



Biomedical knowledge graph construction of *Sus scrofa* and its application in anti-PRRSV traditional Chinese medicine discovery

Mingyang Cui^{1†}, Zhigang Hao^{2†}, Yanguang Liu¹, Bomin Lv¹, Hongyu Zhang¹, Yuan Quan^{1*} and Li Qin^{2*}

Abstract

As a new data management paradigm, knowledge graphs can integrate multiple data sources and achieve quick responses, reasoning and better predictions in drug discovery. Characterized by powerful contagion and a high rate of morbidity and mortality, porcine reproductive and respiratory syndrome (PRRS) is a common infectious disease in the global swine industry that causes economically great losses. Traditional Chinese medicine (TCM) has advantages in low adverse effects and a relatively affordable cost of application, and TCM is therefore conceived as a possibility to treat PRRS under the current circumstance that there is a lack of safe and effective approaches. Here, we constructed a knowledge graph containing common biomedical data from humans and *Sus Scrofa* as well as information from thousands of TCMs. Subsequently, we validated the effectiveness of the *Sus Scrofa* knowledge graph by the t-SNE algorithm and selected the optimal model (i.e., transR) from six typical models, namely, transE, transR, DistMult, ComplEx, RESCAL and RotatE, according to five indicators, namely, MRR, MR, HITS@1, HITS@3 and HITS@10. Based on embedding vectors trained by the optimal model, anti-PRRSV TCMs were predicted by two paths, namely, VHC-Herb and VHPC-Herb, and potential anti-PRRSV TCMs were identified by retrieving the HERB database according to the pharmacological properties corresponding to symptoms of PRRS. Ultimately, Dan Shen's (*Salvia miltiorrhiza Bunge*) capacity to resist PRRSV infection was validated by a cell experiment in which the inhibition rate of PRRSV exceeded 90% when the concentrations of Dan Shen extract were 0.004, 0.008, 0.016 and 0.032 mg/mL. In summary, this is the first report on the *Sus Scrofa* knowledge graph including TCM information, and our study reflects the important application values of deep learning on graphs in the swine industry as well as providing accessible TCM resources for PRRS.

Keywords Knowledge graph, Porcine reproductive and respiratory syndrome, Traditional Chinese medicine, Biomedical data, Deep learning

Introduction

Currently, profiting from the development of modern advanced biotechnology and a great reduction in its use cost, a massive amount of biological data is generated at an exponential rate, so it is an important issue for biologists to store and integrate heterogeneous biomedical data and extract helpful information. The application of knowledge graphs (KGs) to the biological domain can meet the above demand to some degree and have a positive influence on the swine industry. In 2012, Google officially proposed the conception of KG, whose original

[†]Mingyang Cui and Zhigang Hao are the co-first authors of this article.

*Correspondence:

Yuan Quan

quanyuan@mail.hzau.edu.cn

Li Qin

qinli@mail.hzau.edu.cn

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

purpose was aimed at optimization of search results, improvement of search quality and better user experience (Singhal 2012). As a heterogeneous data presentation, KG consists of varieties of entities (nodes) and corresponding edges (relations), and it is represented as "head entity-relation-tail entity" triplets that facilitate the integration of multiple data sources (Hogan et al. 2021). Due to the storage of considerable data and differences between relations, KGs can fulfill quick responses and reasoning as well as perform better predictions (Bonner et al. 2022). KG is developing rapidly with great application potential in agricultural biological recommender systems, information retrieval, and human-computer interaction (Chen et al. 2020; MacLean 2021). For instance, AgroLD, a knowledge graph for plant sciences established by the University of Montpellier, integrated over one hundred datasets from 15 data sources and contained 900 million triples, and its target is to provide a domain-specific knowledge platform to solve complex biological problems about the implication of genes in multiple aspects (Larmande et al. 2022). Against the backdrop of swine health, nodes can denote critical elements such as genes, proteins, and chemicals in a KG where edges capture different categories of associations between nodes. However, there is no standardized KG integrating biomedical data related to swine health, which hinders the informatization and intelligence of this field to a certain extent.

With the increase in number and the expansion of scale in swine production, health problems gradually become prominent. Porcine reproductive and respiratory syndrome (PRRS) is one of the most important diseases threatening the prosperity of the global swine industry. Characterized by powerful contagion and a high rate of morbidity and mortality, PRRS is a common viral infectious disease that was first discovered in the United States in 1987 and subsequently in Europe and Asia in the early 1990s (Keffaber 1989; Wensvoort et al. 1991; Baron et al. 1992; Chang et al. 1993; Kuwahara et al. 1994; Cho and Dee 2006). It occurs in pigs of all ages and breeds, and the clinical manifestations of PRRS mainly include pyrexia, diarrhea, tachypnea, dyspnea, cough, lethargy, anorexia, and reduced growth performance along with secondary bacterial infections (Cho and Dee 2006; Karniychuk et al. 2010), resulting in reproductive disorders or failure in sows, lowered semen quality in boars, and respiratory disease in piglets (Wang et al. 2022). Porcine reproductive and respiratory syndrome virus (PRRSV) is the etiological agent of PRRS. Belonging to the family *Arteriviridae* of the order *Nidovirales*, PRRSV is a single-stranded and positive-sense RNA virus that includes an envelope and a genome of 15.4 kb in size (Snijder et al. 2013). There are two kinds of isolates in terms of their genotypes: European genotype isolate (EU; type I) and North American

genotype isolate (NA; type II) (Yin et al. 2021). Because of the extremely frequent mutation and recombination of PRRSV, the goal of controlling and eliminating it is difficult to achieve (Sha et al. 2022). PRRS caused by PRRSV therefore exerts a highly serious influence on the benefits of the global swine industry and leads to economically great losses, restricting sustainable development of this field (Cui et al. 2021). Obviously, it is a high priority to search for effective approaches to relieve and even cure PRRS.

