Published in final edited form as:

Vaccine. 2013 January 21; 31(5): 797–805. doi:10.1016/j.vaccine.2012.11.066.

Systematic annotation and analysis of “virmugens” - virulence factors whose mutants can be used as live attenuated vaccines

Rebecca Racz¹, Monica Chung¹, Zuoshuang Xiang^{1,2,3}, and Yongqun He^{2,3,4,ξ}

¹College of Literature, Science, and the Arts, University of Michigan, Ann Arbor, MI 48109, USA

²Unit for Laboratory Animal Medicine, University of Michigan Medical School, Ann Arbor, MI 48109, USA

³Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI 48109, USA

⁴Center for Computational Medicine and Biology, University of Michigan, Ann Arbor, MI 48109, USA

Abstract

Live attenuated vaccines are usually generated by mutation of genes encoding virulence factors. “Virmugen” is coined here to represent a gene that encodes for a virulent factor of a pathogen and has been proven feasible in animal models to make a live attenuated vaccine by knocking out this gene. Not all virulence factors are virmugens. VirmugenDB is a web-based virmugen database (<http://www.violinet.org/virmugendb>). Currently, VirmugenDB includes 225 virmugens that have been verified to be valuable for vaccine development against 57 bacterial, viral, and protozoan pathogens. Bioinformatics analysis has revealed significant patterns in virmugens. For example, 10 Gram-negative and one Gram-positive bacterial *aroA* genes are virmugens. A sequence analysis has revealed at least 50% of identities in the protein sequences of the 10 Gram-negative bacterial *aroA* virmugens. As a pathogen case study, *Brucella* virmugens were analyzed. Out of 15 verified *Brucella* virmugens, six are related to carbohydrate or nucleotide transport and metabolism, and two involving cell membrane biogenesis. In addition, 54 virmugens from 24 viruses and 12 virmugens from 4 parasites are also stored in VirmugenDB. Virmugens tend to involve metabolism of nutrients (*e.g.*, amino acids, carbohydrates, and nucleotides) and cell membrane formation. Host genes whose expressions were regulated by virmugen mutation vaccines or wild type virulent pathogens have also been annotated and systematically compared. The bioinformatics annotation and analysis of virmugens helps elucidate enriched virmugen profiles and the mechanisms of protective immunity, and further supports rational vaccine design.

Keywords

live attenuated vaccine; virulence factor; virmugen; bacteria; virus; protozoa; database; bioinformatics; meta-analysis

© 2012 Elsevier Ltd. All rights reserved.

^ξCorresponding author: Unit for Laboratory Animal Medicine University of Michigan Medical School 018 Animal Research Facility 1150 W. Medical Dr. Ann Arbor, MI 48109 yongqunh@umich.edu Tel: 734-615-8231 .

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

A live attenuated vaccine contains a live attenuated pathogen that has one or more gene mutations from its wild type pathogen and is able to induce protection in a host. Therefore, live attenuated vaccines have two common traits: (i) attenuated virulence to a host, and (ii) induction of protective immunity in the host against a challenge with a virulent pathogen. The gene mutation is required to attenuate the virulence of the live pathogen, making the vaccine relatively safe for the vaccinated host. Live attenuated vaccines may still grow inside the host, but the growth is greatly limited. To make the attenuated vaccine live allows the attenuated vaccine strain infect the host in a pathway that is often similar to the virulent parent pathogen. This way the vaccine most likely stimulates strong cellular and antibody responses as well as some time of immunity, which can be short term, long term or lifelong. One difference between virulent pathogen infection and vaccination is that after a host is infected with a virulent wild type pathogen, the survived host is normally resistant to any re-infection, and the protective immune response in the host usually lasts for life. However, the immunity resulting from vaccination with a live attenuated vaccine is often short lived although a lifelong immunity may sometimes be induced. However, the exact differences between live attenuated vaccines and their virulent wild type strains in terms of immune response inductions are typically unclear.

The mutations generated in live attenuated vaccines are on genes encoding for virulence factors. Virulence factors are the various parts of a pathogen that allow the microbe to infect, survive, and replicate in the host [1]. While all live attenuated vaccines contain mutations in one or more virulence factors, not all virulence factors can be mutated to create a live attenuated vaccine. A mutation of a virulence factor in a virulent pathogen will typically lead to attenuation. However, the mutation of a virulence factor may not generate a stable attenuated microbe, or the attenuated strain does not stimulate protective immune response in the host against virulent infection. For simplicity, we have coined the term “virmugen” to represent those virulence factors that can be mutated in wild type virulent pathogens for generation of live attenuated vaccines that stimulate protective immunity against virulent pathogen infections. In our definition, the “live attenuated vaccines” do not have to be licensed vaccines. A live attenuated vaccine can be any vaccine candidate that is live, attenuated, and proven effective to induce protection against virulent pathogen infection in a vaccination-challenge experiment using a well-established laboratory animal model [2, 3].

For several decades researchers have been developing live attenuated vaccines, but systematic analysis of virmugens has not been performed. To address this, we have developed VirmugenDB (<http://www.violinet.org/virmugendb>), a web-based virmugen database for collecting and analyzing various virmugens. VirmugenDB stores manually curated virmugens and associated information (e.g., corresponding live attenuated vaccines) from peer-reviewed journals. The collection and analysis of virmugens are important for rational and efficient development of future live attenuated vaccines against the infections of existing and emerging pathogens.

2. Methods

2.1. System and database design

VirmugenDB is implemented using a three-tier architecture built on two HP servers which run the Redhat Linux operating system (Redhat Enterprise Linux ES 5). Users can submit database or analysis queries through the web. These queries are then processed using PHP/SQL (middle-tier, application server based on Apache) against a MySQL (version 5.0) relational database (back-end, database server). The result of each query is then presented to the user in the web browser. Two servers are scheduled to regularly backup each other's

data. VirmugenDB is a new integrated program of the comprehensive VIOLIN vaccine database and analysis system [4].

Figure 1 demonstrates the VirmugenDB database design and workflow. Each virmugen is annotated with evidence from peer-reviewed references. VirmugenDB also includes general information about individual virmugens such as gene name, pathogen strain name, sequences, and 3D structures. VirmugenDB incorporates many software programs that provide predicted results and tools for sequence analysis (Figure 1).

