

A review on host–pathogen interactions: classification and prediction

R. Sen¹ · L. Nayak¹ · R. K. De¹

Received: 27 April 2016 / Accepted: 22 June 2016 / Published online: 29 July 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract The research on host–pathogen interactions is an ever-emerging and evolving field. Every other day a new pathogen gets discovered, along with comes the challenge of its prevention and cure. As the intelligent human always vies for prevention, which is better than cure, understanding the mechanisms of host–pathogen interactions gets prior importance. There are many mechanisms involved from the pathogen as well as the host sides while an interaction happens. It is a vis-a-vis fight of the counter genes and proteins from both sides. Who wins depends on whether a host gets an infection or not. Moreover, a higher level of complexity arises when the pathogens evolve and become resistant to a host’s defense mechanisms. Such pathogens pose serious challenges for treatment. The entire human population is in danger of such long-lasting persistent infections. Some of these infections even increase the rate of mortality. Hence there is an immediate emergency to understand how the pathogens interact with their host for successful invasion. It may lead to discovery of appropriate preventive measures, and the development of rational therapeutic measures and medication against such infections and diseases. This review, a state-of-the-art updated scenario of host–pathogen interaction research, has been done by keeping in

mind this urgency. It covers the biological and computational aspects of host–pathogen interactions, classification of the methods by which the pathogens interact with their hosts, different machine learning techniques for prediction of host–pathogen interactions, and future scopes of this research field.

Keywords Host–pathogen interactions · Pathogen informatics · Machine learning · In silico prediction · Secretion systems · Effector proteins

Introduction

The term ‘host–pathogen interaction’ refers to the ways in which a pathogen (virus, bacteria, prion, fungus, and viroid) interacts with its host. Pathogens adapt to the changes, and find alternative ways to survive and infect a host. They are infectious agents that cause diseases in a host body, when the host immune system fails against them. Questions like how the pathogens function, how their entry point into the host is facilitated through the biological barriers, and how they survive inside a host that is often under treatment or immunized for the same pathogen, can be answered by exploring host–pathogen interactions. Host–pathogen interactions can be described on the population level (virus infections in a human population), on the organismal level (pathogens infecting host), or on the molecular level (pathogen protein binding to a receptor on a human cell). However, before stepping into methodological details of host–pathogen interaction processes, a brief glimpse into the history of this research field is included here to sum up the how(s) and why(s) of recent advancements of this field.

Some of the earliest research works in the domain of host–pathogen interactions are (i) study of host–pathogen

✉ R. K. De
rajat@isical.ac.in

R. Sen
rishikasen_r@isical.ac.in

L. Nayak
losiana.t@isical.ac.in

¹ Machine Intelligence Unit, Indian Statistical Institute, 203, Barrackpore Trunk Road, Kolkata 700108, India

interaction in mouse typhoid caused by *Salmonella typhimurium* [141], (ii) genetic study of physiology of parasitism of the corn rust pathogen *Puccinia sorghi* [31], (iii) a correlation study of α -galactosidase production and host–pathogen interaction between *Phaseolus vulgaris* and *Colletotrichum lindemuthianum* [42], (iv) study of ultrastructural aspects of a host–pathogen relationship of a deuteromycetes fungus, *Pyrenochaeta terrestris* with two *Allium cepa* (onion) varieties with the help of electron microscopy [56], (v) fine structure study of principal infection procedure during infection of Barley by *Erysiphe graminis* [40], (vi) a study on proteins which obstructs the action of the polygalacturonases (polygalacturonide hydrolases, EC 3.2.1.15) released by the fungal plant pathogens *Fusarium oxysporum*, *Colletotrichum lindemuthianum*, and *Sclerotium rolfsii*. These proteins are extracted from the cell walls of Red Kidney bean hypocotyls, tomato stems, and suspension-cultured sycamore cells [1], (vii) a study on proteins secreted by plant pathogens, which impedes enzymes of the host having the ability to attack the pathogen. The study is conducted on an interaction system of a fungal pathogen (*Colletotrichum lindemuthianum*) and its host, the French bean (*Phaseolus vulgaris*) [2], (viii) a study on a single plant protein that efficiently hinders endopolygalacturonases secreted by *Aspergillus niger* and *Colletotrichum lindemuthianum* [46], (ix) a molecular basis study to showcase mutation of *Xanthomonas campestris* to overcome resistance in pepper (*Capsicum annuum*) [59], (x) a study on stress and immunological response in host–pathogen interactions [90].

Some recent research works have focused on (i) the basic notion of virulence and pathogenicity, which defines and suggests a classification system for microbial pathogens based on their capacity to cause damage as a consequence of the host's immune response [17], (ii) model organisms for host–pathogen interactions, i.e., *C. elegans* [70], *D. melanogaster* [91, 131] and zebrafish [53, 127] among others, (iii) molecular cross-talk of host–pathogen interactions where Type III secretion system is mentioned [108], (iv) novel studies involving epigenetics¹ [49], metallobiology [11], quantitative temporal viromics² [134], heterogeneity in same host tissue [14], and computational systems biology [36] of host–pathogen interactions.

All these investigations indirectly show us the trend of development of the host–pathogen interactions research field. The field has started with sporadic research works

of a pathogen and its interaction with a host. The earliest research has been done on host–pathogen interactions with respect to environmental factors, like light, temperature, season, and pathogen/host population among others. Later some organisms, like *C. elegans* and *D. melanogaster* have been found as model organisms to study the pathogen behavior of other complex hosts (human beings) due to their easy body plan, known genome structure and short life cycle. Gradually, certain proteins and then protein clusters have been marked for taking part in host–pathogen interactions. Moreover, definite classification has been found for the mechanism of host–pathogen interactions at the advent of recent developments in imaging and molecular biology techniques.

Moreover, some research works have defined and have given direction to the host–pathogen interactions research field. Discovery of distinct secretion systems [30, 47, 68, 100, 101, 135] has provided the basic background of host–pathogen interaction research. The concerned studies have spanned from genome locus [68] to biochemical and genetic evidence [88]. With discovery of PPI prediction methods [10], the chance of finding host–pathogen protein pairs and their interactions has become more prominent and such studies have given a different direction to the research field. Then methods have been developed for the machine learning based in silico prediction of secretion system associated proteins [4]. There are also a couple of newly proposed methods [54, 84], which provide new glimmer of hope to the research field in controlling pathogenesis in a host as described below.

- Secretion systems Type I [135], Type II [30], Type III [47] and Type V [100] have been discovered in 1980s, which have defined the base for host–pathogen interaction research.
- Kuldau et al. [68] have predicted 11 ORFs from virB locus in 1990. Based on a hydropathy plot, they have analyzed that nine of them encode proteins which may interact with membranes and may form a membrane pore or channel to mediate exit of the T-DNA copy. This is the first indirect indication of a distinct secretion system, later known as Type IV Secretion system (T4SS).
- Pukatzki et al. have functionally defined T6SS in 2006 [101].
- Mougous et al. in 2006 have provided biochemical and genetic evidence that a virulence-associated genetic locus of *P. aeruginosa*, termed as HSI-I, encodes a protein secretion apparatus (T6SS) [88].
- Machine learning-based prediction of PPIs have been done by Bock et al. in 2001 [10]. They have used Support Vector Machine (SVM) to train and predict interactions based on primary structure and related

¹A procedure through which genotypes give rise to phenotypes during development due to changes in underlying DNA sequences, i.e., histone modifications, DNA methylation, DNA silencing via noncoding RNAs and chromatin remodeling proteins.

²Temporal alterations in host and viral proteins throughout the course of a productive infection.

physicochemical properties. This work has provided a shift in research direction from genes to their protein counter parts and their nature of interaction.

- First ever machine learning-based prediction of Type III secretion system associated proteins have been done by Arnold et al. in 2009 by analyzing the amino acid composition and secondary structure composition of a few experimentally verified effector proteins at N-terminal [4].
- A few new studies and methods have proposed new avenues of future host–pathogen interaction research, i.e., a new way of studying host–pathogen interaction by dendritic cell subtypes [84] and chemoproteomic profiling of host and pathogen enzymes for finding candidates (proteases) to disrupt pathogenic mechanisms which have often boosted the host’s defense mechanisms directly or indirectly [54].

The present review tries to encompass the *in silico* prediction of host–pathogen interactions by machine learning and the related aspects. It has been organized into dedicated sections of classification of host–pathogen interactions, availability of host–pathogen interaction data, prediction of host–pathogen interaction domains, image processing-based research techniques, and conclusive remarks. There are several substrates and pathways whereby pathogens can invade a host. The human body has its own natural defense mechanism against some of the common pathogens in the form of an immune system that acts against these pathogens. Pathogens have the capability to adhere to host tissues, to evade host defenses, and to invade host cells. However, deeper understanding has revealed that each pathogen has its own variation of these themes [107]. Host–pathogen interactions take place between a host and a pathogen through the protein(s) and gene(s), and by disrupting normal functioning of pathway(s), forming biofilm(s), inhibiting macrophage activity and by other methods. In this review, we have briefly discussed the various probable factors that directly or indirectly contribute to host–pathogen interactions. Pathogens can either attack a host at the gene level by emitting RNA, or they can release proteins that could lead to pathogenicity, or they can inhibit the mechanism of macrophage. Some pathogens utilize the components of a host system to survive in the host. These components are called host factors. In a few cases, some factors of a pathogen can initiate the autophagy mechanism, which acts in favor of the host. The classification of the host–pathogen interactions is based on traditional pathogen invasion into host.

The review starts with categorization (Fig. 1) of pathogens, and makes a comprehensive list of diseases caused by them. The following section discusses the classification of host–pathogen interactions based on different

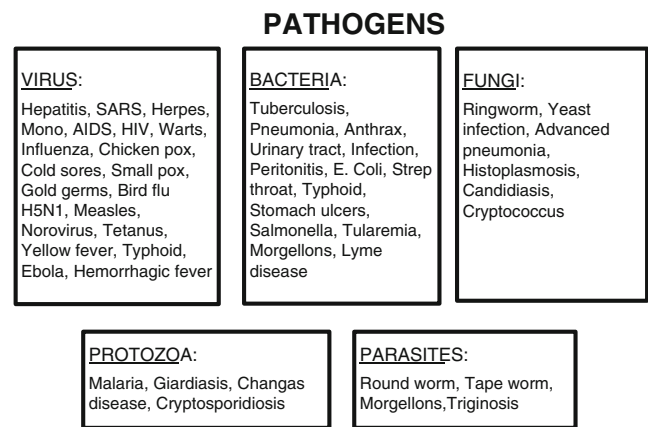


Fig. 1 Classification of some common pathogens and a list of diseases caused by them

biology-based reasoning. Then, the widely used *in silico* prediction methods in the domain of host–pathogen interactions are described. Moreover, an extensive list of the online repositories is given. The review concludes with a brief discussion that includes the merits and demerits of this research field in general, a few scopes for future research and concluding remarks.

