoc-mGneg[56], which was designed for Gram-negative bacteria and included eight subcellular localizations, i.e., cell inner membrane, cell outer membrane, cytoplasm, extracellular, fimbrium, flagellum, nucleoid and periplasm.

Prediction of virulence factors

Known virulence factors, which have been proven to affect pathogen-host interactions, were collected from the PHI-base database[57]. Sequence alignments were performed between *A. veronii* proteins and the known virulence factors by BLASTP. An *A. veronii* protein was predicted as potential virulence factor if the sequence identity ≥ 40% and the coverage ≥ 80% when aligned with a known virulence factor.

Network characteristics analysis of virulence factors

The degree and betweenness centrality of virulence factors and other proteins in the *A. veronii* PPI network were calculated by the Cytoscape plugin NetworkAnalyzer[58], which is commonly used[59, 60]. The number of interactions between the virulence factors was counted. The same number of proteins as the virulence factors was randomly selected from the *A. veronii* PPI network and the number of interactions between the random proteins was also counted. This process was repeated 1000 times. The *A. veronii* PPI network was divided into modules by the Markov cluster algorithm (http://micans.org/mcl/). Only the modules with at least five nodes were further analyzed. Fisher' exact test was used to identify the modules enriched by the virulence factors and annotate the functions of modules.

Prediction of virulence factor-*O. niloticus* protein interactions

A virulence factor has the potential to interact with *O. niloticus* proteins only if it is translocated into host cell. Thus, secreted virulence factors were first predicted by EffectiveDB[61], which integrates various tools to recognize bacterial secreted proteins. Sequences and function annotations of *O. niloticus* proteins were downloaded from the UniProt[44] database. Inparanoid Version 4.1[51] was used to identify the orthologs between *O. niloticus* and above six model organisms, and potential domains of *O. niloticus* proteins were identified by PfamScan[54]. The interactions between the virulence factors and *O. niloticus* proteins were predicted based on experimentally verified PPIs of the six model organisms and experimentally verified domain-domain interactions. Fisher's exact test was used to perform functional enrichment analysis of *O. niloticus* proteins.

Structure modeling of virulence factor-*O.niloticus* protein interactions

Homologous template complexes of virulence factor-*O. niloticus* protein interactions were first searched in the PDB database[62] by BLASTP. Five criteria were considered[63-66]: (1) the alignment between each interacting protein and the template had ≥ 30% sequence identity and covered ≥ 40% of the interacting protein length; (2) the templates of two interacting proteins came from different chains of a protein complex structure in the PDB database and further constituted the template complex; (3) the template

complex with resolution below 5 Å was prioritized; (4) X-ray structure as template complex was preferred over NMR structure; (5) average sequence identity of two interacting proteins with the template complex was given priority over average coverage, except when several template complexes had similar sequence identity, in which case the template complex with a higher coverage was preferred. Further, five models for each protein were generated using Modeller[67] based on the template. Of them, the model with the lowest DOPE score was regarded as the best structure of the protein after truncating unaligned residues at the N and C termini. Finally, the complex structure of two interacting proteins was inferred based on the template complex. The residues from two interacting proteins were defined as interacting sites if their shortest atomic distance was ≤ 4.0 Å. The Jaccard similarity for two sets of interacting sites was calculated by taking the number of their intersection divided by the number of their union.

Results

A. veronii **PPI network**

To construct an *A. veronii* PPI network with high coverage, two common methods were used, including the interolog and domain-based methods. By the interolog method, 13201 *A. veronii* PPIs were obtained, involving in 1904 proteins. Of which, most of the *A. veronii* PPIs (79.74%) were derived from the model organism *E. coli*. On the contrast, the *A. veronii* PPIs derived from *A. thaliana* only accounted for 0.47%. By the domain-based method, 8328 *A. veronii* PPIs among 1479 proteins were obtained after filtering with the strict standards. As a result, 21418 *A. veronii* PPIs predicted by the interolog or domain-based method constituted the *A. veronii* PPI network, involving in 2494 proteins (Supplementary Table S2).

A. veronii **PPI network was of acceptable reliability**

To evaluate the quality of the *A. veronii* PPI network, 1000 random networks were generated. Semantic similarities of GO terms of PPIs were first calculated. Compared with those in any random network, the PPIs in the *A. veronii* PPI network had significantly higher biological process, molecular function or cellular component similarities (Wilcoxon test, $p < 2.20 \times 10^{-16}$; Fig. 1A-C). 22.71% of the PPIs in the *A. veronii* PPI network had a biological process similarity of 1, while in the random networks, the corresponding percentage was only 5.52–6.62%. Similar results were also observed for molecular function and cellular component annotations. The percentages of the PPIs sharing the same molecular function and cellular component annotations were 16.55% and 39.93% in the *A. veronii* PPI network, respectively. By contrast, the corresponding percentages in the random networks were 4.01–4.89% and 17.90-20.64%. These results meant that the *A. veronii* PPI network was of acceptable reliability.

Similarities of gene expression patterns of PPIs were calculated based on 18 samples. Likewise, the absolute PCC values in the *A. veronii* PPI network were significantly higher than those in any random network (Wilcoxon test, $p < 1.00 \times 10^{-3}$ for any random network), suggesting that the PPIs in the *A. veronii* PPI network tended to be co-expressed. Although the percentages of PPIs decreased as absolute PCC values increased in both the *A. veronii* PPI network and the random networks (Fig. 1D), the random

networks displayed a steeper decline when the absolute PCC value was above 0.5. Especially at the high PCC interval of 0.9-1.0, the percentage of the PPIs in the *A. veronii* PPI network was twice as much as that in the random networks (Fig. 1D). Moreover, the percentages of the PPIs with the same subcellular localization were also calculated. As a result, above 50% of the PPIs were co-localized in the *A. veronii* PPI network, whereas the co-localized PPIs only accounted for 38.62–40.33% in the random networks (Fig. 1E). These results further suggested that the *A. veronii* PPI network was of reasonable reliability.