2.2. Semi-automatic VirmugenDB annotation

A semi-automatic approach is used in virmugen curation and analysis (Figure 1 and Supplemental Figure 1). Manual curation of peer-reviewed journal articles emphasizes the retrieval of specific virmugen information and experimental evidence. In order to determine whether or not a gene is a virmugen, experimental evidence must demonstrate attenuation in a host as well as protection from the virulent parent in a challenge study. To improve the efficacy of manual curation, an in-house web-based literature mining and curation system called Limix was used [5, 6]. The interactive Limix data submission and review system: (a) allows a curator to search literature, copy and edit text, and submit data to database, and (b) provides a data reviewer tools to review, edit, and approve the curated data on one comprehensive web interface. A version control and manage system stores different versions of editions and allows a reviewer to compare history versions to trace the detailed changes. Only after approval by a domain expert, the data can be posted publicly (Supplemental Figure 1). The authors and other curators in the team check frequently for possibly new virmugens from published papers in PubMed, and add experimentally verified virmugens to the VirmugenDB. After approval by a domain expert, curated new virmugens from the database will be published in the VirmugenDB website and available for the public to query. We plan to regularly update the customized virmugen BLAST library and the data download site 2-3 times a year.

Different automatic approaches were also developed. Using manually curated PubMed ID, NCBI gene ID (or protein ID or nucleotide ID), or an ID from Protein Data Bank (PDB; www.rcsb.org), our programs in Limix are able to automatically retrieve detailed citation content, gene and protein information (*e.g.*, sequences, names, functions), and 3D structure information, respectively, from different databases. The information is then stored in the VirmugenDB and seamlessly integrated with other content of a specific virmugen (Figure 1 and Supplemental Figure 1).

2.3. Customized BLAST sequence similarity search

To facilitate virmugen-associated sequence analysis, two customized BLAST libraries were generated. One library contains the protein sequences of all virmugens, and the other collects all the DNA sequences of all virmugens. Based on the customized virmugens libraries, a customized BLAST program is generated and allows a user to compare a custom protein or DNA sequence with all sequences collected in the VirmugenDB database.

2.4. Sequence alignment and phylogenetic tree analysis

The AroA protein sequences from 11 bacteria (see Supplemental Table 1 for detailed bacterial strain and protein ID information) were aligned with ClustalW2 [7]. The MEGA 4.0 package [8] was used to generate a phylogenetic tree structure for these 11 AroA proteins. Specifically, the sequences of these 11 AroA proteins were first aligned using the ClustalW program within the MEGA4 software. The phylogenetic tree of these proteins was generated using a Neighbor Joining method [9]. Bootstrap analysis using 500 repetitions provided support for individual nodes [10].

2.5. Methods for generating VirmugenDB data query and display

All data in VirmugenDB are freely available for public queries. The query interface was generated using PHP and HTML. Any query is submitted to the Virmugen MySQL relational database through a SQL query program. The results are returned to the browser in a user-friendly format.

2.6. Data exchange and download

The information of virrugens in VirmugenDB was stored in the Vaccine Ontology (VO; <http://www.violinet.org/vaccineontology>) [11, 12]. Developed based on the Web Ontology Language (OWL) format (<http://www.w3.org/TR/owl-ref/>), the VO content can be processed by software programs and used to support automated reasoning and Semantic Web applications [13]. An example of virrugens representation in VO is demonstrated in Supplemental Figure 2. All virrugens DNA and protein sequence data are freely downloadable in FASTA format from the VirmugenDB website (<http://www.violinet.org/virmugendb>).

3. Results

3.1. VirmugenDB statistics

Currently, VirmugenDB has included 225 virrugens as well as 196 virrugens-related vaccines. Based on the results from vaccination-challenge experiments, these virrugens have individually been verified to be valid for development of protective live attenuated vaccines against 57 bacterial, viral, and protozoan pathogens. It is noted that a majority of the virrugens stored in the database are identified from peer-reviewed literature papers that introduce the development and evaluation of effective live attenuated vaccines using established laboratory animal models. Most of these vaccines are not licensed vaccines. However, each vaccine candidate used in our analysis was proven effective to protect against virulent pathogen challenge in at least one vaccination-challenge experiment.

Below we will analyze these data based on the pathogen categories.

3.2. Analysis of bacterial virrugens in VirmugenDB

In total we have 159 bacterial virrugens for 29 bacterial pathogens. Out of these bacterial virrugens, many “*aro*”, “*pur*”, and “*gua*” genes are found (Supplemental Table 1).

The “*aro*” genes in many pathogens have been mutated to develop protective live attenuated vaccines (Supplemental Table 1). Specifically, 11 *aroA*, three *aroC*, and two *aroD* genes have been mutated and proven to be virrugens. Live attenuated vaccines have been generated by mutating *aroA* in 11 strains belonging to eight genus such as *Bordetella* [14], *Pasteurella* [15], *Salmonella* [16], *Shigella* [17], *Staphylococcus* [18], and *Yersinia* [19]. Except Gram-positive *S. aureus*, all other pathogens are Gram-negative. The mutation of *aroC* or *aroD* has also been used for developing live attenuated vaccines for four pathogens including *Salmonella* [20], *Shigella* [21], *Burkholderia* [22], and *Edwardsiella* [23]. Bacterial *aroA*, *aroC*, and *aroD* genes code for 3-phosphoshikimate 1-carboxyvinyltransferase, chorismate synthase, 3-dehydroquinate dehydratase, respectively. The “*aro*” genes are important to produce aromatic metabolites, mainly aromatic amino acids such as phenylalanine, tyrosine, and tryptophan. If an “*aro*” gene is mutated, the mutant will be unable to synthesize aromatic amino acids and cannot replicate within a host in which aromatic compounds are not freely available. In addition, *aro* mutants often have a defect in cell envelope biosynthesis [24].