Classification of host–pathogen interactions

The components of a host–pathogen interaction can be broadly classified into four stages, i.e., invasion of host through primary barriers, evasion of host defenses by pathogens, pathogen replication in host, and a host’s immunological capability to control/eliminate the pathogen. A pathogen can invade a host only after breaching the primary host defenses. Pathogens contain virulence factors that promote and cause disease. The greater the virulence, the more likely the disease will occur. We have classified the host–pathogen interactions according to these stages. A summary of the methods discussed in this review has been

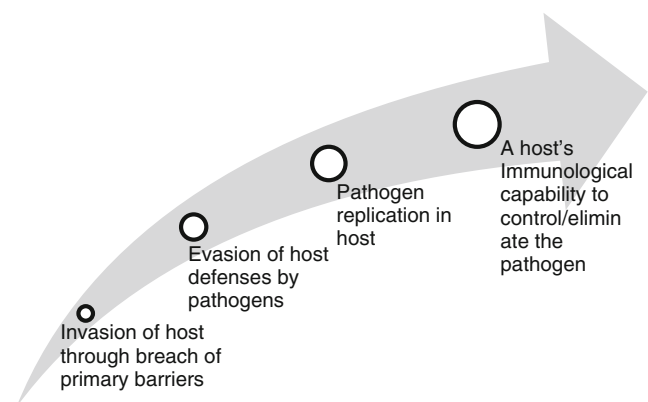


Fig. 2 Classification of host–pathogen interactions

diagrammatically represented in Fig. 2. However, in silico prediction methods used for detection of such interactions have been described in the Section “[Methods for prediction of host–pathogen interactions](#)”. The stages mentioned below are overlapping in nature. They do not have a clear boundary between them. The in silico prediction methods described later cannot be uniquely associated to only one of the stages. Their applicability spans over many or all the stages of host–pathogen interactions.

Invasion of host through breach of primary barriers

One of the main ways in which pathogens invade the host is via protein secretion. Pathogens, particularly Gram-negative bacteria, which cause pathogenesis in host, consist of secretion systems. These secretion systems release proteins, called effectors, into the body of the host when they come in contact with the host. There are at least six specialized secretion systems in Gram-negative bacteria. Type I, Type II, Type III, Type IV, Type V, and Type VI are the prominent ones based on their mechanisms of host infection. Details of these mechanisms can be obtained from Costa et al. [27]. Numerous secreted proteins are crucial in bacterial pathogenesis. We have described a few of them here, i.e., toxins, urease, and multivalent adhesion molecules.

Toxins are substances created by plants and animals that are poisonous to humans. Most toxins that cause problems in humans come from germs such as bacteria. Toxins can be small molecules, peptides, or proteins that are capable of causing disease on contact with or absorption by body tissues interacting with biological macromolecules such as enzymes or cellular receptors. These toxins, once in the body of the host, intervene with the normal functioning of the metabolism of host. Minimized toxin expression in a pathogen has a lesser effect on the stimulation of host’s TCR signaling pathway at the time of attack than that with higher toxin expression. It has been observed that viruses interact with different proteins of individual pathways temporally [117]. The molecules that are secreted by Gram-negative pathogens lead to damage of the host cells. The vesicle released from the enclosure of the growing bacteria serves as a container for the proteins and lipids of the Gram-negative bacteria. This suggests the importance of vesicle-mediated toxin delivery for the onset of infection in the host.

Effector proteins are secreted by pathogenic bacteria for their entry into the host. Effector proteins help a pathogen for invading host tissue, suppressing the host’s immune system, and often help the pathogen in its survival. Effector proteins are crucial for virulence. For example, in *Yersinia pestis* (the causative agent of plague), loss of the T3SS has rendered the bacteria completely avirulent [80]. Naive Bayes classifier and support vector machine have

already been applied to detect effector proteins of T3SS [4, 132]. More details regarding the methodology are given in the Section “[Methods for prediction of host–pathogen interactions](#)”.

Urease (an enzyme) plays an important role in *Mtb*–host interaction [23]. Urease is present in many species of *mycobacterium*, and its presence/absence is frequently used in the speciation of mycobacteria. Urease has been considered to be a virulence factor for several pathogenic microorganisms. Generation of ammonia by urease of urinary pathogens, such as *P. mirabilis*, has contributed to its pathogenesis due to its toxicity to renal epithelium, participation in complement inactivation, and promotion of urinary stone formation [13]. Urease of *H. pylori* alkalizes the bacterial micro-environment in the stomach and is toxic to stomach epithelium [119]. In the case of *Mtb*, urea is readily available to the bacteria in both its intracellular and extracellular locations within the host.

The multivalent adhesion molecule (MAM) is responsible for establishing high-affinity binding to host cells during early stages of infection [63]. MAM7 connects to a host via protein–lipid (phosphatidic acid) and protein–protein (fibronectin) interactions. MAM7 has been found on the outer membrane of the Gram-negative pathogens, which contributes to its virulence.

Evasion of host defenses by pathogens

In order to survive inside the host, the pathogens need to avoid the host defense mechanism. *Mycobacterium tuberculosis* (*Mtb*) showcases that it actively transcribes a number of genes involved in fortification and evasion from a host system [103]. Assessment of the genome of 58 strains of *Staphylococcus aureus* reveals that all the immune evasive proteins are present in all the strains but not all the surface proteins [81]. Remarkably, four strains have surface and immune evasion genes similar to human strain. On the other hand, the putative targets of these proteins vary in different hosts, which proposes that these proteins are not crucial for virulence. Signaling for anti-inflammation by glycolipids and host–system interaction may be considered a method of *Mycobacteria* to evade the host or may be playing a vital role in preventing extreme inflammatory response [128].

Pathogens often affect the essential pathways of their hosts with the aim of evading the host defenses. The NF- κ B family of transcription factors help in the development of APCs (antigen-presenting cells) and the lymphocyte [124]. Once the host is compromised, the NF- κ B pathway gets activated. HIV-1 mostly depends on its host for survival, as it has a few genes of its own. An integrated study of HIV-1 and human signal transduction pathways have been carried out to infer that most of these pathways may get effected by HIV virus during its life cycle [7]. It has assessed and

analyzed all possible paths (perturbed and unperturbed) starting from one protein (start point) terminating into another (end point).

Human proteins potentially targeted by EBV (Epstein–Barr virus) tend to be hubs in the human interactome. This is consistent with the hypothesis that hub protein targeting is an effective mechanism for viruses to convert pathways for their use [16]. Bacterial and viral pathogens are more inclined to interact with hub proteins, and the proteins that are central to multiple pathways in the network [38]. Certain cellular mechanisms, like cell cycle regulation and nuclear transport, participate in these interactions with a different set of pathogens. A study has identified 3073 human-*B. anthracis*, 1383 human-*F. tularensis*, and 4059 human-*Y. pestis* PPIs (protein–protein interactions) [39]. As suggested by Dyer et al. [38], these PPIs have occurred among those hub and bottleneck proteins. The extracellular hydrolytic enzymes, especially the aspartyl proteinases (Saps) secreted by *C. albicans*, are major factors of its pathogenicity [92]. Protein Chaperon 60 and 60.1 have a higher impact on activation of the cytokines than the protein Chaperon 60.2 [75]. In *Staphylococcus aureus*, proteins EsxA and EsxB act as virulent factors to enforce pathogenesis [15]. Mutants that do not secrete these proteins have been observed for failing to enforce strong pathogenesis. Among two closely related families of proteins, PE and PE_PGRS, PE_PGRS of *Mtb* activates a considerable humoral immune response but not PE [29]. Further study suggests that unlike PE, certain PE_PGRS genes are expressed during infection and antibody response. In case of Enterovirus, 71 genes out of 699 get differentially expressed significantly during infection [77]. Lack of the flagella gene in *Salmonella typhimurium* contributes to its virulence. Addition of flagella gene increases the cytotoxicity. However, it does not increase the production of IL-6 (interleukin-6) [96].

One of the crucial host defenses is the macrophage. Hence, macrophage inhibition is a factor using which the pathogen evades the host immune mechanism. Macrophage activation happens due to multiple components, i.e., gene(s) encoding receptor(s), signal transduction molecule(s), transcription factor(s), and bacterial component(s) that activate Toll-like receptor(s) (lipopolysaccharide, muramyl dipeptide, lipoteichoic acid, and heat shock proteins) [94] among others. Pathogens attempt to survive in the host by preventing the macrophages from acting on them. It has been found that pathogens disrupt the enzymatic activity in activated macrophages by disrupting the actin filament network [50].

It has been identified that falsatin is an endogenous protease inhibitor of *Plasmodium falciparum*. Analysis of inhibition of normal functionality of macrophages to engulf pathogens and ingest killed parasites due to the functioning of ornithine decarboxylase has been done by

Nairz et al. [93]. Due to pathogen-specific responses, interleukin-12 production is inhibited for *Mtb*, hence allowing the host to fight against the pathogen. It has been found that 26 to 37 proteins of HIV-1 are associated with MDM (monocyte-derived macrophages) derived from HIV [22]. Inhibition by *Mtb* can be avoided with the help of IFN- γ and transfection of LRG-47 [52]. It has been found that *Mtb* residing in macrophage switches to anaerobic growth [114] to evade host defense for a longer period of time.

The crosstalk of host–pathogen interactions is often governed by miRNAs [48, 111, 112]. The small RNAs, like siRNAs and shRNAs, also play a vital role in host–pathogen interactions. König et al. [62] have studied the association of siRNAs with host–pathogen interactions. They have explored it by combining genome-wide siRNA analysis along with the knowledge from human interactome databases. Pathogens have short linear motifs (SLiM) that have high similarity with host SLiMs. Motif mimicry is used by pathogens to rewire host signaling pathways by co-opting SLiM-mediated protein interactions to affect the host systems [130].

Pneumolysin (an enzyme) is a key virulence factor [78]. It activates multiple genes and signal transduction pathways in eukaryotes. Cytolytic effect of Pneumolysin contributes to lung injury and neural damage. It sometimes induces apoptosis in neurons and other cells. It can also trigger host-mediated apoptosis in macrophages, thus magnifying extermination of pathogens.

Pathogen replication in host

For surviving inside a host, pathogens have multiple ways to facilitate their growth by speedy replication. First of all, they need a few genes and proteins to survive effectively in the host, while many more genes and proteins are required for their survival outside the host. A study on the metabolic network of the pathogen *Salmonella typhimurium* has revealed 1083 genes catalyzing 1087 metabolic and transport reactions. This suggests that a minimal set of potent metabolic pathways within *Salmonella typhimurium* is required for its favorable replication of *Salmonella typhimurium* within the host [104]. Erythrocytic malaria parasites need proteases for a number of their cellular processes [98] in order to survive in the host.

Pathogens have evolved strategies to promote their survival by performing hijacking of the host cells they infect. Viruses implant their DNA sequence into the normal sequence of these hosts in the hope of their better survival [105] inside the hosts. A genome of the strain of *Mtb*, H37Rv, made up of 4000 genes comprising 4,411,529 base pairs, has a high guanine and cytosine content [24]. In this genome, 194 genes are required for the growth of *Mtb* [110]. A large number of these genes is unique to mycobacteria

and its closely related species. This leads to the fact that the mechanism of infection of *Mtb* is different from other pathogenic species.

Some pathogens even respond to more than one micro-environment for their replication and survival. The genes responsible for Smn (secretion in mycobacterium) protein secretion in a mutation of *Mtb*, which is *Mycobacterium smegmatis*, are homologs of their *Mtb* counterpart [26]. This suggests that some strains may have similar secretion mechanisms. Four essential gene products (Sm3866, Sm3869, Sm3882c, and Sm3883c) are needed for Smn secretion. *Mtb* exists in various metabolic states. This fact indicates that it may be responsive to more than one micro-environment [45].

The genome of *Mycobacterium tuberculosis* possesses a large family of Ser/Thr protein kinases (STPKs). STPKs have been found to play an important role in cell division and cell envelope biosynthesis [87]. The outer membrane of the bacteria facilitates the interaction between a host and a pathogen [67]. *C. albicans* have the capability to colonize and infect the majority of the tissues of the human host, which indicates that it can have functionally distinct proteinases (enzymes performing proteolysis) so as to have enough flexibility to multiply and survive in the host.