Virulence factors had higher degree and betweenness centrality in the A. veronii PPI network

A total of 242 potential virulence factors were predicted, of which 195 virulence factors were mapped onto the *A. veronii* PPI network. The degree and betweenness centrality were compared between the virulence factors and other proteins in the *A. veronii* PPI network. The results showed that the virulence factors had significantly higher degree and betweenness centrality than the other proteins (Wilcoxon test, $p = 9.33\times10^{-10}$ for degree, Fig. 2A and $p = 3.04\times10^{-10}$ for betweenness centrality, Fig. 2B). Average degree and betweenness centrality of the virulence factors were 26.75 and 3.00×10^{-3}, respectively. By contrast, the corresponding values of the other proteins were 16.36 and 1.00×10^{-3}, respectively. It has been reported that the proteins with high degree or betweenness centrality play crucial roles in many cellular processes[25, 26], implying the functional importance of the virulence factors. The degree of 28 out of 195 virulence factors ranked in the top 10% of degree distribution, which are often called hubs. The average PCC values between 27 (96.43%) virulence factor and their interacting proteins exceeded 0.30, meaning that the 27 virulence factors were party hubs and they tended to interact with their partners simultaneously. Seven and five out of the 27 virulence factors involved in the biosynthesis of secondary metabolites and the biosynthesis of antibiotics, respectively (e.g., dihydrolipoamide dehydrogenase, pyruvate kinase and glycerol-3-phosphate dehydrogenase). The remaining virulence factors involved in RNA degradation, cell cycle, amino acid metabolism and so on.

Virulence Factors Were Enriched In Two Modules

The number of PPIs formed by 195 virulence factors was counted and a total of 486 PPIs were obtained. While for 195 proteins randomly selected from the *A. veronii* PPI network, they formed at most 261 PPIs and at least 49 PPIs in 1000 trials (Fig. 2C), which were much less than the real number of PPIs formed by the virulence factors. The result suggested that the virulence factors tended to interact. In the case, we guessed that the virulence factors should be enriched in certain network modules. To test the conjecture, the *A. veronii* PPI network was divided into 90 modules, involving 1331 proteins and 100 virulence factors. Of them, two modules were significantly enriched by the virulence factors (Fisher's exact test, $p = 2.36\times10^{-7}$ and 8.82×10^{-4}; Fig. 3).

One module consisted of 57 proteins, 33 out of which had biological process annotations and 17 out of which were virulence factors (Fig. 3A). We found that the module was significantly associated with the terms “phosphorelay signal transduction system”, “regulation of transcription, DNA-templated” and “signal transduction by phosphorylation” (Fisher’s exact test, $p = 1.80\times10^{-38}$, 1.30×10^{-13} and 5.56×10^{-4},

respectively). Especially, 16 and 15 out of 17 virulence factors were annotated with the terms “phosphorelay signal transduction system” and “regulation of transcription, DNA-templated”, respectively (one virulence factor was not annotated with any term). This was mainly because that most of the proteins in the module were the members of two-component regulatory systems, including KdpE, AdeR, ArcA, CheB, CheY, CpxR, OmpR and PhoB. Two-component regulatory systems are important mediators of signal transduction and control bacterial virulence[27]. Thus it can be envisaged that the module was essential for the virulence of *A. veronii* and could serve as a target for antimicrobial therapy in the future. In addition, we also analyzed the topology characteristics of the module in the *A. veronii* PPI network. Average degree of the proteins in the module was 23.61, which was higher than that in the *A. veronii* PPI network (17.18). After removing 17 virulence factors, average degree of the proteins in the module was even higher (24.38). The result indicated that the module connected more other modules and had a great effect on the *A. veronii* PPI network. We also analyzed the other module enriched by the virulence factors (Fig. 3B). It consisted of seven proteins, four out of which were virulence factors and could be secreted by type VI secretion system. But specific functions of the proteins in the module were unknown.

Virulence factors may manipulate host biological processes by mimicking and competitively binding host proteins

Although virulence factors could promote bacteria to enter host cells, evade or inhibit host immune responses, and obtain nutrition from hosts, it is not clear which virulence factors directly interact with host proteins. To this end, 40 (20.51%) secreted virulence factors were first predicted. Of which, 36 virulence factors could interact with 1461 *O. niloticus* proteins, forming 2200 interspecies PPIs (Fig. 4A; Supplementary Table S3). In the interspecies PPI network, 33 virulence factors and 383 *O. niloticus* proteins had at least two partners, reflecting the complexity of interspecies PPIs. Virulence factors, such as succinate dehydrogenase flavoprotein subunit (SdhA), thioredoxin 1 (Trx1), thioredoxin 2 (Trx2), S-adenosylmethionine synthetase (MetK), catalase, ATP-dependent Clp protease proteolytic subunit (ClpP) and peroxiredoxin 2 (Prx2), had higher degree in the interspecies PPI network (Fig. 4A), meaning that these virulence factors could interact with more *O. niloticus* proteins. For the *O. niloticus* proteins in the interspecies PPI network, they were mainly involved in “translation”, “cell redox homeostasis”, “protein folding”, “tricarboxylic acid cycle”, “glycolytic process”, “S-adenosylmethionine biosynthetic process”, “one-carbon metabolic process”, “ubiquitin-dependent protein catabolic process”, “ribosome biogenesis” and “glycerol ether metabolic process” (Fig. 4B; Fisher’s exact test, $p < 1.00\times10^{-3}$), implying that *A. veronii* could directly manipulate host metabolic process, component organization and homeostasis to achieve successful infection.

O. niloticus proteins, such as heat shock protein, elongation factor Tu, DNA-directed RNA polymerase subunit, Trx2, ribosomal protein S3, SdhA, peroxiredoxin 1 (Prx1), transcriptional regulator, MetK and ClpP, interacted with at least 5 *A. veronii* proteins. Trx is oxidoreductase and ubiquitously presents in all organisms. In vertebrates, Trx tends to form homo-dimer or homo-oligomer mediated by disulfide bonds to regulate redox[28]. Our result showed that *A. veronii* Trx1 and Trx2 interacted with *O. niloticus* Trx2. It is possible that *A. veronii* Trx1 and Trx2 mimic *O. niloticus* Trx2 to disturb *O. niloticus* redox. Similarly, MetK,

ClpP and Prx are also found in almost every organism. The three proteins perform functions by forming homo-oligomer, tetradecamer complex and homo-dimer, respectively[29–31]. In our data, *A. veronii* MetK, ClpP and Prx2 interacted with *O. niloticus* MetK, ClpP and Prx1, respectively. These results implied that it was an effective strategy for virulence factors to interfere with host defense responses by mimicking and competitively binding host proteins.

Structures And Key Interacting Sites Of Virulence Factor Trx1

Structures and sites of 61 interspecies PPIs were predicted, which were formed by 15 virulence factors and 47 *O. niloticus* proteins. The data were stored in https://drive.google.com/drive/folders/18cHNUOSSJ5ugmFUL1_1QLVossf5ybYUY?usp=sharing. As a case study, Fig. 5 shows the structures of the interactions between *A. veronii* Trx1 and four *O. niloticus* proteins, including Trx2 (Fig. 5A), thioredoxin-interacting protein (Txnip) (Fig. 5B), methionine sulfoxide reductase (Msr) (Fig. 5C), as well as endoplasmic reticulum resident protein 44 (ERp44) (Fig. 5D). The four interactions had 58.32%, 48.72%, 36.93% and 58.29% average sequence identity and 72.63%, 74.19%, 73.97% and 84.65% average coverage to their template complexes, whose PDB IDs were 1W89, 4LL4, 3PIN and 5XWM, respectively.