To better identify the similarities and differences between experimentally verified *aroA* virgugens, the AroA protein sequences were aligned and the phylogram among these proteins was generated using ClustalW2 (Figure 2). The multiple sequence alignment shows more sequence similarities of AroA proteins among Gram-negative bacteria than the comparison between Gram-negative bacteria and Gram-positive *S. aureus* (Figure 2A). The pairwise scores (*i.e.*, the number of identities between the two sequences, divided by the length of the alignment) are equal to or greater than 50% between all Gram-negative bacteria. However, all pairwise scores between the AroA protein sequences of *S. aureus* and Gram-negative bacteria are between 22% and 26%. Although different pairwise scores were found, all AroA proteins from different bacteria share the same function. In the phylogram generated with AroA protein sequences, *S. aureus* is the root of the tree and significantly distant from all other Gram-negative bacteria (Figure 2B), suggesting the evolutionary difference between the same proteins from Gram-positive and Gram-negative bacteria.

Many bacterial “*pur*” genes (including *purC-F*, and *purK-N*) and “*gua*” genes (including *guaA* and *guaB*) have been used for generating live attenuated vaccines (Supplemental Table 1). The “*pur*” gene products catalyze reactions involved in purine biosynthesis. Two of the four bases in nucleic acids, adenine and guanine, are purines. Nucleotides serve as not only a structural unit of DNA and RNA, but also a key component in signaling and metabolism. Bacterial “*gua*” gene encodes for GMP synthase or inosine 5'-monophosphate dehydrogenase, which participate in the synthesis of guanosine monophosphate (GMP) [25].

As a bacterium use case study, we analyzed all curated virgugens in *Brucella* spp. (Table 1). *Brucella* is a facultative, Gram-negative bacterium that causes zoonotic brucellosis in humans and different animal species. Approximately 500,000 new human brucellosis cases are reported annually worldwide, making brucellosis one of the most common zoonotic disease in the world [26]. In total, 12 live attenuated *Brucella* vaccines generated using 15 virgugens have been found to deliver protective immunity (Table 1). Two of the 15 *Brucella* virgugens are *purE* and *purK* that involve nucleotide transport and metabolism and have been described above. Four *Brucella* virgugens (*manA*, *manB*, *pgk*, and *pgm*) encode for carbohydrate transport and metabolism proteins. Carbohydrates play an important role in various biological processes including energy storage and DNA and RNA structure buildup. Omp25 and Omp31, two virgugen proteins, involve in cell membrane biogenesis. Five *Brucella* virgugens are cytoplasmic proteins, three are cytoplasmic membrane proteins, two outer membrane proteins, two periplasmic proteins, and three with unknown subcellular locations (Table 1).

3.3. Analysis of viral virgugens in VirgugenDB

In total we have 54 viral virgugens for 24 viral pathogens. Three groups of viral virgugens have more than one gene: TK/UL23, glycoproteins, and matrix proteins (Table 2). Thymidine kinase (TK or UL23) is mutated in several herpes viruses infecting different host species including humans [27], horse [28], swine [29], cat [30], and chicken [31]. Thymidine kinase is an enzyme important in DNA repair, replication, and nucleotide metabolism [32]. This kinase is responsible for phosphorylation of deoxythymidine in the pathway of DNA synthesis [33]. Many glycoprotein genes from 10 viruses have been mutated for generating live attenuated viral vaccines (Table 2). A glycoprotein is a class of proteins that have a carbohydrate group attached [34]. Glycoproteins are frequently found in membranes and help a virus bind to the host cell. M2 and M2-2, two matrix proteins, have also been identified to be virgugen proteins in influenza virus, human metapneumovirus, and human respiratory syncytial virus (Table 2). The M protein is crucial in the formation of extracellular virus particles [35]. Viral matrix proteins mediate the regulatory switch from transcription to RNA replication and act late in infection by inhibiting viral transcription and up-regulating RNA replication [36]. Therefore, mutation of a matrix protein gene would

reduce the viral pathogenesis by reducing viral replication and late infection capabilities of viruses. The mutation of this protein is only effective for RNA viruses, and DNA viruses do not have the equivalent of an M protein in RNA viruses.

3.4. Analysis of protozoan virgugens in VirgugensDB

Only 12 protozoan virgugens were identified, and no genes are orthologs (Supplemental Table 2). Several virgugens are cell membrane proteins or involve in nutrient metabolism, Py52 and UIS3 are two membrane proteins from *Plasmodium yoelii*. Membrane bound proteins could aid in attachment or invasion of host cells. Mic2 from *Toxoplasma gondii*, is a transmembrane adhesion protein that alters the protozoan's ability to invade the host cell [37]. Orotidine monophosphate decarboxylase (OMPDC) and uridine phosphorylase (UP), two proteins from *T. gondii*, are involved in pyrimidine synthesis and metabolism.

3.5. Host responses to live attenuated vaccines generated by mutation of virgugens

To study host responses induced by live attenuated vaccines, a comparative meta-analysis of results published in different papers has been performed. Host response comparison varied greatly across papers in different studies, ranging from comparison with naïve control subjects, control subjects infected with the wild type pathogen, to comparison with more specific treatments such as other vaccines. When naïve subjects were compared with host or host cells infected with a live attenuated vaccine, every regulation case was almost an up regulation of the host gene in comparison to the controls.

The most interesting comparison was with subjects infected with the wild type pathogen. Out of the 47 cases of host gene regulation that was compared to infection with the wild type, 18 cases (38%) were down regulated (Table 3). Two genes with particularly significant patterns in regulation are interferon gamma (IFN- γ) and interleukin 4 (IL-4). Both of these were mostly down-regulated, with IFN-gamma down-regulated by 7 virgugens (6 vaccines) and IL-4 down-regulated by 4 virgugens (4 vaccines). In contrast, only one virgugen (*sigE*) mutant of *M. tuberculosis* up-regulated the production of IFN- γ from lung [38]. The production of IL-4 was not updated by any virgugen. The productions of IL-6, IL-10, TNF- α , and IgG were up- or down-regulated in the infected hosts depending on the virgugens and pathogen targeted. For example, compared to host infected with virulent parent strains, IL-6 was down-regulated in live attenuated *F. tularensis* FTL0552 mutant, but up-regulated in an *E. coli* and a *Salmonella* mutant. Interestingly, TNF- α was down-regulated in a *M. tuberculosis* *adD26* mutant but up-regulated in *M. tuberculosis* *sigE* mutant. These suggest that TNF- α can be up- or down-regulated in live attenuated mutant vaccines for the same pathogen.