Sometimes a host itself unknowingly facilitates/inhibits the survival of its pathogens. These facilities are referred to as the host factors. These factors help in pathogen replication, transcription, integration, growth, 198 propagation, pathogen entry, and host–pathogen interactions among others. A set of 295 cellular cofactors (of host) are essential for replication of influenza virus in the early stage [61]. Among these cofactors, 181 are highly significant in host–pathogen interactions, 219 help in efficient influenza virus growth, 23 have role in viral entry, and ten are required for post-entry steps of virus replication. Small molecule inhibitors of multiple factors, including vATPase and CAMK2B, go against influenza virus replication. A set of 116 Dengue Virus Host Factors (DVHF) are needed for the propagation of DENV-2 (dengue virus type 2) [115]. Among 82 human homologs of dipteran DVHF, 42 have been identified to be human DVHF. A set of 311 host factors have been found to be responsible for the growth of HIV-1 [143]. Considering HIV dependency factors obtained previously in [12, 143], it is observed that the cardinality of the set of intersection is 311 host factors. Six newly identified host factors are AKT1, PRKAA1, CD97, NEIL3, BMP2k, and SERPINB6 [143]. A set of 250 such factors in HIV has been identified [12]. Rab6 and Vps53 play a role in viral entry, and TNPO3 is important for viral integration and Med28 for viral transcription. HDF genes show a stronger presence in the immune cell, thus allowing the viruses to evolve in the host cells that perform the life-cycle functions needed for them to survive. A set

of 213 host factors and 11 HIV-encoded proteins have been found to be responsible for HIV-1 replication [12]. Among them, a few proteins help in regulation of ubiquitin conjugation, DNA damage response, proteolysis, and RNA splicing. Forty new factors play a vital role in the process of initiation and/or kinetics of DNA synthesis. Fifteen proteins with different functions have been found to play a significant role in nuclear import or viral DNA integration.

Pathogens like *M. laprae* cannot survive independently. Hence, they convert the glial cells of a host into progenitor cells and using these progenitor cells, it can survive and spread infection inside the host [55]. It alters the genetic structure of the adult Schwann cells to form the progenitor cells. However, it is still unknown how long *M. laprae* can survive in the de-differentiated Schwann cells, as they will eventually differentiate back into adult Schwann cells.

Often apoptosis of host factors has been found to be involved in bacterial growth and sustenance inside host [144]. Apoptosis contributes to the processes of the host-cell deletion method, triggering the inflammation and defense mechanism. Apoptosis by the pathogen *Bordetella pertussis* allows *Bordetella* to survive in the introductory stages of infection. After the pathogen has successfully colonized the tissue of the host, it stops producing the toxin adenylate cyclase hemolysin.

Biofilm formation plays a major role in host–pathogen interactions. This is a mechanism of pathogens by which they form a biofilm for their survival in the host, often utilizing degraded host proteins. *Leucobacter chromiireducens* subsp. *solipictus* strain TAN 31504 forms biofilm. Exposure to TAN 31504 leads to change in a few innate immunity-related genes in *C. elegans* [89]. Esp (a serine protease secreted by *S. epidermidis*) degrades 75 proteins of *Staphylococcus aureus* by proteolytic activity, which include 11 proteins essential for the formation of biofilm [121]. Esp also degrades several human receptor proteins involved in colonization and infection by the pathogen for the benefit of the host.

A host's immunological capability to control/eliminate the pathogen

In order to prevent occurrence of infection/disease, the host body launches immune response with respect to the pathogenic invasion, i.e., high expression of certain genes [122], autophagy [118, 129], role of dendritic cells [84, 106], glycoconjugates [86, 87], and iron [32, 93] in activation/alteration of host immune system.

Host genes play an important role in its (hosts) immune response. Mutated β -catenin homolog bar 1 or homeobox gene *egl-5* of *C. elegans* has resulted in defective response and hypersensitivity to *Staphylococcus aureus* [57]. Bar-1 and the *egl-5* genes function parallel to the immune response

pathway taken up by *C. elegans*. Over-expression of egl-5 resulted in the modification of NF- κ B-dependent TLR2 (Toll-like receptor 2) signaling in epithelial cells, suggesting the role played by these two genes in immune defense of a host. Pro-16 in E cadherin is responsible for host specificity towards the human pathogen *Listeria monocytogenes* [73]. E-cadherin of mouse, which is 85 % similar to E-cadherin of human, denotes the entry of bacterial pathogen, *Listeria monocytogenes*, by not allowing E-cadherin to interact with bacterial surface protein internalin. If Proline (Pro) in the position 16 of amino acid in human is replaced by Glutamic acid (Glu), then interaction with internalin is disabled. However, in mouse, if Glu is substituted by Pro, then interaction with internalin is enabled. On *Mtb* interaction with mice, a group of 67 genes in an immuno-competent host has shown a higher level of expression than the immuno-deficient host often in 21 days. This shows that 67 genes are responsible for immunity of mice (host) [122].

Autophagy is another mechanism of the hosts defense against pathogen. Autophagy can be used in the elimination of *Mtb* [129]. LRG-47 initiates autophagy according to the study carried out by Singh et al. [118]. IRGM (immunity-related GTPase family M protein) also plays role in autophagy and degradation of intracellular bacillary load.

Dendritic cells (DCs) play a vital role in the activation of the immune system on encountering a pathogen [106]. DCs are summoned to the lamina propria of the small intestine after bacterial infection. The number of DCs summoned depends on the pathogenicity of microorganisms confronted. Infection stimulates the release of a variety of soluble factors, including chemokines, which facilitate the summoning of DCs, and cytokines that are strong arbitrators of DC activation. Pathogens, viruses, and their components can activate DCs directly. One of the important characteristics of DCs is their ability to migrate. During some infections, this property may have a harmful as well as a favorable side. Relocation of pathogen-laden DCs from the periphery into lymph nodes leads to the activation of T cells. On the other hand, this contributes to the spread of infection within the host.

Glycoconjugates can alter the immune system of the human body. Immunomodulatory components of *Mtb* are phosphatidyl-myo-inositol (PMI), lipomannan (LM), and lipoarabinomannan (LAM). Apart from LM and LAM, mannose also contributes to the synthesis of multiple glycosylated proteins and also polymethylated polysaccharides in *Mycobacteria* [86]. These molecules are synthesized by both pathogenic and non-pathogenic species. Many of the genes involved in biosynthesis of these glycoconjugates are important for survival of *Mycobacteria* [109, 110]. Only serine-threonine kinases have been predicted to take part in the regulation process of Mycobacterial glycosyltransferases [3, 87]. The interaction of *Mycobacteria* with the

pattern recognition receptors may be an influencing factor for the functioning of the inflammatory signals, hence determining the way in which the immune system reacts [3, 87].

Iron plays a crucial role in the secretion of cytokines and in the activity of the transcription factors, affecting the immune response [32, 93]. Iron homeostasis is controlled by immune cell-derived mediators and acute-phase proteins. An effective method of host defense is to restrict the supply of iron to the pathogens. Pathogens have evolved to utilize iron, as it is found abundant in the host. The control of iron homeostasis is one of the main issues, as it can be controlled by the host or the pathogen for their benefit.

With such diverse mechanisms involved at each step of pathogen infection, predicting host–pathogen interactions are extremely crucial. However, prediction of interactions among the huge number of host and pathogen proteins do pose a real-time experimental problem. Hence, many in silico prediction methods have been devised to abate such issues. They effectively provide the primary screening of the possible interactions and provide a list of highly probable interactions, which can then be experimentally verified. In the following section, we have listed and described a few of these.

Methods for prediction of host–pathogen interactions

Predictions in the domain of host–pathogen interactions play a vital role in designing rational-therapeutic measures including drugs. Sometimes, experimental procedures can be cumbersome, time-consuming, and expensive. Experimenting with all possibilities takes a lot of time. Prediction methods with the help of machine learning can overcome such problems. They can be used to predict the putative data first, which satisfies certain conditions. Then the predicted set can be verified experimentally, which will engage far less time and resources. The respective subsections describe some of the widely used techniques for in silico prediction of host–pathogen interactions. One or more of these methods can be used for prediction of genes, proteins, factors, and pathways among others of both the host and pathogen. Experimental- and data-related aspects of these techniques have been covered in Section “[Classification of host–pathogen interactions](#)”.

Biological reasoning based prediction of host–pathogen interactions

The most extensively explored way by which a pathogen interacts with the host is by PPIs. Pathogen proteins interact with host proteins for invading the host. Proteins of a

pathogen can affect a host and its environment in multiple ways. They can directly bind with host protein(s) and affect downward cascades of reactions preventing normal function(s) of host. They can even compromise a host's immunological defenses by misguiding and weakening it. They can even utilize the components of a crumbling harsh anaerobic environment of a immune-compromised host. Hence, predicting the putative PPIs between a pathogen and its host(s) is of paramount importance. In order to foretell whether a host protein can interact with a pathogen protein or vice-versa, the following categories of methods can be used.

Homology-based prediction

An interaction between a pair of proteins in one species is anticipated to be conserved in its related species [79]. Prediction of host–pathogen PPIs in *Homo sapiens* (as host) and *Plasmodium falciparum* (as pathogen) [64] considers interaction templates of human and *P. falciparum* genomic sequences to bring out the probable set of PPIs. Then a homology detection algorithm as shown in Fig. 3 is applied to these PPIs to filter out non-homologous ones. The new set thus formed is made to pass through the filter of stage-specific and tissue-specific expression data of *P. falciparum* and *Homo sapiens* respectively, and further filtered using the concept of predicted localized data. A study by Lee et al. [74] has considered orthologous pair of genes from 18 different species to predict PPIs. Further analyzing them, 81 genes are found to be conserved in all the 18 species and 243 genes are missing in *P. falciparum* but found in the rest of the 17 species. Hence, these 81 genes and their related PPIs are probably conserved.

Homology-based approaches to host–pathogen PPI prediction are widely used for their sheer simplicity and biological background support. Since the data needed for implementing the prediction are only the template PPIs and

protein sequences, these approaches are adaptable and can be applied to multiple different host–pathogen systems.

Similar is the case of molecular interaction between GBP (galactose-binding protein) and LPS (Gram-negative bacterial lipopolysaccharide). GBP from *Carcinoscorpius rotundicauda* performs as an anti-microbial defense [76]. Most importantly, GBP shares architectural and functional homology to human proteins. Therefore, there is a probability of some human protein and LPS interactions. Moreover, there are 6 Tectonic domains containing LPS binding sites in GBP. GBP acts as a bridge between LPS and CRP (C-reactive protein) by indulging in GBP-LPS and GBP-CRP interactions with the aim at forming a stable pathogen recognition molecule. These interactions have indicated that Tectonin domains can differentiate between host and pathogen proteins.

Homology-based approaches have their own set of weaknesses. In an infection, two proteins in a predicted PPI may actually have very low probability to be present together. Therefore, host–pathogen PPIs predicted completely on the homology basis, without taking into consideration other biological properties of the proteins involved, may not be very dependable. Further information is needed to increase the accuracy of the prediction. An investigation by Wuchty and Stefan [138] has described filtering of the PPIs predicted by the homology-based approach using a Random Forest classifier. Then the result has been filtered according to expression and molecular characteristics. It has led to a potent subset of proteins that indeed interact.

Structure-based prediction

When a pair of proteins has structures that are similar to a known interacting pair of proteins, it is justifiable to believe that the former are likely to interact in a way similar to the latter. Likewise, several investigations have used structural information to recognize the similarity between query proteins (i.e., proteins in the host and pathogen) and template PPIs (i.e., known interacting protein pairs), and conclude that host–pathogen protein pairs, which match some template PPIs, indeed interact. The method is depicted in Fig. 4.

A computational method for prediction of PPIs representing host–pathogen interactions has been devised by Davis et al. [28]. Their proposed method has first scanned the host and pathogen genome, searched for structural similarity to the already known protein complexes, and then analyzed their probable interactions, using the physical structures of the proteins. The result finally has undergone a filtering by tissue-specific expression data of host proteins and stage-specific expression data of pathogen proteins, leading to a potent set of proteins that have a high probability to interact.