A. veronii Trx1 interacted with *O. niloticus* Trx2 by the 33th, 34th, 64th, 71th, 75-79th residues (Fig. 5A); interacted with *O. niloticus* Txnip by the 34th, 35th, 37th, 64th, 76-80th, 94-96th residues (Fig. 5B); interacted with *O. niloticus* Msr by the 28-32th, 34-41th, 44th, 61th, 63th, 64th, 70-80th, 93th, 95-99th residues (Fig. 5C); interacted with *O. niloticus* ERp44 by the 35th, 37th, 39th, 40th, 74-79th, 97th residues (Fig. 5D). The Jaccard similarity between any two sets of interacting sites was as high as 0.23–0.40, indicating that *A. veronii* Trx1 tended to bind to host proteins by the same interaction interface. Especially, the 76-79th residues of *A. veronii* Trx1 were involved in each interspecies PPI, which could be potential targets to develop new antibacterial agents.

Discussion

In recent years, more and more aquatic animal diseases are being caused by *A. veronii.* However, molecular mechanisms underlying its pathogenicity remain largely unknown. In this study, an intraspecies and an interspecies PPI networks were constructed based on the interolog and domain-based methods, which contributed to the identification of new virulence factors and global understanding of pathogenic mechanisms. To ensure the reliability of PPI networks, multiple strategies were adopted. For example, the coverage of protein domains was strictly limited when using the domain-based method. Especially, two proteins were defined as a PPI only if all domains from the two proteins interacted with each other. Although the GO annotation, gene expression pattern and subcellular localization information have demonstrated the accuracy of PPI networks, we still have to admit that the PPI networks have false positives and false negatives. The degree and betweenness centrality between the virulence factors and other proteins were compared in this study. To our knowledge, this was the first time to study the network characteristics of virulence factors.

By analyzing the interactions formed by the virulence factors, we found that the virulence factors tended to connect with each other and were enriched in two network modules. One of the modules consisted of 57 proteins, 17 out of which were virulence factors. While 9 out of the remaining 40 proteins directly interacted with the virulence factors. According to the “guilt-by-association” principle, the 9 proteins were likely to be also virulence factors, although they were not predicted based on the sequence homology. The 9 proteins included three copies of CheY, CreB, PhoB, CitB, CpxA, CheB and an unknown protein. Except for CpxA that is a histidine kinase, the other 8 proteins are response regulators in two-component regulatory systems. It has been reported that many two-component regulatory systems, such as PhoP/PhoQ and EnvZ/OmpR, play important roles in virulence[32–35]. Thus, the 9 proteins could be potential drug targets.

Among the 57 proteins, 5, 31, 6, 2 and 10 proteins were histidine kinases, response regulators, diguanylate cyclases, phosphodiesterases and unknown proteins, respectively. Similarly, based on the “guilt-by-association” principle, the 10 unknown proteins could also belong to certain one of the four types of proteins. Histidine kinase senses environmental stimulus and the corresponding response regulator mediates cellular response. The two proteins constitute the two-component regulatory system, which is a key mediator of bacterial signal transduction[36]. Diguanylate cyclase synthesizes cyclic-diGMP and phosphodiesterase degrades cyclic-diGMP[37]. Cyclic-di-GMP as the second messenger transmits extracellular signals to intracellular environment. Since histidine kinases, response regulators, diguanylate cyclases and phosphodiesterases co-exist in the same module, then implying that cyclic-di-GMP and two-component regulatory systems could work together to regulate *A. veronii* signal transduction. In *Xanthomonas campestris*, it has been demonstrated that cyclic-di-GMP binds to histidine kinase RavS to control two-component regulatory system RavS/RavR phosphotransfer[38]. While in *Legionella pneumophila*, two-component system Lpg0278/Lpg0277 modules cyclic-diGMP metabolism[39].

Each virulence factor interacted with 11 *O. niloticus* proteins on average, which may be one of the reasons that pathogens with smaller genomes can overcome host with larger genomes. The *O. niloticus* proteins targeted by the virulence factors were mainly involved in cellular metabolic process, component organization and homeostasis, implying that *A. veronii* successfully infected host by manipulating these cellular processes. It has been shown that group A *Streptococcus* delivers virulence factors into host cells during infection. The delivered virulence factors modulate host metabolism to cause endoplasmic reticulum stress and induce the formation of asparagine. Further, the asparagine can be sensed by group A *Streptococcus* to increase the rate of its growth[40]. In fact, to block nutritional source for pathogens, many host cells are normally in a metabolically quiescent state during pathogen infection. This forces pathogens have to reprogram host cell metabolism in its specific manner to obtain nutrients and energy[41]. In this process, virulence factors play an important role.

Our result showed that *A. veronii* virulence factors hijacked host pathways by mimicking *O. niloticus* proteins, which was a commonly used strategy in pathogen-host interactions[42]. Virulence factors could mimic host global proteins, domains or short linear motifs to compete endogenous interfaces of host[43].

In this study, we only focused on the mimicry of global proteins, which generated more tight interactions between virulence factors and host proteins. The virulence factors identified in this study could be used as preferred drug targets. ClpP was a typical example. Many researchers have begun to design antibacterial drugs based on ClpP[30], which demonstrated potential application value of the virulence factors. Taken together, our results could lay the foundation for the development of drugs and offer new hints for the treatment of bacterial diseases.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated during the current study are available in https://drive.google.com/drive/folders/18cHNUOSSJ5ugmFUL1_1QLVossf5ybYUY?usp=sharing and its supplementary information files.

Funding

This work was supported by the National Natural Science Foundation of China (32060153), the Hainan Natural Science Foundation (319QN161) and the Priming Scientific Research Foundation of Hainan University (KYQD(ZR)1929).

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

HL conceived the study. HL, XM and DW performed the analyses. HL and YT drafted the manuscript. ZZ revised the manuscript. ZL supervised the study. All authors read and approved the final manuscript.

Acknowledgements

The authors thank the lab members for experimental and analytical assistance.