3.6. VirgugensDB data download, exchange, and web query

The virgugen data collected in VirgugensDB are freely available for download and exchange as described in the Methods section. The manually curated and pre-computed VirgugensDB data can be efficiently queried and visualized (Supplemental Figure 3A-D). Customized BLAST sequence analysis is also provided (Supplemental Figure 3E).

4. Discussion

For a specific vaccine research and development laboratory, the selection of a pathogen gene to mutate for the purpose of generating a live attenuated vaccine is usually *ad hoc* without systematic understanding of what will happen. However, as identified in this study, many virgugens encode for proteins involving in basic amino acid, carbohydrate, and nucleotide transport and metabolism, or cell membrane formation (*e.g.*, viral glycoproteins). The virulence of these mutants is decreased inside host due to their reduced capability of binding

to/infecting host cells or replicating inside host. However, protective immunity can still be induced.

It is still unclear why mutating a particular virulence mutant causes specific host gene response. When IFN-gamma and IL-4 were down-regulated, many papers simply stated that because the virulence mutant vaccine was attenuated, the mutant induced less inflammation and therefore a less robust immune response [39, 40]. On the other hand, papers that described up-regulation of host genes hypothesized that because the vaccine was protective, it makes sense that a more robust immune response would occur in host infected with a live attenuated vaccine [38, 41, 42]. Our meta-analysis results show up- or down-regulation of the same host genes may occur in different virulence mutants of the same pathogen (Table 3). More research needs to be done on specific virulence mutants and how they react with the host in order to figure out why the host responds in a certain way when the virulence mutants are mutated.

Numerous Omics data associated with host responses to virulent pathogens and live attenuated vaccines have been published and available online. Although the large amount of Omics data is available, many Omics data have noise and require additional experiments for verification. Therefore, our database will include only those experimentally verified host gene responses. Those Omics data analysis results without experimental verifications will not normally be included in our database. However, the Omics data may be very helpful for generating new hypotheses and knowledge. The combination of the experimentally verified results and Omics data analysis will help further advancing the analysis of host genes specifically related to protective immunity.

To explain how virulence mutants can function as live attenuated vaccines, we have developed a “virulence working model” (Figure 3). This model suggests that compared to the wild type virulent parent strain, a virulence mutant may stimulate a specific host gene response associated with the induction of protective immunity or successful host defense activity that results in the killing of the attenuated vaccine strain. On the other hand, a series of the host responses against virulent parent strain may be associated with pathological processes that lead to the outcome of a specific disease. It is crucial to differential different sets of host responses to wild type virulent parent strain and virulence mutants.

In order to have a clearer understanding of how these virulence mutants function with respect to the host immune system, a majority of the live attenuated vaccines found in VirulenceDB have a mutation in a single gene. The single gene criterion was purposeful, as it allows for examination and analysis of the true function of that particular gene within host infection and immune response. In many cases, live attenuated vaccines are created through multigenic mutations. However, with multiple genes mutated, it is often difficult to determine how these genes are interacting with each other or their individual roles in infection and immunity. Therefore, VirulenceDB prefers to collect virulence mutants whose corresponding vaccines include only single gene mutations. Occasionally those live attenuated vaccines with multigenic mutations were included as evidence for obtaining and justifying new virulence mutants. In these cases, a careful annotation was taken to ensure that each gene was mutated as a virulence factor. Also, it should be noted that virulence is usually multigenic with different virulence factors working together. The mutation of one gene may disrupt a pathogenesis pathway involving many virulence factors. Sometimes a deletion of a single virulence factor may not lead to obvious decrease of virulence, probably due to the presence of an alternative pathway(s) in the pathogen. In this case, a simultaneous deletion of two or even more virulence factors may be required to achieve the status of attenuation.

It is common that an attenuated vaccine candidate does not provide significant protection. This was the case with a *manBA* mutant from *Brucella abortus* [43], an *ofs* mutant from *Streptococcus suis* [44], and *3abc* and *7ab* mutant in Feline Infectious Peritonitis [45]. The failure to elicit significant protection may be due to an organism that is too attenuated to stimulate an immune response. A live attenuated vaccine must be carefully balanced to ensure that the organism is attenuated enough so that it does not cause illness, but not too attenuated to induce an immune response [26].

To the best of our knowledge, the authors have not been able to identify any other immunoinformatics database that is similar to our VirmugenDB. The VirmugenDB database can be used for many applications. For example, the virumugens can be used for rational vaccine design based on successful virmugen vaccines in other pathogens. Different virumugens stimulate different immune responses, which addresses the need for specific immune responses to confer protection against disease. Many diseases are complex in their interaction with the immune system, thus making vaccine design difficult. Usage of different types of virumugens induces different immune responses, which can be tailored for use in vaccines against specific pathogens and diseases that require different responses.

5. Conclusions

VirmugenDB is the first database that collects virumugens, *i.e.*, the virulent factors whose mutation leads to the generation of successful live attenuated vaccines. The analysis of VirmugenDB data allows us to identify specific patterns of virumugens and host responses induced by virmugen mutants. VirmugenDB is targeted to become a central and vital source of virumugens and will support researchers in the areas of vaccinology, microbiology, and immunology with curated data and bioinformatics tools. In the time of extensive vaccine research, VirmugenDB is a timely repository and will have significant impact for rational vaccine development and better understanding of fundamental protective immune mechanisms.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by NIH-NIAID grant R01AI081062.