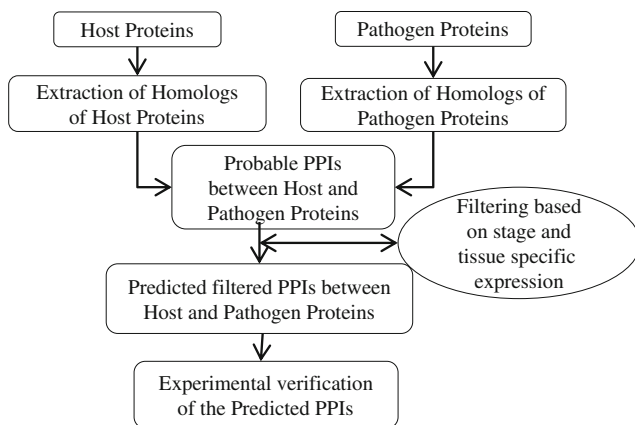


Fig. 3 Homology-based predictions of host–pathogen interactions

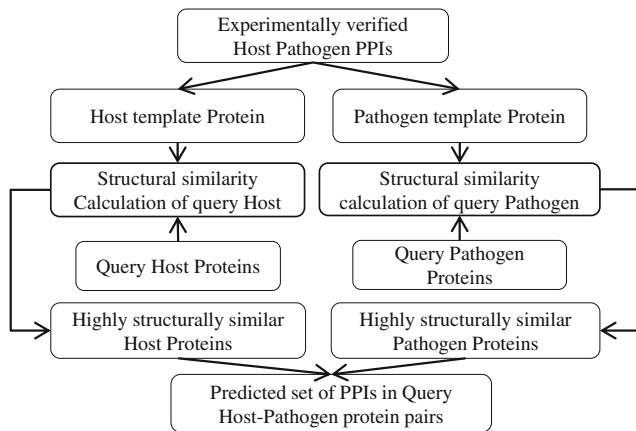


Fig. 4 Structure-based predictions of host–pathogen interactions

Mapping of PPIs between the dengue virus and its human and insect host has been carried out by Doolittle et al. [34]. They have also predicted the interactions depending on structural similarity of the host and the pathogen proteins. It has also focused on predictions relevant to stress, unfolded protein response and interferon pathways. Another work by Dolittle et al. [33] has predicted PPIs between HIV-1 and *Homo sapiens* based on structural similarity. It has modeled a network of interactions between HIV-I and human proteins. Structurally similar proteins from host and HIV-1 have been retrieved and from this structurally similar set of proteins, the known interactions have been mapped. The resultant subset has again been screened with factors like cellular co-localization and RNAi screen to get a more determined set that has higher probability to interact. The result has highlighted a more potent set of proteins with higher chances of forming PPIs, representing the interactions among human and HIV-1.

Domain/motif interaction-based prediction

Here, the methodology for prediction of host–pathogen PPIs involves integration of known intra-species PPIs with protein domain profiles, and thereby predicting PPIs between a host and a pathogen [37]. For a set of intra-species PPIs, the functional domains are identified for each interacting protein. For each pair of functional domains, Bayesian statistics is used to compute the possibility of two proteins containing that pair of domain will interact. The method is shown in Fig. 5. It has been applied to *Homo sapiens-Plasmodium falciparum* host–pathogen system, and has successfully predicted 516 PPIs. Human proteins anticipated to interact with the same *Plasmodium* protein are close to each other in the human PPI network, and *Plasmodium* pairs predicted to interact with the same human protein are co-expressed in DNA micro-array datasets measured during various stages of the *Plasmodium* life cycle.

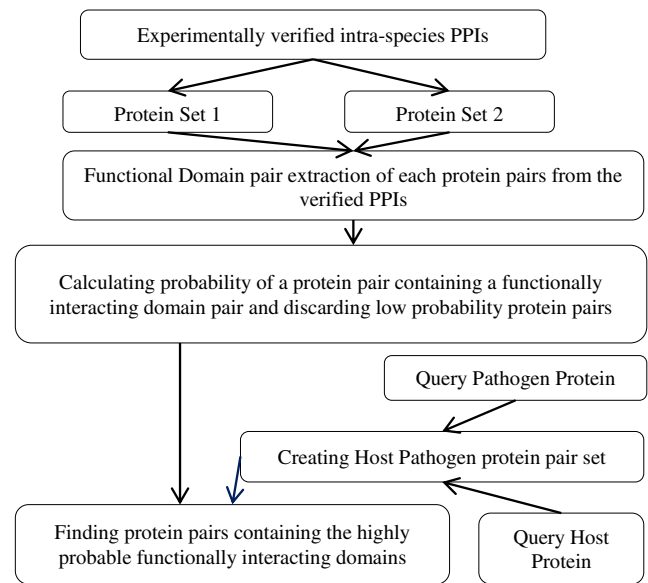


Fig. 5 Domain/motif-based prediction of host–pathogen interactions

Prediction of PPIs, based on motifs conserved in HIV-1, has been performed by Evans et al. [43] and Bertoletti et al. [8]. The similarity between the binding motifs shared by virus and host proteins plays an important part in the crosstalk between virus and host. Similarly, the study by Bertoletti et al. [8] has attempted to predict PPIs based on motifs conserved in HIV-1. It has also highlighted the role of chemokines as a factor for liver inflammation.

Machine learning-based predictions of host–pathogen interactions

Machine learning-based prediction methods are extensively used for detecting host–pathogen interactions, as shown in Table 1. This table lists a few machine learning methods used for the prediction of various aspects of host–pathogen interactions in different species. Moreover, the particular domain knowledge is also included in this table. The sub-area of research in some cases is referred to as “pathogen informatics”. Supervised learning has been used for the prediction of PPIs in the host–pathogen domain by Tastan et al. [123]. The work has considered 35 features, including tissue distribution, gene expression profile, gene ontology, graph properties of human interactome, sequence similarity, post-translational modification similarity to neighbor, and HIV-1 protein-type features among others. Then, the authors have selected the top three and top six features that are of maximum importance to classify the given data set into interacting and non-interacting classes. The Random Forest classifier has been used as a tool for supervised learning with these feature set for training and resulting in MAP (maximum a posteriori) of 23 %. From this computation,

Table 1 Summary of the machine learning-based tools used in the domain of host–pathogen interactions

Machine-learning method	Species	Reference	Domain
Random Forest classifier	<i>HIV1-Homo sapiens</i>	Tastan et al. [123]	PPI
Naïve Bayes classifier	Phylum <i>Chlamydiae</i> and genera <i>Escheria</i> , <i>Yersinia</i> and <i>Pseudomonas</i>	Arnold et al. [4]	T3SS
Support vector machine	<i>Ralstonia solanacearum</i>	Wang et al. [132]	T3SS
Multi-task classifier using support vector machine	<i>Yersinia pestis</i> , <i>Francisella tularensis</i> , <i>Salmonella</i> and <i>Bacillus anthracis</i>	Kshirsagar et al. [66]	PPI
Semi-supervised learning using multi-layer perceptron	<i>HIV1-Homo sapiens</i>	YanJun et al. [102]	PPI
Random Forest classifier	<i>Homo sapiens-Plasmodium falciparum</i>	Wuchty [138]	PPI
Group lasso with $L1/L2$ regularization	<i>Homo sapiens-Salmonella</i> , <i>Homo sapiens-Yersinia</i>	Kshirsagar et al. [65]	PPI
Support vector machine	None	Thieu et al. [126]	Data mining

it has been concluded that graph and neighbor similarity features contribute to a better classification.

Prediction of proteins secreted by Type III (T3) secretion system has been carried out by Arnold et al. [4]. The authors have examined the amino acid composition and the secondary structure of the N-terminal of 100 experimentally verified effector proteins, and used them for identification of T3 secretion signal. They have used Naive Bayes algorithm for classification. The training samples have been grouped depending on how similar they are, and this similarity has been measured by the Smith–Waterman local alignment algorithm. The input feature set has included frequencies of amino acid, amino acid properties, and short combinations of them. Finally, the feature-selection strategies have been applied to identify the most important feature to do away with computational complexity. In another attempt for prediction, the authors have used derived features from the secondary structure elements. They have used PSIPred software [82] to predict the structure. From the predicted structures, the features of the input vector have been formulated.

In another attempt to predict bacterial Type III secreted (T3S) effectors, a distinct N terminal position-specific amino acid composition feature has been found in more than 50 % of T3S proteins [132]. Bi-profile Bayes method has been used in this particular work for feature extraction. Then, the entire dataset along with the new feature has been analyzed with a new SVM-based classifier. The new classifier has classified T3S and non-T3S proteins successfully.

In order to establish a relation among a host and multiple pathogens, Kshirsagar et al. [66] have developed a method taking the similarity in infection initiated by four different pathogens in human host. The authors have used the machine-learning technique in the form of multi-task classification frameworks. The host–bacteria PPIs have been used as the input to the multi-task classifier, which has then classified the PPIs into interacting and non-interacting classes. Considering the biological hypothesis of similar pathogens targeting the same critical biological processes in a host, the classifier has minimized the empirical error on the training set and favored models that are biased towards the biological hypothesis. A bias term has been incorporated into the classifier in the form of a regularizer to overcome it.

A semi-supervised multi-task method has been used on *Homo sapiens*–HIV 1 dataset [102] to predict host–pathogen PPIs. The method has involved both supervised and semi-supervised learning. The supervised classifier has worked on labeled PPIs data. The semi-supervised classifier has shared network layers of the supervised classifier and got trained with partially labeled PPIs. This entire framework has been used to improve the recognition of interacting pairs. The supervised classifier has done multi-tasking with

a semi-supervised classifier so that weak positive labels could ameliorate the supervised classification.

For prediction of PPIs between *Homo sapiens* and *Plasmodium falciparum*, a Random Forest classifier has assessed a set of PPIs and then filtered the result according to expression and molecular characteristics, leading to a subset of proteins, which indeed interact among themselves [138]. It has been observed here that the separate sets and a combined set of predicted and experimentally verified interactions have shared similar characteristics. In another investigation, Kshirsagar et al. [65] have tried to improve the supervised learning-based prediction of PPIs between *Salmonella*-human and *Yersinia*-human. This has been done by replacing the missing values of the dataset by the values generated by cross species information along with group lasso technique with regularization (obtained 77.6 % precision). In order to impute values, localized nearest-neighbor approach (which uses sequence similarity) has been used as the basis to compute locality.

Data mining also forms an integral part of machine learning. Retrieved data about host–pathogen interactions in a few cases reflects information in two different ways, i.e., feature-based (SVM) [126] and language-based [19]. The investigation by Chaussabel et al. [19] used the hierarchical clustering algorithm by taking the literature available to identify a functionally and transcriptionally homologous pair of genes as input. Removal of noise from the

PPI databases was done by removing PPIs that have less probability of taking place. Each such PPI has then been given a score. Then, these PPIs have been hierarchically clustered to obtain the PPIs likeliness of occurrence. In this way, it has been found that out of 12,122 binary PPIs obtained from BioGRID, 7504 PPIs are less likely to take place.