References

1. Wang D, Li H, Khan WU, Ma X, Tang H, Tang Y, et al. SmpB and tmRNA orchestrate purine pathway for the trimethoprim resistance in *Aeromonas veronii*. Front Cell Infect Microbiol. 2020;10:239.
2. Dong HT, Techatanakitarnan C, Jindakittikul P, Thaiprayoon A, Taengphu S, Charoensapsri W, et al. *Aeromonas jandaei* and *Aeromonas veronii* caused disease and mortality in Nile tilapia, *Oreochromis niloticus* (L.). J Fish Dis. 2017;40(10):1395–1403.
3. Roberts MTM, Enoch DA, Harris KA, Karas JA. *Aeromonas veronii* biovar sobria bacteraemia with septic arthritis confirmed by 16S rDNA PCR in an immunocompetent adult. J Med Microbiol. 2006;55(Pt 2):241–243.
4. Mencacci A, Cenci E, Mazzolla R, Farinelli S, D'Alo F, Vitali M, et al. *Aeromonas veronii* biovar veronii septicaemia and acute suppurative cholangitis in a patient with hepatitis B. J Med Microbiol. 2003;52(Pt 8):727–730.
5. Sharma AK, Dhasmana N, Dubey N, Kumar N, Gangwal A, Gupta M, et al. Bacterial virulence factors: secreted for survival. Indian J Microbiol. 2017;57(1):1–10.
6. Green ER, Mecsas J. Bacterial secretion systems: an overview. Microbiol Spectr. 2016;4(1): VMBF-0012-2015.
7. Dangl JL, Horvath DM, Staskawicz BJ. Pivoting the plant immune system from dissection to deployment. Science. 2013;341(6147):746–751.
8. Cui W, Chen L, Huang T, Gao Q, Jiang M, Zhang N, et al. Computationally identifying virulence factors based on KEGG pathways. Mol Biosyst. 2013;9(6):1447–1452.
9. Zheng LL, Li YX, Ding J, Guo XK, Feng KY, Wang YJ, et al. A comparison of computational methods for identifying virulence factors. PLoS One. 2012;7(8):e42517.
10. Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M. The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. PLoS Comput Biol. 2007;3(4):e59.
11. Han JD, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, et al. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature. 2004;430(6995):88–93.
12. Jeong H, Mason SP, Barabasi AL, Oltvai ZN. Lethality and centrality in protein networks. Nature. 2001;411(6833):41–42.
13. Barabási A-L, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet. 2004;5(2):101–113.
14. Arabidopsis Interactome Mapping Consortium. Evidence for network evolution in an Arabidopsis interactome map. Science. 2011;333(6042):601–607.
15. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, et al. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. Nature. 2000;403(6770):623–627.
16. Li S, Armstrong CM, Bertin N, Ge H, Milstein S, Boxem M, et al. A map of the interactome network of the metazoan *C. elegans*. Science. 2004;303(5657):540–543.
17. Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, et al. A protein interaction map of *Drosophila melanogaster*. Science. 2003;302(5651):1727–1736.

18. Butland G, Peregrín-Alvarez JM, Li J, Yang W, Yang X, Canadien V, et al. Interaction network containing conserved and essential protein complexes in Escherichia coli. Nature. 2005;433(7025):531–537.
19. Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, et al. Towards a proteome-scale map of the human protein-protein interaction network. Nature. 2005;437(7062):1173–1178.
20. Walhout AJ, Sordella R, Lu X, Hartley JL, Temple GF, Brasch MA, et al. Protein interaction mapping in *C. elegans* using proteins involved in vulval development. Science. 2000;287(5450):116–122.
21. Wang TY, He F, Hu QW, Zhang Z. A predicted protein-protein interaction network of the filamentous fungus *Neurospora crassa*. Mol Biosyst. 2011;7(7):2278–2285.
22. Zhang K, Li Y, Li T, Li ZG, Hsiang T, Zhang Z, et al. Pathogenicity genes in *Ustilaginoidea virens* revealed by a predicted protein-protein interaction network. J Proteome Res. 2017;16(3):1193–1206.
23. Li S, Musungu B, Lightfoot D, Ji P. The interactomic analysis reveals pathogenic protein networks in *Phomopsis longicolla* underlying seed decay of soybean. Front Genet. 2018;9:104.
24. Remmele CW, Luther CH, Balkenhol J, Dandekar T, Müller T, Dittrich MT. Integrated inference and evaluation of host-fungi interaction networks. Front Microbiol. 2015;6:764.
25. Li H, Zhou Y, Zhang Z. Competition-cooperation relationship networks characterize the competition and cooperation between proteins. Sci Rep. 2015;5:11619.
26. Li H, Zhou Y, Zhang Z. Network analysis reveals a common host-pathogen interaction pattern in Arabidopsis immune responses. Front Plant Sci. 2017;8:893.
27. Tiwari S, Jamal SB, Hassan SS, Carvalho P, Almeida S, Barh D, et al. Two-component signal transduction systems of pathogenic bacteria as targets for antimicrobial therapy: an overview. Front Microbiol. 2017;8:1878.
28. Campos-Acevedo AA, Sotelo-Mundo RR, Perez J, Rudino-Pinera E. Is dimerization a common feature in thioredoxins? The case of thioredoxin from Litopenaeus vannamei. Acta Crystallogr D Struct Biol. 2017;73(Pt 4):326–339.
29. Garrido F, Estrela S, Alves C, Sanchez-Perez GF, Sillero A, Pajares MA. Refolding and characterization of methionine adenosyltransferase from Euglena gracilis. Protein Expr Purif. 2011;79(1):128–136.
30. Moreno-Cinos C, Goossens K, Salado IG, Van Der Veken P, De Winter H, Augustyns K. ClpP protease, a promising antimicrobial target. Int J Mol Sci. 2019;20(9):2232.
31. Teixeira F, Tse E, Castro H, Makepeace KAT, Meinen BA, Borchers CH, et al. Chaperone activation and client binding of a 2-cysteine peroxiredoxin. Nat Commun. 2019;10(1):659.
32. Schaefers MM. Regulation of virulence by two-component systems in pathogenic *Burkholderia*. Infect Immun. 2020;88(7): e00927-19.
33. Lu H-F, Wu B-K, Huang Y-W, Lee M-Z, Li M-F, Ho H-J, et al. PhoPQ two-component regulatory system plays a global regulatory role in antibiotic susceptibility, physiology, stress adaptation, and virulence in *Stenotrophomonas maltophilia*. BMC Microbiol. 2020;20(1):312.