References

- [1]. Casadevall A, Pirofski LA. Virulence factors and their mechanisms of action: the view from a damage-response framework. *J Water Health*. 2009; 7(Suppl 1):S2–S18. [PubMed: 19717929]
- [2]. Brinkman RR, Courtot M, Derom D, Fostel J, He Y, Lord P, et al. Modeling biomedical experimental processes with OBI. *Journal of Biomedical Semantics*. Jun 22.2010 1(Suppl 1):S7. doi: 10.1186/2041-1480-1-S1-S7. [PubMed: 20626927]
- [3]. Richt JA, Garcia-Sastre A. Attenuated influenza virus vaccines with modified NS1 proteins. *Curr Top Microbiol Immunol*. 2009; 333:177–95. [PubMed: 19768406]
- [4]. Xiang Z, Todd T, Ku KP, Kovacic BL, Larson CB, Chen F, et al. VIOLIN: vaccine investigation and online information network. *Nucleic Acids Res*. Jan; 2008 36(Database issue):D923–8. [PubMed: 18025042]
- [5]. Xiang Z, Zheng W, He Y. BBP: Brucella genome annotation with literature mining and curation. *BMC Bioinformatics*. Jul 16.2006 7(1):347. [PubMed: 16842628]
- [6]. Yang B, Sayers S, Xiang Z, He Y. Protegen: a web-based protective antigen database and analysis system. *Nucleic Acids Res*. Oct 19.2010

- [7]. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics* (Oxford, England). Nov 1; 2007 23(21):2947–8.
- [8]. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol.* Aug; 2007 24(8):1596–9. [PubMed: 17488738]
- [9]. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol.* Jul; 1987 4(4):406–25. [PubMed: 3447015]
- [10]. Shayman JA, Kelly R, Kollmeyer J, He Y, Abe A. Group XV phospholipase A, a lysosomal phospholipase A. *Prog Lipid Res.* Jan; 2011 50(1):1–13. [PubMed: 21074554]
- [11]. He, Y.; Cowell, L.; Diehl, AD.; Mobley, HL.; Peters, B.; Ruttenberg, A., et al. VO: Vaccine Ontology. The 1st International Conference on Biomedical Ontology (ICBO 2009) Nature Precedings; Buffalo, NY, USA. 2009. <http://precedings.nature.com/documents/3553/version/1>
- [12]. Ozgur A, Xiang Z, Radev D, He Y. Mining of vaccine-associated IFN- γ gene interaction networks using the Vaccine Ontology. *Journal of Biomedical Semantics.* 2011 In press: <http://www.jbiomedsem.com/qc/content/2/S/S8>.
- [13]. Xiang Z, Courtot M, Brinkman RR, Ruttenberg A, He Y. OntoFox: web-based support for ontology reuse. *BMC Res Notes.* 2010; 3:175. [PubMed: 20569493]
- [14]. Stevenson A, Roberts M. Use of a rationally attenuated *Bordetella bronchiseptica* as a live mucosal vaccine and vector for heterologous antigens. *Vaccine.* May 22; 2002 20(17-18):2325–35. [PubMed: 12009288]
- [15]. Dagleish MP, Hodgson JC, Ataei S, Finucane A, Finlayson J, Sales J, et al. Safety and protective efficacy of intramuscular vaccination with a live *aroA* derivative of *Pasteurella multocida* B:2 against experimental hemorrhagic septicemia in calves. *Infect Immun.* Dec; 2007 75(12):5837–44. [PubMed: 17875632]
- [16]. Dougan G, Maskell D, Pickard D, Hormaeche C. Isolation of stable *aroA* mutants of *Salmonella typhi* Ty2: properties and preliminary characterisation in mice. *Mol Gen Genet.* May; 1987 207(2-3):402–5. [PubMed: 3039297]
- [17]. Noriega FR, Wang JY, Losonsky G, Maneval DR, Hone DM, Levine MM. Construction and characterization of attenuated delta *aroA* delta *virG* *Shigella flexneri* 2a strain CVD 1203, a prototype live oral vaccine. *Infect Immun.* Nov; 1994 62(11):5168–72. [PubMed: 7927802]
- [18]. Buzzola FR, Barbagelata MS, Caccuri RL, Sordelli DO. Attenuation and persistence of and ability to induce protective immunity to a *Staphylococcus aureus aroA* mutant in mice. *Infect Immun.* Jun; 2006 74(6):3498–506. [PubMed: 16714581]
- [19]. Bowe F, O'Gaora P, Maskell D, Cafferkey M, Dougan G. Virulence, persistence, and immunogenicity of *Yersinia enterocolitica* O:8 *aroA* mutants. *Infect Immun.* Oct; 1989 57(10):3234–6. [PubMed: 2777382]
- [20]. Tacket CO, Levine MM. CVD 908, CVD 908-htrA, and CVD 909 live oral typhoid vaccines: a logical progression. *Clin Infect Dis.* Jul 15; 2007 45(Suppl 1):S20–3. [PubMed: 17582563]
- [21]. Karnell A, Cam PD, Verma N, Lindberg AA. *AroD* deletion attenuates *Shigella flexneri* strain 2457T and makes it a safe and efficacious oral vaccine in monkeys. *Vaccine.* 1993; 11(8):830–6. [PubMed: 8356844]
- [22]. Srilunchang T, Proungvitaya T, Wongratanaheewin S, Strugnelli R, Homchampa P. Construction and characterization of an unmarked *aroC* deletion mutant of *Burkholderia pseudomallei* strain A2. *Southeast Asian J Trop Med Public Health.* Jan; 2009 40(1):123–30. [PubMed: 19323044]
- [23]. Xiao J, Chen T, Wang Q, Liu Q, Wang X, Lv Y, et al. Search for live attenuated vaccine candidate against edwardsiellosis by mutating virulence-related genes of fish pathogen *Edwardsiella tarda*. *Lett Appl Microbiol.* Oct; 2011 53(4):430–7. [PubMed: 21777261]
- [24]. Sebkova A, Karasova D, Crhanova M, Budinska E, Rychlik I. *aro* mutations in *Salmonella enterica* cause defects in cell wall and outer membrane integrity. *J Bacteriol.* May; 2008 190(9):3155–60. [PubMed: 18310348]
- [25]. McClelland M, Sanderson KE, Spieth J, Clifton SW, Latreille P, Courtney L, et al. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature.* Oct 25; 2001 413(6858):852–6. [PubMed: 11677609]
- [26]. Ficht TA, Kahl-McDonagh MM, Arenas-Gamboa AM, Rice-Ficht AC. Brucellosis: the case for live, attenuated vaccines. *Vaccine.* Nov 5; 2009 27(Suppl 4):D40–3. [PubMed: 19837284]