Online repositories for host–pathogen interactions

Host–pathogen interactions data can be obtained from several databases and repositories. We have summarized some of these repositories in Table 2. Some of these databases are referred to purely for their data content, i.e., genome, proteome, and metabolic pathway data [133], virus–virus, host–virus, and host–host interaction networks [95], PPIs of hosts and pathogens [69], literature-based viral–human protein interactions [18], experimentally verified pathogenic, virulence and effector genes of fungal pathogens [136], human signaling and regulatory pathways [113], information on specific biodefense and public health pathogens [120], 3D viral proteins [116], information on invertebrate vectors of human pathogens [71], and a collection of genus-specific databases [6] among others. Some of these databases even have integrated in-house tools, i.e., BLAST interface [35] and browser [142] for host–pathogen interactions data

Table 2 List of online repositories storing data related to host–pathogen interactions

No.	Name	URL
1	PATRIC [133]	http://patricbrc.org/portal/portal/patric/Home
2	PIG [35]	http://patricbrc.org/portal/portal/patric/HPITool
3	VirHostNet [95]	http://virhostnet.prabi.fr/
5	HPIDB [69]	http://agbase.msstate.edu/hpi/main.html
6	GPS-Prot [44]	http://gpsprot.org/
7	VirusMint [18]	http://mint.bio.uniroma2.it/virusmint/Welcome.do
8	PHIDIAS [139]	http://www.phidias.us/introduction.php
9	MvirDB [142]	http://mvirdb.llnl.gov/
10	PHI-base [136, 137]	http://www.phi-base.org/
11	PID [113]	http://pid.nci.nih.gov/
12	BioHealthBase [120]	http://www.biohealthbase.org/
13	VPDB [116]	http://www.vpdb.bicpu.edu.in/
14	VectorBase [71]	https://www.vectorbase.org/
15	EuPathDB [6]	http://eupathdb.org/eupathdb/
16	PHISTO [125]	http://www.phisto.org/
17	ViPR [99]	http://www.viprbc.org/brc/home.spg?decorator=vipr
18	EDWIP [97]	http://cricket.inhs.uiuc.edu/edwipweb/edwipabout.htm
19	HoPaCI-db [9]	http://mips.helmholtz-muenchen.de/HoPaCI
20	VFDB [21]	http://www.mgc.ac.cn/VFs/main.htm
21	AquaPathogen X [41]	http://pubs.usgs.gov/fs/2012/3015/

analysis. Moreover, we have described some tools [44] used in analysis and visualization of these kinds of data.

The PATHosystems Resource Integration Center (PATRIC) [133] includes a relational database, analytical pipelines, and a website that supports querying, browsing, data visualization, and allowing the download of raw and curated data in standard formats. Currently, the database houses complete sequences for viral and bacterial genomes, hence providing an all-inclusive bioinformatics resource for pathogens.

The Pathway Interaction Gateway (PIG) provides a text-based search and a BLAST interface for searching the host–pathogen PPIs. Each entry in PIG incorporates information on the functional annotations and the domains present in the interacting proteins [35].

VirHostNet (Virus–Host Network) [51, 95] is a public knowledge base specialized in the management and analysis of integrated virus–virus, host–host, and virus–host interaction networks coupled with their functional annotations. VirHostNet contains data of virus–host and virus–virus interactions constituting more than 180 distinct viral species. The VirHostNet Web interface provides suitable tools which allow effective query and visualization of infected cellular networks.

HPIDB (Host–Pathogen Interaction Database) [69] basically contains experimentally confirmed and predicted PPIs of hosts and pathogens.

GPS-Prot [44] is a software tool that permits users to easily create an all-inclusive and integrated HIV–host networks. Its web-based format, which requires no software installation or data downloads, gives it an extra edge over other visualization tools. GPS-Prot enables users to quickly generate networks that amalgamate both genetic and protein–protein interactions between HIV and its human host into a single representation.

VirusMint [18] contains protein interactions between viral (papilloma viruses, HIV-1, Epstein–Barr, hepatitis B, hepatitis C, herpes, and Simian virus 40) and human proteins reported in the literature. VirusMINT presently stores interactions constituting more than 490 unique viral proteins from more than 110 different viral strains.

PHIDIAS (a Pathogen Host Interaction Data Integration and Analysis System) [139] is a database and analysis system to curate, analyze, and address different scientific issues in the areas of host–pathogen interactions (PHI, or called host–pathogen interactions or HPI).

MvirDB [142] integrates DNA and protein sequence information from multiple databases. Entries in MvirDB are hyper-linked back to their original sources. A blast tool enables the user to blast against all DNA or protein sequences in MvirDB, and a browser tool enables the user to explore the database to retrieve virulence factor

descriptions, sequences, and classifications, and to download sequences of interest.

PHI-base [136], a web-accessible database currently catalogs experimentally verified virulence and effector genes from fungal and oomycete pathogens. These pathogens interact with animal, plant, and fungi as hosts.

PID [113] is a freely available collection of curated and peer-reviewed pathways composed of human molecular signaling and regulatory events and key cellular processes. PID offers a range of search features to facilitate pathway exploration.

BioHealthBase [120] is a public bioinformatics database and analysis resource for study of specific biodefense and public health pathogens like *Francisella tularensis*, *Mycobacterium tuberculosis*, *Influenza virus*, *Microsporidia* species and ricin toxin. It serves as a substantial integrated repository of data imported from public databases and data derived from various computational algorithms and information curated from the scientific literature. Its 3D visualization capacity allows researchers to view proteins with their key structural and functional features highlighted.

VPDB (Viral Protein Structural Database) [116] is an interactive database for three-dimensional viral proteins. It provides an all-inclusive resource, with an emphasis on the description of derived data from structural biology. At present, VPDB includes viral protein structures from more than 277 viruses with more than 465 virus strains.

VectorBase [71, 72, 85] is a web-accessible data repository storing information about invertebrate vectors of human pathogens. It annotates and maintains vector genomes, providing an integrated resource for the research community. It hosts data related to nine genomes, i.e., mosquitoes (3 *Anopheles gambiae* genome), *Aedes aegypti* and *Culex quinquefasciatus*, body louse (*Pediculus humanus*), tick (*Ixodes scapularis*), tsetse fly (*Glossina morsitans*) and kissing bugs (*Rhodnius prolixus*). The data spans across genomic features, expression data, population genetics, and ontologies.

EuPathDB [5, 6] is an integrated database covering the eukaryotic pathogens of the genera *Giardia*, *Cryptosporidium*, *Neospora*, *Leishmania*, *Toxoplasma*, *Plasmodium*, *Trypanosoma* and *Trichomonas*. These groups are supported by a taxon-specific database built upon the same infrastructure. EuPathDB portal provides an entry point to all these resources, and the opportunity to leverage orthology for searches across genera.

Similarly, a number of other databases, like PHISTO [125], ViPR [99], HoPaCI-DB [9], VFDB [21, 140] [20], EDWIP [97], AquaPathogen X [41], are available, which help in the host–pathogen interactions domain research.

Discussions and future scopes

In this section, we discuss multiple facets of host–pathogen interactions research, the shortcoming of the previously defined methodologies as discussed in Sections “[Classification of host–pathogen interactions](#)” and “[Methods for prediction of host–pathogen interactions](#)”, and the future scopes associated with the aforesaid methodologies, which takes both the host and pathogen points of view into account. We discuss the ways in which a pathogen can attack its host, the proteins emitted by a pathogen responsible for perturbing normal functionality of host, the genes responsible for such proteins, silencing and hijacking gene mechanism of pathogens, inhibiting the functions of macrophages, along with genes and proteins needed for their survival inside a host. From the host’s point of view, we also discuss the factors of pathogen that activates immune response. Salient features of the discussion are given in Table 3.

The genes of multiple strains of an organism have been studied in several investigations [58, 81, 96] to understand the infection mechanism of these strains on the host and to locate the difference between them. In order to survive in a host, a pathogen can either perform hijacking [105] or it can use the existing environment to survive [12]. The effect of the genes in different strains of a pathogen has been studied. There is still uncertainty in the generalization/specialization of interactions in different strains of pathogens. A study has suggested that different strains of the same pathogen have different methods of invasion [81]. On the contrary, a counter example has also been provided in [26], which indicates that two strains of *Mycobacterium* have homologous genes required for Snm.

Influenza, DENV-2, and HIV have been in the limelight for identification of the host factors. Other pathogens too need to be taken into account. Inhibition of macrophage is a prospective aspect of research in bioinformatics. The inhibition mechanism needs to be studied in more pathogens apart from the mostly studied ones to find similarity between the inhibition mechanisms among these organisms.

Machine learning-based prediction methods have been applied mainly to PPIs. However, protein–ligand interactions, and hence prediction of pathways (excluding signal transduction pathways) via machine learning methods, have not been attempted much. Different pathogens become drug resistant and form new pathways, and these newly formed pathways can perturb the present host pathways in an unknown way. Similarly, machine learning algorithms in the field of pathway predictions are needed, which would mainly consider protein–ligand binding. Along with reaction dynamics are needed to be known too, as pathways are nothing but chain of reactions. Prediction of Type III secreted bacterial proteins by machine learning techniques is also a

Table 3 Summary of host protection and pathogen-attacking mechanisms

Host protection mechanism	Pathogen-attacking mechanism
Protein–protein interactions (GBP galactose-binding protein)	Protein–protein interactions (target hub protein)
shRNAs (pathogen gene knockdown)	microRNAs (protection against cellular micro-viral response, gene silencing)
Autophagy	MAM (multivariate adhesion molecule, high binding affinity with host during infection)
siRNAs (inhibit HIV-1 replication)	Pneumolysin (virulence factor)
Macrophages	Inhibition of macrophage
Restricting supply of iron	Glial cells of host (convert it into progenitor cells then survive in the host)
None	Motif mimicry (utilized by pathogens to rewire host pathways by co-opting SLiM-mediated protein interactions)
None	Biofilm formation
None	Hijacking (implant own sequence in normal sequence of host)

challenging task. However, a major drawback in the area of prediction of host–pathogen PPIs, are the unavailability of data sets for different pathogens. Moreover, there is always this lurking issue of biological validation of the predicted PPIs.

Some of the organisms studied for the exploration of host–pathogen PPIs are *Homo sapiens-Plasmodium falciparum* [37, 64, 74, 138], *Homo sapiens-Dengue virus* [34], *Homo sapiens-HIV 1* [8, 33, 43]. However, there are many more host–pathogen pairs waiting in the line for these kinds of studies. In addition, homology-based approaches have their own inherent weaknesses. In a real scenario, two proteins in a predicted PPI may actually have little opportunity to be present close enough to interact with each other. Therefore, host–pathogen PPIs predicted entirely on the basis of homology, without considering other biological characteristics of the proteins involved, may not be reliable. Additional information must be used to increase the accuracy of the prediction and make the predictions biologically sound. Keeping this in mind, the study by Wuchty [138] has filtered the predicted PPIs based on homology using gene expression and molecular characteristics. It has led to the formation of a concrete set of PPIs closer to the biological scenario. The prediction of PPIs by comparative modeling [28] has very stringent filters leading to the formation of a smaller and robust set of PPIs.

Supervised, unsupervised and semi supervised learning have been mostly used for prediction of host–pathogen PPIs. The organisms for which these predictions have been made are mainly *Homo sapiens-HIV1* [102, 123], *Homo sapiens-Plasmodium falciparum* [138], and *Homo sapiens-Saccharomyces cerevisiae* [25]. Both Tastan et al. and Yanjun et al. [102, 123] have applied their respective algorithms on the same dataset, which basically restricts the contribution of the articles. The performance of the Random Forest-based classifier is negligibly better than the Multi-Layer Perceptron classifier [102]. Some research articles have selected the top six and top three features among 35 features to predict whether a protein is interacting or not [123]. This is not a novel way of prediction since the interaction between proteins depends on all of its features even if by a negligible amount, which should not be ignored.

A flaw is often noticed in the choice of a dataset. In a semi-supervised based learning approach to identify PPIs [102], the negative dataset is way more extensive than the positive one. The negative (non-interacting) data set has approximately 16,000 pairs of proteins while the experimentally verified positive (interacting) dataset has only 158 pairs of proteins. Training with such a dataset might lead to a biased classifier, and the classifier would be inclined to predict most test pairs as non-interacting. Moreover, the logic used behind selecting a non-interacting dataset is based on

a random list of pairs of proteins that do not fall into the positive set. It is always a risk, since there is no experimental evidence that the selected negative pairs will not interact at all. There may be several interacting pairs present among the negative set. Another study has been done for predicting proteins secreted by a Type III secretion system based only on structural and compositional aspects of the proteins [4]. These studies should include other factors like expression and molecular characteristics.