34. Lv M, Hu M, Li P, Jiang Z, Zhang LH, Zhou J. A two-component regulatory system VfmIH modulates multiple virulence traits in *Dickeya zeae*. Mol Microbiol. 2019;111(6):1493–1509.
35. Bhagirath AY, Li Y, Patidar R, Yerex K, Ma X, Kumar A, et al. Two component regulatory systems and antibiotic resistance in Gram-negative pathogens. Int J Mol Sci. 2019;20(7):1781.
36. Groisman EA. Feedback control of two-component regulatory systems. Annu Rev Microbiol. 2016;70:103–124.
37. Hengge R. Trigger phosphodiesterases as a novel class of c-di-GMP effector proteins. Philos Trans R Soc Lond B Biol Sci. 2016;371(1707):20150498.
38. Cheng ST, Wang FF, Qian W. Cyclic-di-GMP binds to histidine kinase RavS to control RavS-RavR phosphotransfer and regulates the bacterial lifestyle transition between virulence and swimming. PLoS Pathog. 2019;15(8):e1007952.
39. Hughes ED, Byrne BG, Swanson MS. A two-component system that modulates cyclic di-GMP metabolism promotes *Legionella pneumophila* differentiation and viability in low-nutrient conditions. J Bacteriol. 2019;201(17):e00253-00219.
40. Baruch M, Belotserkovsky I, Hertzog BB, Ravins M, Dov E, McIver KS, et al. An extracellular bacterial pathogen modulates host metabolism to regulate its own sensing and proliferation. Cell. 2014;156(1–2):97–108.
41. Eisenreich W, Rudel T, Heesemann J, Goebel W. How viral and intracellular bacterial pathogens reprogram the metabolism of host cells to allow their intracellular replication. Front Cell Infect Microbiol. 2019;9:42.
42. Paulus JK, van der Hoorn RAL. Tricked or trapped-Two decoy mechanisms in host-pathogen interactions. PLoS Pathog. 2018;14(2):e1006761.
43. Samano-Sanchez H, Gibson TJ. Mimicry of short linear motifs by bacterial pathogens: a drugging opportunity. Trends Biochem Sci. 2020;45(6):526–544.
44. The UniProt Consortium. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 2019;47(D1):D506-D515.
45. Oughtred R, Stark C, Breitkreutz BJ, Rust J, Boucher L, Chang C, et al. The BioGRID interaction database: 2019 update. Nucleic Acids Res. 2019;47(D1):D529-D541.
46. Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, et al. The MIntAct project–IntAct as a common curation platform for 11 molecular interaction databases. Nucleic Acids Res. 2014;42(Database issue):D358-363.
47. Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D. The Database of Interacting Proteins: 2004 update. Nucleic Acids Res. 2004;32(Database issue):D449-451.
48. Licata L, Briganti L, Peluso D, Perfetto L, Iannuccelli M, Galeota E, et al. MINT, the molecular interaction database: 2012 update. Nucleic Acids Res. 2012;40(Database issue):D857-861.
49. Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, et al. The Arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. Genesis.

2015;53(8):474–485.

50. Goel R, Harsha HC, Pandey A, Prasad TS. Human Protein Reference Database and Human Proteinpedia as resources for phosphoproteome analysis. Mol Biosyst. 2012;8(2):453–463.
51. Sonnhammer ELL, Östlund G. InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. Nucleic Acids Res. 2014;43(D1):D234-D239.
52. Mosca R, Ceol A, Stein A, Olivella R, Aloy P. 3did: a catalog of domain-based interactions of known three-dimensional structure. Nucleic Acids Res. 2014;42(Database issue):D374-379.
53. Finn RD, Miller BL, Clements J, Bateman A. iPfam: a database of protein family and domain interactions found in the Protein Data Bank. Nucleic Acids Res. 2014;42(Database issue):D364-373.
54. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. Nucleic Acids Res. 2019;47(D1):D427-D432.
55. Yu G, Li F, Qin Y, Bo X, Wu Y, Wang S. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. Bioinformatics. 2010;26(7):976–978.
56. Cheng X, Xiao X, Chou KC. pLoc-mGneg: Predict subcellular localization of Gram-negative bacterial proteins by deep gene ontology learning via general PseAAC. Genomics. 2017.
57. Urban M, Cuzick A, Seager J, Wood V, Rutherford K, Venkatesh SY, et al. PHI-base: the pathogen–host interactions database. Nucleic Acids Res. 2019;48(D1):D613-D620.
58. Assenov Y, Ramirez F, Schelhorn SE, Lengauer T, Albrecht M. Computing topological parameters of biological networks. Bioinformatics. 2008;24(2):282–284.
59. Chen L, Xin X, Zhang J, Redmile-Gordon M, Nie G, Wang Q. Soil characteristics overwhelm cultivar effects on the structure and assembly of root-associated microbiomes of modern maize. Pedosphere. 2019;29(3):360–373.
60. Danielson RE, McGinnis ML, Holub SM, Myrold DD. Soil fungal and prokaryotic community structure exhibits differential short-term responses to timber harvest in the Pacific Northwest. Pedosphere. 2020;30(1):109–125.
61. Eichinger V, Nussbaumer T, Platzer A, Jehl MA, Arnold R, Rattei T. EffectiveDB–updates and novel features for a better annotation of bacterial secreted proteins and Type III, IV, VI secretion systems. Nucleic Acids Res. 2016;44(D1):D669-674.
62. Goodsell DS, Zardecki C, Di Costanzo L, Duarte JM, Hudson BP, Persikova I, et al. RCSB Protein Data Bank: Enabling biomedical research and drug discovery. Protein Sci. 2020;29(1):52–65.
63. Li H, Jiang S, Li C, Liu L, Lin Z, He H, et al. The hybrid protein interactome contributes to rice heterosis as epistatic effects. Plant J. 2020;102(1):116–128.
64. Yang X, Yang S, Qi H, Wang T, Li H, Zhang Z. PlaPPISite: a comprehensive resource for plant protein-protein interaction sites. BMC Plant Biol. 2020;20(1):61.
65. Li H, Yang S, Wang C, Zhou Y, Zhang Z. AraPPISite: a database of fine-grained protein-protein interaction site annotations for *Arabidopsis thaliana*. Plant Mol Biol. 2016;92(1–2):105–116.

66. Mosca R, Ceol A, Aloy P. Interactome3D: adding structural details to protein networks. Nat Methods. 2013;10(1):47–53.
67. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol. 1993;234(3):779–815.