- [27]. Field HJ, Wildy P. The pathogenicity of thymidine kinase-deficient mutants of herpes simplex virus in mice. *J Hyg (Lond)*. Oct; 1978 81(2):267–77. [PubMed: 212476]
- [28]. Slater JD, Gibson JS, Field HJ. Pathogenicity of a thymidine kinase-deficient mutant of equine herpesvirus 1 in mice and specific pathogen-free foals. *J Gen Virol*. May; 1993 74(Pt 5):819–28. [PubMed: 8388018]
- [29]. Weigel RM, Hahn EC, Scherba G. Survival and immunization of raccoons after exposure to pseudorabies (Aujeszky's disease) virus gene-deleted vaccines. *Vet Microbiol*. Mar 20; 2003 92(1-2):19–24. [PubMed: 12488067]
- [30]. Yokoyama N, Maeda K, Tohya Y, Kawaguchi Y, Shin YS, Ono M, et al. Pathogenicity and vaccine efficacy of a thymidine kinase-deficient mutant of feline herpesvirus type 1 in cats. *Arch Virol*. 1996; 141(3-4):481–94. [PubMed: 8645090]
- [31]. Han MG, Kweon CH, Mo IP, Kim SJ. Pathogenicity and vaccine efficacy of a thymidine kinase gene deleted infectious laryngotracheitis virus expressing the green fluorescent protein gene. *Arch Virol*. May; 2002 147(5):1017–31. [PubMed: 12021870]
- [32]. Li J, Le L, Sibley DA, Mathis JM, Curiel DT. Genetic incorporation of HSV-1 thymidine kinase into the adenovirus protein IX for functional display on the virion. *Virology*. Aug 1; 2005 338(2): 247–58. [PubMed: 15996701]
- [33]. Zimmermann W, Broll H, Ehlers B, Buhk HJ, Rosenthal A, Goltz M. Genome sequence of bovine herpesvirus 4, a bovine Rhadinovirus, and identification of an origin of DNA replication. *J Virol*. Feb; 2001 75(3):1186–94. [PubMed: 11152491]
- [34]. Spiro RG. Glycoproteins: Structure, Metabolism and Biology. *N Engl J Med*. Sep 19.1963 269:616–21. CONCL. [PubMed: 14044400]
- [35]. Rodriguez L, Cuesta I, Asenjo A, Villanueva N. Human respiratory syncytial virus matrix protein is an RNA-binding protein: binding properties, location and identity of the RNA contact residues. *J Gen Virol*. Mar; 2004 85(Pt 3):709–19. [PubMed: 14993657]
- [36]. Pham QN, Biacchesi S, Skidopoulos MH, Murphy BR, Collins PL, Buchholz UJ. Chimeric recombinant human metapneumoviruses with the nucleoprotein or phosphoprotein open reading frame replaced by that of avian metapneumovirus exhibit improved growth in vitro and attenuation in vivo. *J Virol*. Dec; 2005 79(24):15114–22. [PubMed: 16306583]
- [37]. Huynh MH, Carruthers VB. Toxoplasma MIC2 is a major determinant of invasion and virulence. *PLoS Pathog*. Aug.2006 2(8):e84. [PubMed: 16933991]
- [38]. Hernandez, Pando R.; Aguilar, LD.; Smith, I.; Manganelli, R. Immunogenicity and protection induced by a Mycobacterium tuberculosis sigE mutant in a BALB/c mouse model of progressive pulmonary tuberculosis. *Infect Immun*. Jul; 2010 78(7):3168–76. [PubMed: 20457786]
- [39]. Aguilar LD, Infante E, Bianco MV, Cataldi A, Bigi F, Pando RH. Immunogenicity and protection induced by Mycobacterium tuberculosis mce-2 and mce-3 mutants in a Balb/c mouse model of progressive pulmonary tuberculosis. *Vaccine*. Mar 20; 2006 24(13):2333–42. [PubMed: 16388878]
- [40]. Darji A, Mohamed W, Domann E, Chakraborty T. Induction of immune responses by attenuated isogenic mutant strains of Listeria monocytogenes. *Vaccine*. Jun 1; 2003 21(Suppl 2):S102–9. [PubMed: 12763691]
- [41]. Jayakar HR, Parvathareddy J, Fitzpatrick EA, Bina XR, Bina JE, Re F, et al. A galU mutant of Francisella tularensis is attenuated for virulence in a murine pulmonary model of tularemia. *BMC Microbiol*. 2011; 11:179. [PubMed: 21819572]
- [42]. Sammons-Jackson WL, McClelland K, Manch-Citron JN, Metzger DW, Bakshi CS, Garcia E, et al. Generation and characterization of an attenuated mutant in a response regulator gene of Francisella tularensis live vaccine strain (LVS). *DNA Cell Biol*. Jul; 2008 27(7):387–403. [PubMed: 18613792]
- [43]. Kahl-McDonagh MM, Arenas-Gamboa AM, Ficht TA. Aerosol infection of BALB/c mice with Brucella melitensis and Brucella abortus and protective efficacy against aerosol challenge. *Infect Immun*. Oct; 2007 75(10):4923–32. [PubMed: 17664263]
- [44]. Kock C, Beineke A, Seitz M, Ganter M, Waldmann KH, Valentin-Weigand P, et al. Intranasal immunization with a live Streptococcus suis isogenic of mutant elicited suilysin-neutralization

- titers but failed to induce opsonizing antibodies and protection. *Vet Immunol Immunopathol.* Dec 15; 2009 132(2-4):135–45. [PubMed: 19505730]
- [45]. Haijema BJ, Volders H, Rottier PJ. Live, attenuated coronavirus vaccines through the directed deletion of group-specific genes provide protection against feline infectious peritonitis. *J Virol.* Apr; 2004 78(8):3863–71. [PubMed: 15047802]
- [46]. He Y, Xiang Z, Mobley HL. Vaxign: the first web-based vaccine design program for reverse vaccinology and applications for vaccine development. *Journal of Biomedicine and Biotechnology.* 2010; Volume 2010 (2010), Article ID 297505, 15 pages.