One notable thing is that a few attempts have been made on metabolic pathways. For host–pathogen interactions, most of the work has been done with signal transduction pathways. If enzyme(s) from a pathogen is introduced into a host, they get involved with more than one host pathway. There is no tool available that would take a list of protein (enzyme) names and provide the pathway (just one pathway based on these enzymes) based only on those enzymes (at least 90 %). Moreover, a pathogen can be associated with more than one disease. Such diseases, for which a pathogen is responsible, need to be looked into. The scenario becomes more complex when a host suffers from two or more diseases simultaneously, which implies the presence of multiple pathogens responsible for multiple diseases in a host in real time. Such real-time simulation studies are hardly done.

An important aspect that needs to be considered is that some pathogenic proteins prevent the working of macrophage. This is a serious problem in host–pathogen domain. Drugs are needed that would facilitate the working behavior of a macrophage. Drugs are also needed for the prevention of formation of intracytoplasmic vesicle that HIV-1 uses [22] to prevent identification by macrophages. Formation of biofilms [89, 121] is another domain that needs to be looked into. Breaking the biofilm formed by pathogens is indeed recommended to avoid the spread of infection. More attention is needed in this domain, given the rate at which new infectious pathogens are emerging along with their variety of degree of infection.

Hardly any research has been done based on the automated image processing-based techniques available for predicting host–pathogen interactions. A study by Mech et al. [83] has come up with a technique of a more robust analysis of microscopy images of macrophages that is made to coexist with different *A. fumigatus* strains. Usually, the images are manually analyzed, which is both time-consuming and error prone. The authors used the feature set which includes size, shape, number of cells, and cell–cell contacts. By analyzing the images, it has been found that different mutants of *A. fumigatus* have an impact on the ability of the macrophages to adhere and phagocytose the conidia. It has been observed that the rate of phagocytosis is higher in pksP mutants of *A. fumigates*, while it is not the same case in the other strains.

Conclusions

In this review, we have covered various aspects of host–pathogen interactions. Interaction of a pathogen with its host(s) is always a unique mechanism. Each one of the pathogenic species has specific mechanism(s) to interact with their host. The different mechanisms of a number of species have been included in this review along with the similarities and similar factors in the attacking mechanism(s) of pathogens. The review has introduced a brief history and introduction of the host–pathogen interactions research field followed by classification of host–pathogen interactions based on gene(s), protein(s), host-factor(s), involved pathway(s), and inhibition mechanism of macrophage(s). It has listed prediction methods used in the host–pathogen interactions domain based on biological reasoning (homology, structure, and motif interaction), machine learning (unsupervised, semi-supervised, and supervised) and sometimes both methods. Various data sources used for research in this domain have also been listed. The review concludes with a general discussion of the topic and future scopes followed by a conclusion. The field of host–pathogen interactions is emerging as a crucial area of infectious disease research in the post-genomic era. It is a budding research field where new discoveries are getting announced almost each day around the globe. The discovery of dynamics of the host–pathogen interactions will aptly facilitate further development in the field of discovering new drugs and new therapies for different diseases.

Acknowledgments LN acknowledges University Grants Commission, India, for a UGC Post-Doctoral Fellowship (No. F.15-1/2013-14/PDFWM-2013-14-GE-ORI-19068(SA-II)).

Compliance with Ethical Standards

Funding This work is not funded by any funding agency.

Conflict of interest The authors declare that they have no conflicts of interest.

Ethical approval This is a review article. Ethical approval is not required as no human subject is involved.

Informed consent Informed consent is not needed for this work as it is a review article and no human subject is involved.

Author's contributions RS conceptualized the whole review. She prepared the initial manuscript. LN and RKD gave theoretical input and modified it. RS, LN, and RKD read and corrected the final manuscript.

References

- Albersheim P, Anderson AJ (1971) Proteins from plant cell walls inhibit polygalacturonases secreted by plant pathogens. *Proc Nat Acad Sci* 68(8):1815–1819
- Albersheim P, Valent BS (1974) Host–pathogen interactions VII. Plant pathogens secrete proteins which inhibit enzymes of the host capable of attacking the pathogen. *Plant Physiol* 53(5):684–687
- Alderwick LJ, Dover LG, Seidel M, Gande R, Sahn H, Eggeling L, Besra GS (2006) Arabinan-deficient mutants of *Corynebacterium glutamicum* and the consequent flux in decaprenylmonophosphoryl-D-arabinose metabolism. *Glycobiology* 16(11):1073–1081
- Arnold R, Brandmaier S, Kleine F, Tischler P, Heinz E, Behrens S, Niinikoski A, Mewes HW, Horn M, Rattei T (2009) Sequence-based prediction of type III secreted proteins. *PLoS Pathogen* 5(4):e1000376
- Aurrecochea C, Barreto A, Brestelli J, Brunk BP, Cade S, Doherty R, Fischer S, Gajria B, Gao X, Gingle A et al (2013) EuPathDB: the eukaryotic pathogen database. *Nucleic Acids Res* 41(D1):D684–D691
- Aurrecochea C, Brestelli J, Brunk BP, Fischer S, Gajria B, Gao X, Gingle A, Grant G, Harb OS, Heiges M et al (2010) EuPathDB: a portal to eukaryotic pathogen databases. *Nucleic Acids Res* 38(suppl 1):D415–D419
- Balakrishnan S, Tastan O, Carbonell J, Klein-Seetharaman J (2009) Alternative paths in HIV-1 targeted human signal transduction pathways. *BMC Genom* 10(3):1
- Bertoletti A, Maini MK, Ferrari C (2010) The host–pathogen interaction during HBV infection: immunological controversies. *Antiv Therapy* 15(3):15
- Bleves S, Dunger I, Walter MC, Frangoulidis D, Kastenmüller G, Voulhoux R, Ruepp A (2014) HoPaCI-DB: host-Pseudomonas and Coxiella interaction database. *Nucleic Acids Res* 42(D1):D671–D676
- Bock JR, Gough DA (2001) Predicting protein–protein interactions from primary structure. *Bioinformatics* 17(5):455–460
- Botella H, Stadthagen G, Lugo-Villarino G, de Chastellier C, Neyrolles O (2012) Metallobiology of host–pathogen interactions: an intoxicating new insight. *Trends Microbiol* 20(3):106–112
- Brass AL, Dykxhoorn DM, Benita Y, Yan N, Engelman A, Xavier RJ, Lieberman J, Elledge SJ (2008) Identification of host proteins required for HIV infection through a functional genomic screen. *Science* 319(5865):921–926
- Breitenbach JM, Hausinger RP (1988) *Proteus mirabilis* urease. Partial purification and inhibition by boric acid and boronic acids. *Biochem J* 250(3):917–920
- Bumann D (2015) Heterogeneous host–pathogen encounters: act locally, think globally. *Cell Host Microbe* 17(1):13–19
- Burts ML, Williams WA, DeBord K, Missiakas DM (2005) EsxA and EsxB are secreted by an ESAT-6-like system that is required for the pathogenesis of *Staphylococcus aureus* infections. *Proc Nat Acad Sci USA* 102(4):1169–1174
- Calderwood MA, Venkatesan K, Xing L, Chase MR, Vazquez A, Holthaus AM, Ewence AE, Li N, Hirozane-Kishikawa T, Hill DE et al (2007) Epstein–Barr virus and virus human protein interaction maps. *Proc Nat Acad Sci* 104(18):7606–7611
- Casadevall A, Pirofski La (1999) Host–pathogen interactions: redefining the basic concepts of virulence and pathogenicity. *Infec Immun* 67(8):3703–3713
- Chatr-aryamontri A, Ceol A, Peluso D, Nardoza A, Panni S, Sacco F, Tinti M, Smolyar A, Castagnoli L, Vidal M et al (2009)