Figures

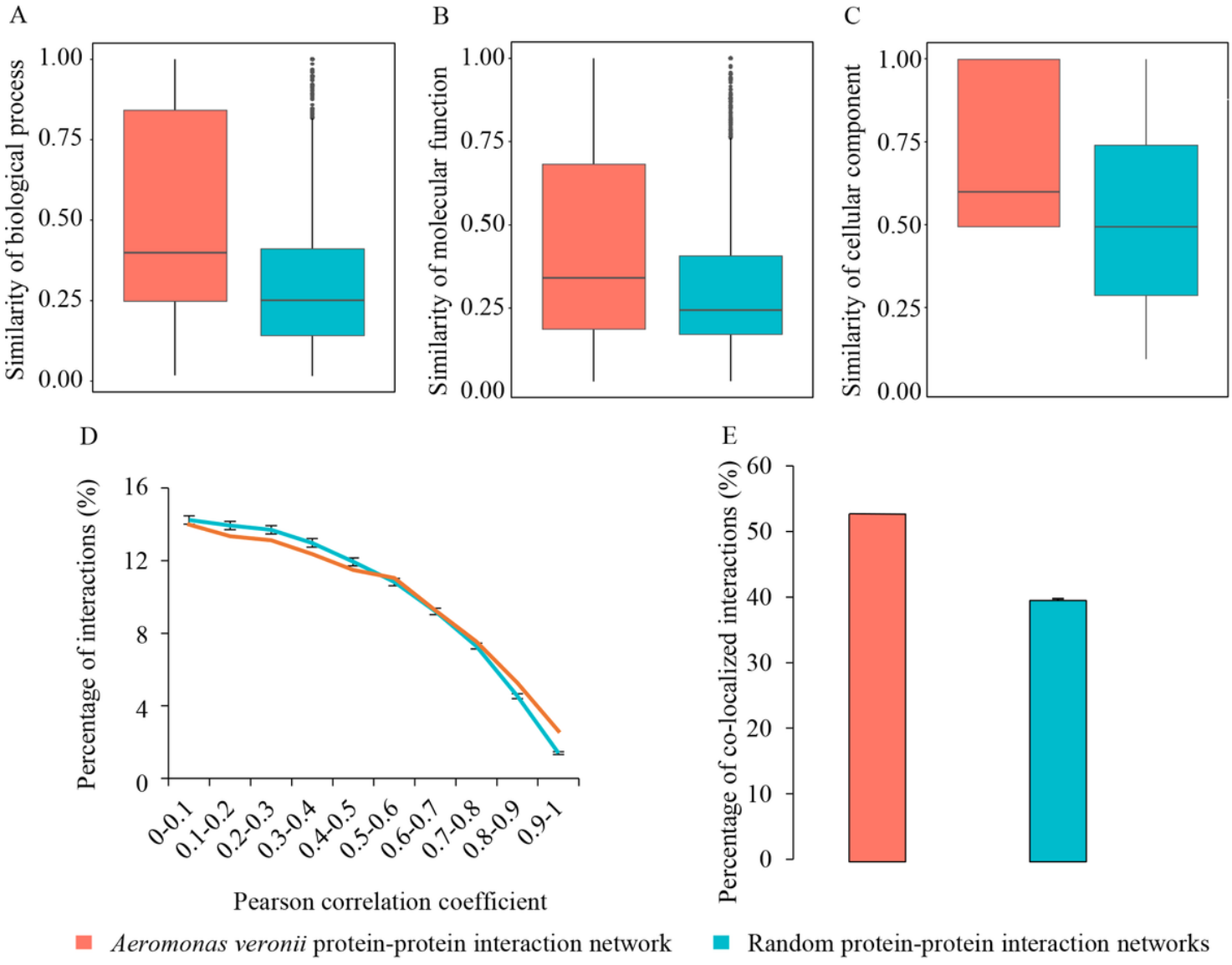

Figure 1

Reliability assessment of Aeromonas veronii protein-protein interaction (PPI) network. (A-C) Semantic similarities of gene ontology terms of interacting proteins in the A. veronii PPI network and one random network. The range of the box is from the first quartile to the third. The black line represents the median. The filled circle represents the outlier. (D) Percentages of interacting proteins with different Pearson correlation coefficients in the A. veronii PPI network and average percentages of those in 1000 random

networks. (E) Percentage of co-localized interacting proteins in the A. veronii PPI network and average percentage of those in 1000 random networks. The error bar in (D and E) represents the standard deviation of the percentages in random networks.

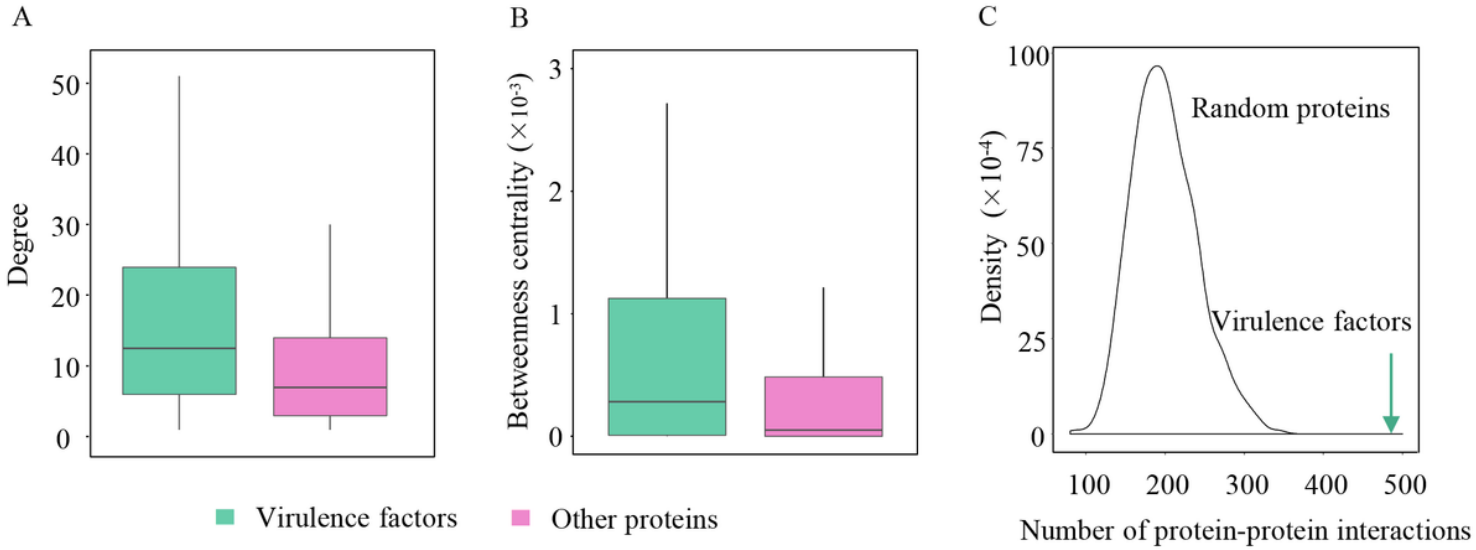

Figure 2

Analysis of topological properties. (A) Degree distributions of virulence factors and other proteins in the Aeromonas veronii protein-protein interaction (PPI) network. (B) Betweenness centrality distributions of virulence factors and other proteins in the A. veronii PPI network. The range of the box in (A and B) is from the first quartile to the third. The black line represents the median. (C) Statistics of the number of PPIs. The arrow points to the number of interactions formed by virulence factors. The same number of proteins as the virulence factors was randomly selected from the A. veronii PPI network and the number of interactions formed by the random proteins was counted. This process was repeated 1000 times. The curve represents the distribution of the number of interactions formed by the random proteins.