Highlights

- On virgogens, virulence factors whose mutants are used as live attenuated vaccines.
- Curation and analysis of bacterial, viral, and protozoan virgogens.
- Specific virgogen patterns are identified.
- Host genes specifically induced by virgogen vaccines are annotated.
- A web-based virgogen database is developed for query and analysis.

Figure 1. VirmugenDB annotation workflow and system design

A semi-automatic process was performed to annotate virulence factor data. Manual curation includes peer-reviewed publications from PubMed. A PubMed ID (PMID) is extracted and used to retrieve detailed citation information (e.g., authors, journal, and date). The evidence that proves the status of virulence factor for each protein is curated from published experimental studies. Vaccines associated with the virulence factors are also curated. PDB IDs are manually retrieved when available to provide 3D structure information of individual protective antigens. Internally developed script uses an input sequence ID from a NCBI database (e.g., NCBI Entrez Gene database) to automatically retrieve different types of information. The extracted DNA and protein sequences are further used for bioinformatics analyses using different methods. A web-based VirmugenDB query system is available for public query and analysis of Virmugen data.

Figure 2. AroA amino acid sequence analysis

(A) Multiple sequence alignment results from ClustalW2. An “*” sign (asterisk) indicates positions which have a single, fully conserved residue. A “:” sign (colon) indicates conservation between groups of strongly similar properties - scoring > 0.5 in the Gonnet PAM 250 matrix. A “.” sign (period) indicates conservation between groups of weakly similar properties - scoring ≤ 0.5 in the Gonnet PAM 250 matrix. (B) Phylogeny analysis of AroA proteins from different bacteria. The detailed method is described in the Method section. The numbers next to the branches are the percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates). The tree is drawn to scale. The scale bar indicates nucleotide substitutions per site.

Figure 3. Virmugen working model

Compared to the wild type virulent parent strain, a live attenuated virmugen mutant has a mutated virmugen that encodes for a virulence factor. The virmugen virulence factors often involve synthesis and metabolism of various nutrients including amino acids, carbohydrates, and nucleotides. Some virmengens are critical for cell membrane formation. When a virmugen mutant or a virulent parent strain infects a host, similar and different host immune responses will occur. Specifically, a virulent pathogen will initiate a series of pathological host processes that lead to a host disease. The host immune system is often hijacked by the pathogen, allowing the pathogen to survive and replicate inside the host. In contrast, the host immune system will form a strong defense process(es) to eventually eliminate the live attenuated virmugen mutant inside the host. In terms of the induction of protective immunity, common protective immune responses may be generated by the host for both virmugen mutant and its parent virulent strain. In this case, the levels of immune responses between these two strains may be different. In addition, the host immune system may be able to generate a protective immunity that is unique to the live attenuated virmugen mutant (not for virulent parent strain). All the differences can be traced to the original mutation of a virmugen(s).

Table 1

Brucella virgugens

#	Gene	Protein name	Gene or protein ID	Cellular Location (Prob.)	Strain(s) mutated	Mouse Model	PMID
1	asp24	acid shock protein	2353000 *	Unknown (0.25)	B. abortus 2308, and B. melitensis 16M	BALB/c	17664263
2	bp26	Outer membrane protein BP26	265991532 *	Periplasmic (1)	B. melitensis Rev.1	BALB/c	15246618
3	exsA	ABC transporter ATPase	3787220	Cytoplasmic Membrane (1)	2308	BALB/c	12183550
4	manA	ManA family protein	260563601 *	Cytoplasmic (0.896)	16M	BALB/c	16790778
5	manB	phosphonannomutase	1198671	Cytoplasmic (0.926)	16M	BALB/c	17664263
6	mucR	transcriptional regulatory protein MUCR	1197075	Cytoplasmic (0.896)	16M	BALB/c	21708998
7	omp25	25 kDa outer-membrane immunogenic protein precursor	1196960	Outer Membrane (1)	B. ovis 63/290, 16M	BALB/c	12151196
8	omp31	outer membrane protein Omp31	7676420	Outer Membrane (1)	Rev.1	BALB/c	15246618
9	pgk	phosphoglycerate kinase	3788256	Cytoplasmic (0.997)	2308	BALB/c, C57BL/6, 129/Sv, and IRF-1 KO	20194591
10	pgm	phosphoglucomutase	3788847	Cytoplasmic (0.896)	2308	BALB/c	14573645
11	purE	5'-phosphoribosyl-5-amino-4-imidazole carboxylase	600729 *	Unknown (0.2)	16M	BALB/c	10531243
12	purK	N5-carboxyaminoimidazole ribonucleotide synthase	20141763 *	Cytoplasmic Membrane (0.788)	16M	BALB/c	7642258

#	Gene	Protein name	Gene or protein ID	Cellular Location (Prob.)	Strain(s) mutated	Mouse Model	PMID
13	virB2	type IV secretion system protein VirB2	3827981	Cytoplasmic Membrane (0.946)	2308	BALB/c	17664263
14	vjbR	Transcriptional activator, LuxR family	1198888	Unknown (0.2)	16Mand B. abortus S19	BALB/c	21530111
15	znuA	zinc ABC transporter	3827700	Periplasmic (0.976)	2308	BALB/c	16790759

* Note: denotes a protein ID; Subcellular localization was predicted by Vaxign [46].