- VirusMINT: a viral protein interaction database. *Nucleic Acids Res* 37(suppl 1):D669–D673
19. Chaussabel D, Semnani RT, McDowell MA, Sacks D, Sher A, Nutman TB (2003) Unique gene expression profiles of human macrophages and dendritic cells to phylogenetically distinct parasites. *Blood* 102(2):672–681
 20. Chen L, Xiong Z, Sun L, Yang J, Jin Q (2011) VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. *Nucleic Acids Res*:gkr989
 21. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q (2005) VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res* 33(suppl 1):D325–D328
 22. Chertova E, Chertov O, Coren LV, Roser JD, Trubey CM, Bess JW, Sowder RC, Barsov E, Hood BL, Fisher RJ et al (2006) Proteomic and biochemical analysis of purified human immunodeficiency virus type 1 produced from infected monocyte-derived macrophages. *J Virol* 80(18):9039–9052
 23. Clemens DL, Lee BY, Horwitz MA (1995) Purification, characterization, and genetic analysis of *Mycobacterium tuberculosis* urease, a potentially critical determinant of host–pathogen interaction. *J Bacteriol* 177(19):5644–5652
 24. Cole S, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon S, Eiglmeier K, Gas S, Barry Cr et al (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393(6685):537–544
 25. Collins SR, Kemmeren P, Zhao XC, Greenblatt JF, Spencer F, Holstege FC, Weissman JS, Krogan NJ (2007) Toward a comprehensive atlas of the physical interactome of *Saccharomyces cerevisiae*. *Molec Cell Proteom* 6(3):439–450
 26. Converse SE, Cox JS (2005) A protein secretion pathway critical for *Mycobacterium tuberculosis* virulence is conserved and functional in *Mycobacterium smegmatis*. *J Bacteriol* 187(4):1238–1245
 27. Costa TR, Felisberto-Rodrigues C, Meir A, Prevost MS, Redzej A, Trokter M, Waksman G (2015) Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nat Rev Microbiol* 13(6):343–359
 28. Davis FP, Barkan DT, Eswar N, McKerrow JH, Sali A (2007) Host–pathogen protein interactions predicted by comparative modeling. *Protein Sci* 16(12):2585–2596
 29. Delogu G, Brennan MJ (2001) Comparative immune response to PE and PE.PGRS antigens of *Mycobacterium tuberculosis*. *Infect Immun* 69(9):5606–5611
 30. d’Enfert C, Ryter A, Pugsley A (1987) Cloning and expression in *Escherichia coli* of the *Klebsiella pneumoniae* genes for production, surface localization and secretion of the lipoprotein pullulanase. *EMBO J* 6(11):3531
 31. Dickson J, Syamananda R, Flangas A (1959) The genetic approach to the physiology of parasitism of the corn rust pathogens. *Amer J Botany*:614–620
 32. Doherty CP (2007) Host–pathogen interactions: the role of iron. *J Nutr* 137(5):1341–1344
 33. Doolittle JM, Gomez SM (2010) Structural similarity-based predictions of protein interactions between HIV-1 and *Homo sapiens*. *Virol J* 7(1):82
 34. Doolittle JM, Gomez SM (2011) Mapping protein interactions between Dengue virus and its human and insect hosts. *PLoS Neglect Tropical Dis* 5(2):e954
 35. Driscoll T, Dyer MD, Murali T, Sobral BW (2009) PIG-the pathogen interaction gateway. *Nucleic Acids Res* 37(suppl 1):D647–D650
 36. Durmuş S, Çakır T, Özgür A, Guthke R (2015) A review on computational systems biology of pathogen–host interactions. *Frontiers Microbiol* 6
 37. Dyer MD, Murali T, Sobral BW (2007) Computational prediction of host–pathogen protein–protein interactions. *Bioinformatics* 23(13):i159–i166
 38. Dyer MD, Murali T, Sobral BW (2008) The landscape of human proteins interacting with viruses and other pathogens. *PLoS Pathogen* 4(2):e32
 39. Dyer MD, Neff C, Dufford M, Rivera CG, Shattuck D, Bassaganya-Riera J, Murali T, Sobral BW (2010) The human-bacterial pathogen protein interaction networks of *Bacillus anthracis*, *Francisella tularensis*, and *Yersinia pestis*. *PLoS One* 5(8):e12,089
 40. Edwards H, Allen P et al (1970) A fine-structure study of the primary infection process during infection of barley by *Erysiphe graminis* f. sp. hordei. *Phytopathology* 60(10):1504–1509
 41. Emmenegger E, Kentop E, Thompson T, Pittam S, Ryan A, Keon D, Carlino J, Ranson J, Life R, Troyer R et al (2011) Development of an aquatic pathogen database (Aquapathogen X) and its utilization in tracking emerging fish virus pathogens in North America. *J Fish Dis* 34(8):579–587
 42. English PD, Albersheim P (1969) Host–pathogen interactions: I. A correlation between α -galactosidase production and virulence. *Plant Physiol* 44(2):217–224
 43. Evans P, Dampier W, Ungar L, Tozeren A (2009) Prediction of HIV-1 virus-host protein interactions using virus and host sequence motifs. *BMC Med Genom* 2(1):1
 44. Fahey ME, Bennett MJ, Mahon C, Jäger S, Pache L, Kumar D, Shapiro A, Rao K, Chanda SK, Craik CS et al (2011) GPS-Prot: a web-based visualization platform for integrating host–pathogen interaction data. *BMC Bioinform* 12(1):298
 45. Fenhalls G, Stevens L, Moses L, Bezuidenhout J, Betts JC, van Helden P, Lukey PT, Duncan K (2002) In situ detection of *Mycobacterium tuberculosis* transcripts in human lung granulomas reveals differential gene expression in necrotic lesions. *Infect Immun* 70(11):6330–6338
 46. Fisher ML, Anderson AJ, Albersheim P (1973) Host–pathogen interactions VI. A single plant protein efficiently inhibits endopolygalacturonases secreted by *Colletotrichum lindemuthianum* and *Aspergillus niger*. *Plant Physiol* 51(3):489–491
 47. Galan JE, Curtiss R (1989) Cloning and molecular characterization of genes whose products allow *Salmonella typhimurium* to penetrate tissue culture cells. *Proc Nat Acad Sci* 86(16):6383–6387
 48. Ghosh Z, Mallick B, Chakrabarti J (2009) Cellular versus viral microRNAs in host–virus interaction. *Nucleic Acids Res* 37(4):1035–1048
 49. Gómez-Díaz E, Jordà M, Peinado MA, Rivero A (2012) Epigenetics of host–pathogen interactions: the road ahead and the road behind. *PLoS Pathogen* 8(11):e1003,007
 50. Guérin I, de Chastellier C (2000) Pathogenic mycobacteria disrupt the macrophage actin filament network. *Infect Immun* 68(5):2655–2662
 51. Guirimand T, Delmotte S, Navratil V (2015) VirHostNet 2.0: surfing on the web of virus/host molecular interactions data. *Nucleic Acids Res* 43(D1):D583–D587
 52. Gutierrez MG, Master SS, Singh SB, Taylor GA, Colombo MI, Deretic V (2004) Autophagy is a defense mechanism inhibiting BCG and *Mycobacterium tuberculosis* survival in infected macrophages. *Cell* 119(6):753–766
 53. Meijer HA, Spaik PH (2011) Host–pathogen interactions made transparent with the zebrafish model. *Current Drug Targets* 12(7):1000–1017
 54. Hatzios SK, Abel S, Martell J, Hubbard T, Sasabe J, Munera D, Clark L, Bachovchin DA, Qadri F, Ryan ET et al (2016) Chemoproteomic profiling of host and pathogen enzymes active in cholera. *Nat Chem Biol* 12(4):268–274

55. Hess S, Rambukkana A (2015) Bacterial-induced cell reprogramming to stem cell-like cells: new premise in host–pathogen interactions. *Current Opinion Microbiol* 23:179–188
56. Hess W (1969) Ultrastructure of onion roots infected with *Pyrenochaeta terrestris*, a fungus parasite. *Amer J Botany*:832–845
57. Irazoqui JE, Ng A, Xavier RJ, Ausubel FM (2008) Role for β -catenin and HOX transcription factors in *Caenorhabditis elegans* and mammalian host epithelial–pathogen interactions. *Proc Nat Acad Sci* 105(45):17,469–17,474
58. de Jong HK, Parry CM, van der Poll T, Wiersinga WJ (2012) Host–pathogen interaction in invasive salmonellosis. *PLOS Pathogen* 8(10):e1002,933
59. Kearney B, Ronald PC, Dahlbeck D, Staskawicz BJ (1988) Molecular basis for evasion of plant host defence in bacterial spot disease of pepper. *Nature* 332(6164):541–543
60. Kierszenbaum F, Wirth JJ, McCANN PP, Sjoerdsma A (1987) Impairment of macrophage function by inhibitors of ornithine decarboxylase activity. *Infect Immun* 55(10):2461–2464
61. König R, Stertz S, Zhou Y, Inoue A, Hoffmann HH, Bhattacharyya S, Alamares JG, Tscherne DM, Ortigoza MB, Liang Y et al (2010) Human host factors required for influenza virus replication. *Nature* 463(7282):813–817
62. König R, Zhou Y, Elleder D, Diamond TL, Bonamy GM, Irelan JT, Chiang Cy, Tu BP, De Jesus PD, Lilley CE et al (2008) Global analysis of host–pathogen interactions that regulate early-stage HIV-1 replication. *Cell* 135(1):49–60
63. Krachler AM, Ham H, Orth K (2011) Outer membrane adhesion factor multivalent adhesion molecule 7 initiates host cell binding during infection by Gram-negative pathogens. *Proc Nat Acad Sci* 108(28):11,614–11,619
64. Krishnadev O, Srinivasan N (2008) A data integration approach to predict host–pathogen protein–protein interactions: application to recognize protein interactions between human and a malarial parasite. *Silico Biol* 8(3, 4):235–250
65. Kshirsagar M, Carbonell J, Klein-Seetharaman J (2012) Techniques to cope with missing data in host–pathogen protein interaction prediction. *Bioinformatics* 28(18):i466–i472
66. Kshirsagar M, Carbonell J, Klein-Seetharaman J (2013) Multi-task learning for host–pathogen protein interactions. *Bioinformatics* 29(13):i217–i226
67. Kuehn MJ, Kesty NC (2005) Bacterial outer membrane vesicles and the host–pathogen interaction. *Genes Develop* 19(22):2645–2655
68. Kulldau GA, De Vos G, Owen J, McCaffrey G, Zambryski P (1990) The virB operon of *Agrobacterium tumefaciens* pTiC58 encodes 11 open reading frames. *Molecular Gen Genet* MGG 221(2):256–266
69. Kumar R, Nanduri B (2010) HPIDB—a unified resource for host–pathogen interactions. *BMC Bioinform* 11(6):1
70. Kurz CL, Ewbank JJ (2000) *Caenorhabditis elegans* for the study of host–pathogen interactions. *Trends Microbiol* 8(3):142–144
71. Lawson D, Arensburger P, Atkinson P, Besansky NJ, Bruggner RV, Butler R, Campbell KS, Christophides GK, Christley S, Dialynas E et al (2007) VectorBase: a home for invertebrate vectors of human pathogens. *Nucleic Acids Resh* 35(suppl 1):D503–D505
72. Lawson D, Arensburger P, Atkinson P, Besansky NJ, Bruggner RV, Butler R, Campbell KS, Christophides GK, Christley S, Dialynas E et al (2009) VectorBase: a data resource for invertebrate vector genomics. *Nucleic Acids Res* 37(suppl 1):D583–D587
73. Lecuit M, Dramsi S, Gottardi C, Fedor-Chaiken M, Gumbiner B, Cossart P (1999) A single amino acid in E-cadherin responsible for host specificity towards the human pathogen *Listeria monocytogenes*. *EMBO J* 18(14):3956–3963
74. Lee SA, Chan Ch, Tsai CH, Lai JM, Wang FS, Kao CY, Huang CYF (2008) Ortholog-based protein–protein interaction prediction and its application to inter-species interactions. *BMC Bioinform* 9(12):1
75. Lewthwaite JC, Coates AR, Tormay P, Singh M, Mascagni P, Poole S, Roberts M, Sharp L, Henderson B (2001) *Mycobacterium tuberculosis* chaperonin 60.1 is a more potent cytokine stimulator than chaperonin 60.2 (Hsp 65) and contains a CD14-binding domain. *Infect Immun* 69(12):7349–7355
76. Low DHP, Frecer V, Le Saux A, Srinivasan GA, Ho B, Chen J, Ding JL (2010) Molecular interfaces of the galactose-binding protein Tectonin domains in host–pathogen interaction. *J Biol Chem* 285(13):9898–9907
77. Lui YLE, Tan TL, Timms P, Hafner LM, Tan KH, Tan EL (2014) Elucidating the host–pathogen interaction between human col-orectal cells and invading Enterovirus 71 using transcriptomics profiling. *FEBS Open Bio* 4(1):426–431
78. Marriott HM, Mitchell TJ, Dockrell DH (2008) Pneumolysin: a double-edged sword during the host–pathogen interaction. *Current Molec Med* 8(6):497–509
79. Matthews LR, Vaglio P, Reboul J, Ge H, Davis BP, Garrels J, Vincent S, Vidal M (2001) Identification of potential interaction networks using sequence-based searches for conserved protein–protein interactions or “interologs”. *Genome Res* 11(12):2120–2126
80. Mattoo S, Lee YM, Dixon JE (2007) Interactions of bacterial effector proteins with host proteins. *Current Opinion Immunol* 19(4):392–401
81. McCarthy AJ, Lindsay JA (2010) Genetic variation in *Staphylococcus aureus* surface and immune evasion genes is lineage associated: implications for vaccine design and host–pathogen interactions. *BMC Microbiol* 10(1):1
82. McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. *Bioinformatics* 16(4):404–405
83. Mech F, Thywißen A, Guthke R, Brakhage AA, Figue MT (2011) Automated image analysis of the host–pathogen interaction between phagocytes and *Aspergillus fumigatus*. *PLoS One* 6(5):e19,591
84. Mege JL (2016) Dendritic cell subtypes: a new way to study host–pathogen interaction. *Virulence* 7(1):5–6
85. Megy K, Emrich SJ, Lawson D, Campbell DM, Dialynas E, Hughes DS, Koscielny G, Louis C, MacCallum RM, Redmond SN et al (2012) VectorBase: improvements to a bioinformatics resource for invertebrate vector genomics. *Nucleic Acids Res* 40(D1):D729–D734
86. Mishra AK, Driessen NN, Appelmek BJ, Besra GS (2011) Lipoarabinomannan and related glycoconjugates: structure, biogenesis and role in *Mycobacterium tuberculosis* physiology and host–pathogen interaction. *FEMS Microbiol Rev* 35(6):1126–1157
87. Molle V, Kremer L (2010) Division and cell envelope regulation by Ser/Thr phosphorylation: *Mycobacterium* shows the way. *Molecul Microbiol* 75(5):1064–1077
88. Mougous JD, Cuff ME, Raunser S, Shen A, Zhou M, Gifford CA, Goodman AL, Joachimiak G, Ordoñez CL, Lory S et al (2006) A virulence locus of *Pseudomonas aeruginosa* encodes a protein secretion apparatus. *Science* 312(5779):1526–1530
89. Muir RE, Tan MW (2008) Virulence of *Leucobacter chromiireducens* subsp. solipictus to *Caenorhabditis elegans*: characterization of a novel host–pathogen interaction. *Appl Environ Microbiol* 74(13):4185–4198
90. Murray PJ, Young RA (1992) Stress and immunological recognition in host–pathogen interactions. *J Bacteriol* 174(13):4193