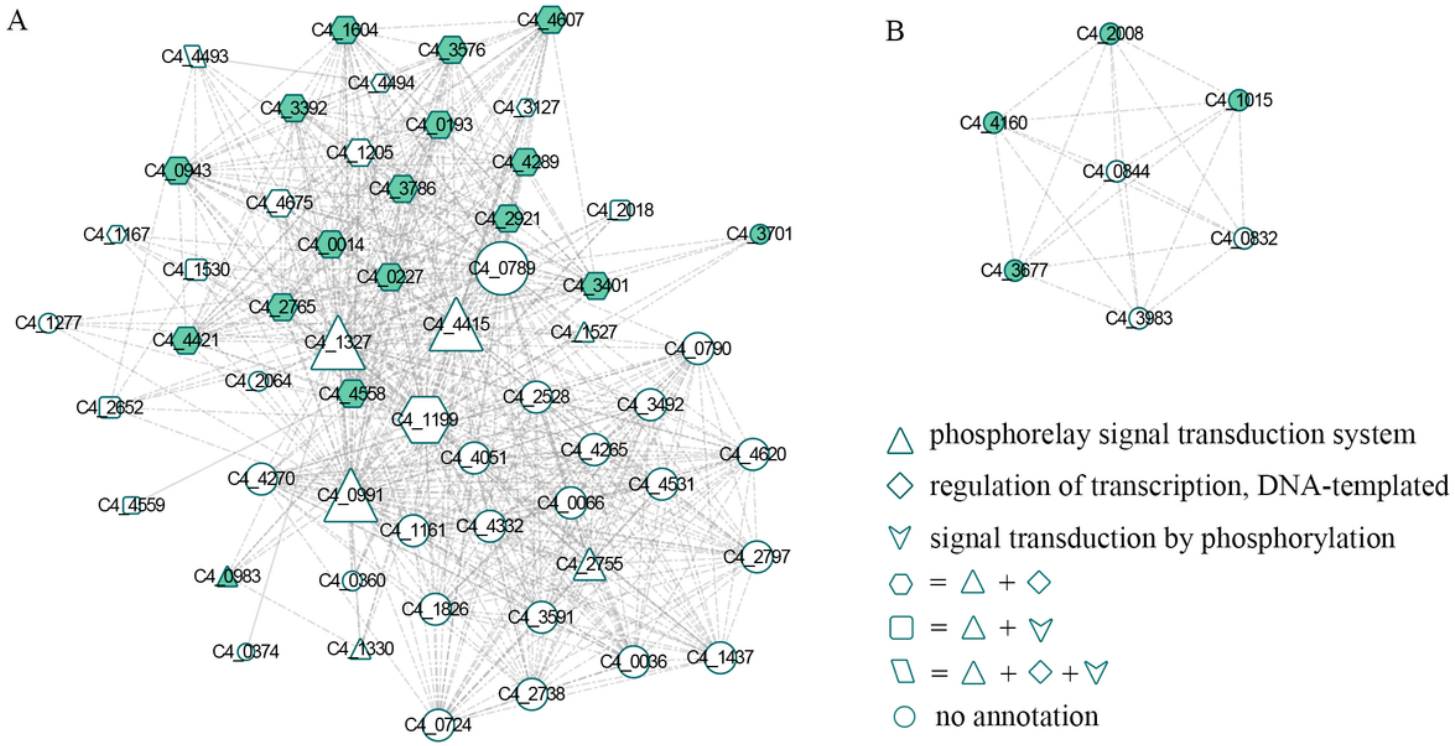

Figure 3

Two network modules enriched by virulence factors. The green and white nodes represent the virulence factor and the other protein, respectively. The larger node represents the protein with higher degree. The proteins with different function annotations are represented by different shapes. The solid, dash-dotted and parallel lines represent the interactions predicted by the interolog, domain-based and both methods, respectively.