Table 2

Viral virgogens found in at least two different viruses

Gene	Gene or protein ID	Pathogen	Strain	PMID of article
Thymidine kinase from herpesvirus				
TK	59214 *	Equid herpesvirus 1	Ab4	8388018
TK	2703374	Herpes simplex virus type 1	Strain 17	212476
UL23	1487380	Bovine herpesvirus 1	N/A	2994602
UL23	8658565	Feline herpesvirus 1	C-27	8645090
UL23	3239035	Gallid herpesvirus 1	N91B01	12021870
UL23	2952559	Pseudorabies virus	Suid herpesvirus 1	12488067
Viral glycoproteins				
Glycoprotein D	19548971 *	Herpes simplex virus 1	KHS2	18243431
Glycoprotein E(rs)	130457 *	Classical swine fever virus	Brescia	17904607
Glycoprotein E1	325461 *	Classical swine fever virus	Brescia	19203774
Glycoprotein E1	17148717 *	Western equine encephalomyelitis virus	Ag80-646	12641414
Glycoprotein E1	798795 *	Venezuelan equine encephalitis virus	125573	7676619
Glycoprotein E2	920146	Classical swine fever virus	Eystrup	15837238
Glycoprotein E2	29611992 *	Western equine encephalomyelitis virus	71V-1658	12641414
Glycoprotein E	307634504 *	Pseudorabies virus	Suid herpesvirus 1 LXB88	12488067
Glycoprotein E	330789 *	Equid herpesvirus	Equid herpesvirus 1 Ab1	17085880
Glycoprotein G	1489813	Bovine Respiratory Syncytial Virus	ATue51908	12414977
Glycoprotein G	15422177 *	Pseudorabies virus	Suid herpesvirus 1	12488067
Glycoprotein G	1489856	Rabies virus	ERAg3m	21514343
Glycoprotein H	138315 *	Herpes simplex virus 1	strain 17	8289395
Glycoprotein I	330788 *	Equid herpesvirus	Equid herpesvirus 1 Ab1	17085880
Glycoprotein M	1478255 *	Pseudorabies virus	Suid herpesvirus 1 strain Kaplan	9292000
Viral Matrix Proteins				
M2	956528	Influenza virus	A/Puerto Rico/8/34(H1N1)	19321619
M2-2	75549952 *	Human metapneumovirus	CAN97-83	16160190
M2-2	81925031 *	Human respiratory syncytial virus	A2	12922094

* Note: denotes a protein ID

Table 3

Host genes up- or down-regulated by virmugen vaccines as compared to infection with the wild type pathogen

Pathogen	Virmugen	Host Gene	Vaccine Name	Cell or serum	PMID
Host gene down-regulated in cell/tissue infected with virmugen mutant vaccine compared to wild type					
M. gallisepticum	lpd	IgG	lpd mutant	Serum	18342996
F. tularensis	FTL0552	Ccl2	FTL0552 mutant	Lung homogenate	18613792
L. monocytogenes	actA, plbC	IFN-gamma	actA/plbC mutant	Spleen	12763691
M. tuberculosis	fadD26	IFN-gamma	fadD26 mutant	Lung homogenate	15958066
F. tularensis	FTL0552	IFN-gamma	FTL0552 mutant	Lung homogenate	18613792
M. tuberculosis	mce2	IFN-gamma	mce2 mutant	Lung	16388878
M. tuberculosis	mce3	IFN-gamma	mce3 mutant	Lung	16388878
Y. enterocolitica	sodA	IFN-gamma	sodA mutant	T cells	10496939
B. bronchiseptica	bseN, cyaA	IL-10	bseN and cyaA double mutant	Splenocytes	17452472
M. tuberculosis	fadD26	IL-4	fadD26 mutant	Lung homogenate	15958066
M. tuberculosis	mce2	IL-4	mce2 mutant	Lung	16388878
M. tuberculosis	mce3	IL-4	mce3 mutant	Lung	16388878
M. tuberculosis	sigE	IL-4	sigE mutant	Lung	20457786
F. tularensis	FTL0552	IL-6	FTL0552 mutant	Lung homogenate	18613792
M. tuberculosis	fadD26	TNF-alpha	fadD26 mutant	Lung homogenate	15958066
F. tularensis	FTL0552	TNF-alpha	FTL0552 mutant	Lung homogenate	18613792
Host gene up-regulated in cell/tissue infected with virmugen mutant vaccine compared to wild type					
M. tuberculosis	sigE	APOBEC3	sigE mutant	THP-1	18657035
M. tuberculosis	sigE	C19orf66	sigE mutant	THP-1	18657035
M. tuberculosis	sigE	CD70	sig Emutant	THP-1	18657035
M. tuberculosis	sigE	DefB3	sigE mutant	Lung	20457786
M. tuberculosis	sigE	GPR182	sigE mutant	THP-1	18657035
M. tuberculosis	sigE	HIST1H2BE	sigE mutant	THP-1	18657035
M. tuberculosis	sigE	HIST2H4A	sigE mutant	THP-1	18657035
M. tuberculosis	sigE	IFN-gamma	sigE mutant	Lung	20457786

Pathogen	Virulence	Host Gene	Vaccine Name	Cell or serum	PMID
F. tularensis	capB	IgG	capB mutant	Serum	20643859
Salmonella	ippA, ippB, msbB	IgG1	ippA/ippB/msbB mutant	Serum	17997275
F. tularensis	capB	IgG2a	capB mutant	Serum	20643859
F. tularensis	wbtA	IgG2a	wbtA mutant	Serum	17296751
F. tularensis	capB	IgM	capB mutant	Serum	20643859
M. tuberculosis	sigE	IL-10	sigE mutant	Lung	20457786
F. tularensis	FTL0552	IL-12	FTL0552 mutant	Lung homogenate	18613792
F. tularensis	galU	IL-1beta	galU mutant	THP-1	21819572
Salmonella	IppA, IppB, msbB	IL-6	ippA/ippB/msbB mutant	T cell supernatant	17997275
E. coli	rfaL	IL-6	rfaL mutant	Macrophage	19522648
M. tuberculosis	fadD26	Nos2	fadD26 mutant	Lung homogenate	15958066
M. tuberculosis	sigE	PCDHB1	sigE mutant	THP-1	18657035
M. tuberculosis	sigE	PTCD1	sigE mutant	THP-1	18657035
M. tuberculosis	sigE	STAG3	sigE mutant	THP-1	18657035
M. tuberculosis	sigE	TBX21	sigE mutant	THP-1	18657035
M. tuberculosis	sigE	TNF-alpha	sigE mutant	Lung	20457786
M. tuberculosis	sigE	TSC22D4	sigE mutant	THP-1	18657035