91. Mylonakis E, Aballay A (2005) Worms and flies as genetically tractable animal models to study host–pathogen interactions. *Infect Immun* 73(7):3833–3841
92. Naglik J, Albrecht A, Bader O, Hube B (2004) *Candida albicans* proteinases and host/pathogen interactions. *Cell Microbiol* 6(10):915–926
93. Nairz M, Schroll A, Sonnweber T, Weiss G (2010) The struggle for iron—a metal at the host–pathogen interface. *Cell Microbiol* 12(12):1691–1702
94. Nau GJ, Richmond JF, Schlesinger A, Jennings EG, Lander ES, Young RA (2002) Human macrophage activation programs induced by bacterial pathogens. *Proc Nat Acad Sci* 99(3):1503–1508
95. Navratil V, de Chasse B, Meyniel L, Delmotte S, Gautier C, André P, Lotteau V, Rabourdin-Combe C (2009) VirHost-Net: a knowledge base for the management and the analysis of proteome-wide virus–host interaction networks. *Nucleic Acids Res* 37(suppl 1):D661–D668
96. Olsen JE, Hoegh-Andersen KH, Casadesús J, Rosenkranz J, Chadfield MS, Thomsen LE (2013) The role of flagella and chemotaxis genes in host pathogen interaction of the host adapted *Salmonella enterica* serovar Dublin compared to the broad host range serovar S. Typhimurium *BMC Microbiol* 13(1):1
97. Onstad DW (1997) Ecological database of the world's insect pathogens (edwip)
98. Pandey KC, Singh N, Arastu-Kapur S, Bogyo M, Rosenthal PJ (2006) Falstatin, a cysteine protease inhibitor of *Plasmodium falciparum*, facilitates erythrocyte invasion. *PLoS Pathogen* 2(11):e117
99. Pickett BE, Sadat EL, Zhang Y, Noronha JM, Squires RB, Hunt V, Liu M, Kumar S, Zaremba S, Gu Z et al (2012) ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic Acids Res* 40(D1):D593–D598
100. Pohlner J, Halter R, Beyreuther K, Meyer TF (1986) Gene structure and extracellular secretion of *Neisseria gonorrhoeae* IgA protease. *Nature* 325(6103):458–462
101. Pukatzki S, Ma AT, Sturtevant D, Krastins B, Sarracino D, Nelson WC, Heidelberg JF, Mekalanos JJ (2006) Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the *Dictyostelium* host model system. *Proc Nat Acad Sci* 103(5):1528–1533
102. Qi Y, Tastan O, Carbonell JG, Klein-Seetharaman J, Weston J (2010) Semi-supervised multi-task learning for predicting interactions between HIV-1 and human proteins. *Bioinformatics* 26(18):i645–i652
103. Rachman H, Strong M, Ulrichs T, Grode L, Schuchhardt J, Mollenkopf H, Kosmiadi GA, Eisenberg D, Kaufmann SH (2006) Unique transcriptome signature of *Mycobacterium tuberculosis* in pulmonary tuberculosis. *Infect Immun* 74(2):1233–1242
104. Raghunathan A, Reed J, Shin S, Palsson B, Daefler S (2009) Constraint-based analysis of metabolic capacity of *Salmonella typhimurium* during host–pathogen interaction. *BMC Syst Biol* 3(1):38
105. Rappoport N, Linial M (2012) Viral proteins acquired from a host converge to simplified domain architectures. *PLoS Comput Biol* 8(2):e1002364
106. Rescigno M, Borrow P (2001) The host–pathogen interaction: new themes from dendritic cell biology. *Cell* 106(3):267–270
107. Rupp S, Sohn K (2009) Host–pathogen interactions: methods and protocols. Humana Press
108. Sansonetti P (2002) Host–pathogen interactions: the seduction of molecular cross talk. *Gut* 50(suppl 3):iii2–iii8
109. Sasseti CM, Boyd DH, Rubin EJ (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Molec Microbiol* 48(1):77–84
110. Sasseti CM, Rubin EJ (2003) Genetic requirements for mycobacterial survival during infection. *Proc Nat Acad Sci* 100(22):12,989–12,994
111. Scaria V, Hariharan M, Maiti S, Pillai B, Brahmachari SK (2006) Host-virus interaction: a new role for microRNAs. *Retrovirology* 3(1):68
112. Scaria V, Hariharan M, Pillai B, Maiti S, Brahmachari SK (2007) Host–virus genome interactions: macro roles for microRNAs. *Cell Microbiol* 9(12):2784–2794
113. Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, Hannay T, Buetow KH (2009) PID: the pathway interaction database. *Nucleic Acids Res* 37(suppl 1):D674–D679
114. Schnappinger D, Ehrt S, Voskuil MI, Liu Y, Mangan JA, Monahan IM, Dolganov G, Efron B, Butcher PD, Nathan C et al (2003) Transcriptional adaptation of *Mycobacterium tuberculosis* within macrophages insights into the phagosomal environment. *J Exper Med* 198(5):693–704
115. Sessions OM, Barrows NJ, Souza-Neto JA, Robinson TJ, Hershey CL, Rodgers MA, Ramirez JL, Dimopoulos G, Yang PL, Pearson JL et al (2009) Discovery of insect and human dengue virus host factors. *Nature* 458(7241):1047–1050
116. Sharma OP, Jadhav A, Hussain A, Kumar MS (2011) VPDB: viral protein structural database. *Bioinformatics* 6(8):324
117. Singh I, Tastan O, Klein-Seetharaman J (2010) Comparison of virus interactions with human signal transduction pathways. In: Proceedings of the First ACM international conference on bioinformatics and computational biology. ACM, pp 17–24
118. Singh SB, Davis AS, Taylor GA, Deretic V (2006) Human IRGM induces autophagy to eliminate intracellular mycobacteria. *Science* 313(5792):1438–1441
119. Smoot D, Mobley H, Chippendale G, Lewison J, Resau J (1990) *Helicobacter pylori* urease activity is toxic to human gastric epithelial cells. *Infect Immun* 58(6):1992–1994
120. Squires B, Macken C, Garcia-Sastre A, Godbole S, Noronha J, Hunt V, Chang R, Larsen CN, Klem E, Biersack K et al (2008) BioHealthBase: informatics support in the elucidation of influenza virus host–pathogen interactions and virulence. *Nucleic Acids Res* 36(suppl 1):D497–D503
121. Sugimoto S, Iwamoto T, Takada K, Okuda Ki, Tajima A, Iwase T, Mizunoe Y (2013) *Staphylococcus epidermidis* Esp degrades specific proteins associated with *Staphylococcus aureus* biofilm formation and host–pathogen interaction. *J Bacteriol* 195(8):1645–1655
122. Talaat AM, Lyons R, Howard ST, Johnston SA (2004) The temporal expression profile of *Mycobacterium tuberculosis* infection in mice. *Proc Nat Acad Sci USA* 101(13):4602–4607
123. Tastan O, Qi Y, Carbonell JG, Klein-Seetharaman J (2009) Prediction of interactions between HIV-1 and human proteins by information integration. In: Pacific Symposium on biocomputing. NIH Public Access, p 516
124. Tato C, Hunter C (2002) Host–pathogen interactions: subversion and utilization of the NF- κ B pathway during infection. *Infect Immun* 70(7):3311–3317
125. Tekir SD, Çakır T, Ardic E, Sayılırbaş AS, Konuk G, Konuk M, Saryer H, Uğurlu A, Karadeniz İ, Özgür A et al (2013) PHISTO: pathogen–host interaction search tool. *Bioinformatics* 29(10):1357–1358
126. Thieu T, Joshi S, Warren S, Korkin D (2012) Literature mining of host–pathogen interactions: comparing feature-based supervised learning and language-based approaches. *Bioinformatics* 28(6):867–875
127. Tobin DM, May RC, Wheeler RT (2012) Zebrafish: a see-through host and a fluorescent toolbox to probe host–pathogen interaction. *PLoS Pathogens* 8(1)

128. Torrelles JB, Schlesinger LS (2010) Diversity in *Mycobacterium tuberculosis* mannosylated cell wall determinants impacts adaptation to the host. *Tuberculosis* 90(2):84–93
129. Vergne I, Singh S, Roberts E, Kyei G, Master S, Harris J, Haro Sd, Naylor J, Davis A, Delgado M et al (2006) Autophagy in immune defense against *Mycobacterium tuberculosis*. *Autophagy* 2(3):175–178
130. Via A, Uyar B, Brun C, Zanzoni A (2015) How pathogens use linear motifs to perturb host cell networks. *Trends Biochem Sci* 40(1):36–48
131. Vodovar N, Acosta C, Lemaitre B, Boccard F (2004) *Drosophila*: a polyvalent model to decipher host–pathogen interactions. *Trends Microbiol* 12(5):235–242
132. Wang Y, Zhang Q, Sun Ma, Guo D (2011) High-accuracy prediction of bacterial type III secreted effectors based on position-specific amino acid composition profiles. *Bioinformatics* 27(6):777–784
133. Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, Gillespie JJ, Gough R, Hix D, Kenyon R et al (2013) PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res*:gkt1099
134. Weekes MP, Tomasec P, Huttlin EL, Fielding CA, Nusinow D, Stanton RJ, Wang EC, Aicheler R, Murrell I, Wilkinson GW et al (2014) Quantitative temporal viromics: an approach to investigate host–pathogen interaction. *Cell* 157(6):1460–1472
135. Welch R, Dellinger E, Minshew B, Falkow S (1981) Haemolysin contributes to virulence of extra-intestinal *E. coli* infections. *Nature* 294(5842):665–667
136. Winnenburg R, Baldwin TK, Urban M, Rawlings C, Köhler J, Hammond-Kosack KE (2006) PHI-base: a new database for pathogen host interactions. *Nucleic Acids Res* 34(suppl 1):D459–D464
137. Winnenburg R, Urban M, Beacham A, Baldwin TK, Holland S, Lindeberg M, Hansen H, Rawlings C, Hammond-Kosack KE, Köhler J (2008) PHI-base update: additions to the pathogen–host interaction database. *Nucleic Acids Res* 36(suppl 1):D572–D576
138. Wuchty S (2011) Computational prediction of host–parasite protein interactions between *P. falciparum* and *H. sapiens*. *PLoS One* 6(11):e26,960
139. Xiang Z, Tian Y, He Y et al (2007) PHIDIAS: a pathogen–host interaction data integration and analysis system. *Genom Biol* 8(7):R150
140. Yang J, Chen L, Sun L, Yu J, Jin Q (2008) VFDB 2008 release: an enhanced web-based resource for comparative pathogenomics. *Nucleic Acids Res* 36(suppl 1):D539–D542
141. Zelle MR (1942) Genetic constitutions of host and pathogen in mouse typhoid. *J Infect Dis* 71(2):131–152
142. Zhou C, Smith J, Lam M, Zemla A, Dyer MD, Slezak T (2007) MvirDB—a microbial database of protein toxins, virulence factors and antibiotic resistance genes for bio-defence applications. *Nucleic Acids Res* 35(suppl 1):D391–D394
143. Zhou H, Xu M, Huang Q, Gates AT, Zhang XD, Castle JC, Stec E, Ferrer M, Strulovici B, Hazuda DJ et al (2008) Genome-scale RNAi screen for host factors required for HIV replication. *Cell Host Microbe* 4(5):495–504
144. Zychlinsky A, Sansonetti PJ (1997) Apoptosis as a proinflammatory event: what can we learn from bacteria-induced cell death? *Trends Microbiol* 5(5):201–204