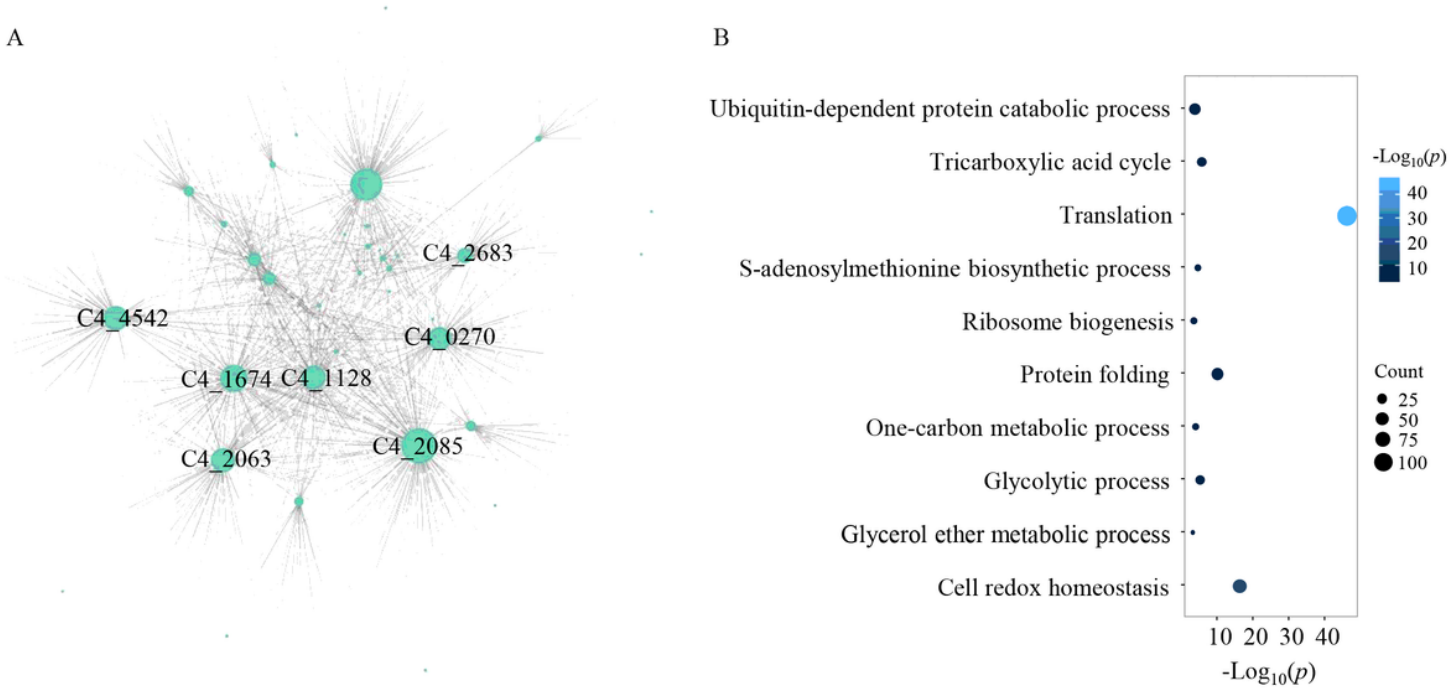

Figure 4

Interspecies protein-protein interaction (PPIs) between Aeromonas veronii and Oreochromis niloticus. (A) The interspecies PPI network consisting of 36 virulence factors and 1461 O. niloticus proteins. The green and white nodes represent the virulence factor and the O. niloticus protein, respectively. The larger node represents the protein with higher degree, such as C4_2085 (succinate dehydrogenase flavoprotein subunit), C4_4642 (thioredoxin 1), C4_2063 (thioredoxin 2), C4_1128 (S-adenosylmethionine synthetase), C4_0270 (catalase), C4_2683 (ATP-dependent Clp protease proteolytic subunit) and C4_1674 (peroxiredoxin 2). (B) Functional enrichment analysis of the O. niloticus proteins in the interspecies PPI network.

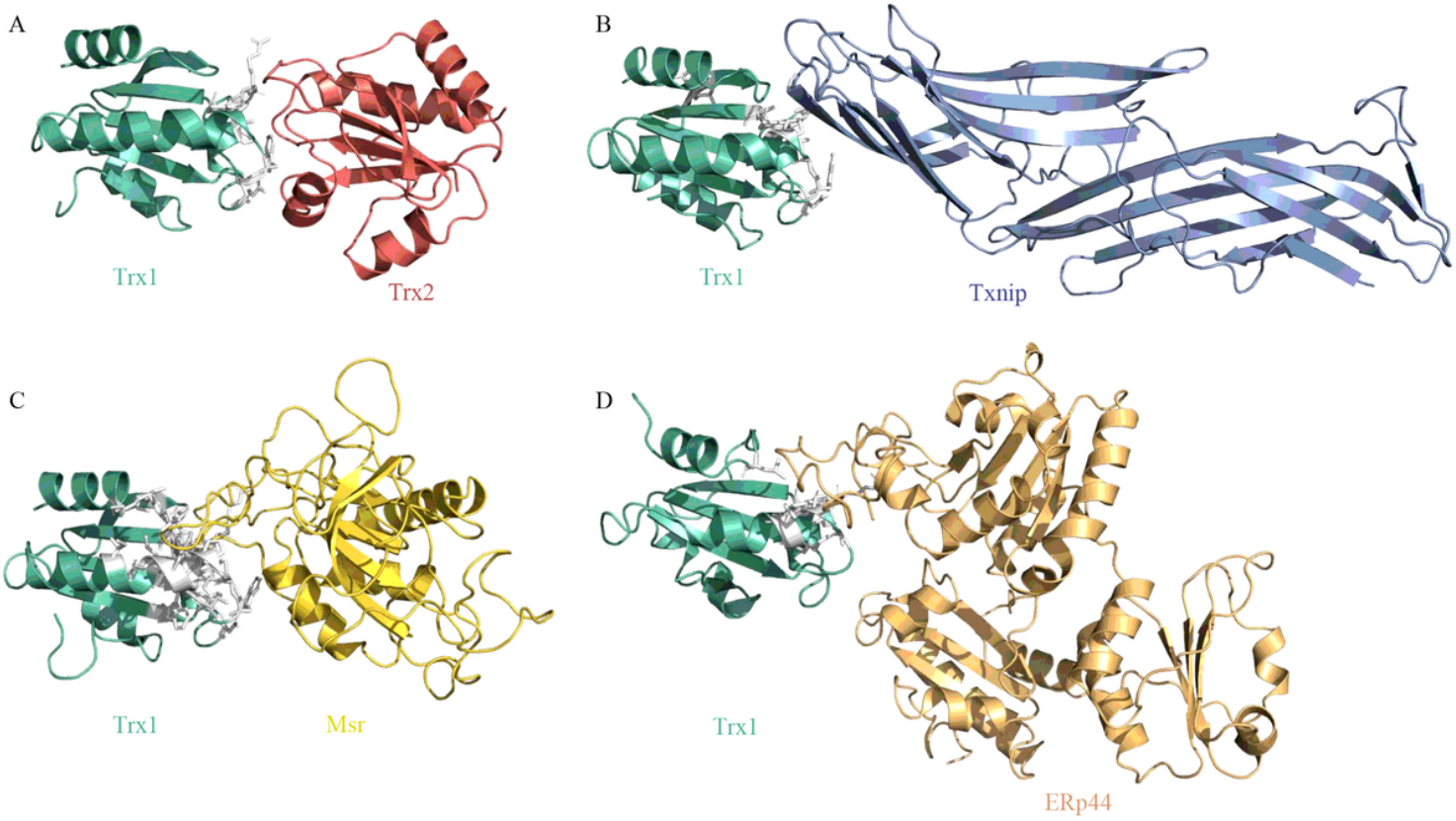

Figure 5

Protein complex structures formed by Aeromonas veronii thioredoxin 1 (Trx1) and four Oreochromis niloticus proteins. (A) thioredoxin 2 (Trx2), (B) thioredoxin-interacting protein (Txnip), (C) methionine sulfoxide reductase (Msr) and (D) endoplasmic reticulum resident protein 44 (ERp44). White sticks represent the interacting sites of Trx1. Trx1 binds to the four O. niloticus proteins using the same interface.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.xlsx
- TableS2.xlsx
- TableS3.xlsx