The measure for prophylaxis and treatment of swine diseases, such as PRRS, is to make full use of antibiotics and chemical synthetic drugs (Tantituvanont et al. 2009; Li et al. 2017). However, overuse of antibiotics for a long time leads to a series of negative problems, such as drug resistance in pathogenic microorganisms and drug residues in pork products, which may threaten public health (Holman and Chénier 2015; Mencia-Ares et al. 2021; Soares et al. 2022). Although chemical synthetic drugs function quickly for disease, their frequent use also has adverse or side effects (Karimi et al. 2015). In addition, vaccination, including the injection of attenuated and inactivated vaccines, has become one of the main strategies to prevent and treat PRRS, whereas it cannot constrain the occurrence and spread of PRRSV and offers effective protection because of the high rate of mutation and recombination of PRRSV (Dokland 2010; Zhang et al. 2022).

With a characteristic and unique theoretical architecture and sufficient clinical experience, Traditional Chinese Medicine (TCM) has a long history in disease prevention and treatment, and it is regarded as a valuable medicinal resource known to China and gradually to the world (Gong et al. 2014; Cui et al. 2021). Studies have shown that, with the advantages of low toxicity for creatures, few drug residues in pork, and little drug resistance in pathogenic microorganisms, TCMs possess multiple pharmacological activities, such as anti-inflammatory, antipyretic and antiviral activities, and some TCMs make a certain positive difference in animal production performance (Gong et al. 2014; Abdallah et al. 2019). Simultaneously, TCMs have relatively affordable applied costs, accessible resources and convenience of use (Hsu and Chung 2012). The above results show that TCMs provide abundant and beneficial drug resources for the prevention and treatment of swine diseases such as PRRS and have brilliant development prospects in the future.

In this study, to compensate for the vacancy of KGs in swine health, we constructed a *Sus scrofa* knowledge graph that not only includes conventional swine biomedical data but also integrates thousands of TCMs. After effectiveness validation of the *Sus Scrofa* knowledge graph, the optimal embedding model was selected from six classical models,

and its embedding vectors were used to predict anti-PRRSV TCMs. By reviewing the HERB database, potential TCMs were identified based on their pharmacological activity corresponding to symptoms of PRRS. Finally, Dan Shen (*Salvia miltiorrhiza Bunge*) was chosen to experimentally validate its capacity to resist PRRSV infection. In summary, our research built a *Sus scrofa* knowledge graph, which is the first report on swine health KGs with TCM information, and successfully predicted anti-PRRSV TCMs, reflecting the application value of deep learning on graphs in the global swine industry and providing drug resources for swine health problems such as PRRS.

Results

Construction of *Sus Scrofa* knowledge graph

In this study, we first constructed a *Sus scrofa* knowledge graph that integrates and normalizes swine health information from six existing biomedical databases, including the Ensembl database, KEGG database, STITCH database, STRING database, TCMID database and HVIDB database. The *Sus Scrofa* knowledge graph is a comprehensive biological knowledge graph that not only includes various kinds of human biological data and swine biomedical information involving genes, proteins, biological processes, and chemicals but also adds to the gene and protein homologous relations between humans and swine as well as the interaction between human proteins and virus proteins. In addition, it is important and characteristic for *Sus Scrofa* knowledge graph to contain thousands of TCMs and their ingredients. The *Sus Scrofa* knowledge graph includes 393,840 entities of 9 entity types and 4,631,878 relations of 16 relation types. Among 393,840 entities, there are 5,872 TCM entities, 233,586 chemical entities, 2092 virus protein entities, 89,468 human-relevant entities and 62,822 swine-related entities. A total of 89,468 human-relevant entities were composed of 73,231 proteins, 347 pathways and 15,890 genes. A total of 62,822 swine-related entities comprised 46,064 proteins, 343

pathways and 16,415 genes. In terms of 16 relation types, there are 5 relation types of interaction, including 752,626 protein-chemical associations in humans, 892,290 protein-protein associations in humans, 149,532 human protein-virus protein associations, 1,789,600 protein-chemical associations in swine, and 825,275 protein-protein associations in swine. Sixteen relation types also had 6 relation types for pathways, including 33,280 pathway-gene associations, 3,715 pathway-compound associations, and 1,663 pathway-pathway associations in humans and 33,292 pathway-gene associations, 3,698 pathway-compound associations, and 1,641 pathway-pathway associations in swine. The rest of the relation types and their numbers are 25,818 gene-protein associations in humans, 31,895 gene-protein associations in swine, 12,293 human gene-swine gene associations (homologous relations), 45,270 human protein-swine protein associations (homologous relations) and 29,990 herb-compound associations.

Validation of *Sus Scrofa* knowledge graph

Subsequently, six knowledge graph embedding models, including TransE, TransR, RESCAL, DistMult, ComplEx and RotatE, were used to acquire vector representations or embeddings of entities and relations in the *Sus Scrofa* knowledge graph. The T-distributed stochastic neighbor embedding (t-SNE) algorithm, a dimensionality reduction method, can alter high-dimensional vectors to graphical representations in a two-dimensional space. For t-SNE, adjacent nodes possess similar embedding representations. The t-SNE algorithm was adopted to visualize the effectiveness of entities and relations embedding representations that were learned by these six models. Figures 1, 2, 3, 4, 5 and 6 suggest that the embedding vectors of the identical type of entities and relations in the *Sus Scrofa* knowledge graph have comparably good colocalization in a two-dimensional graph, reflecting the effectiveness of the *Sus Scrofa* knowledge graph to some degree.

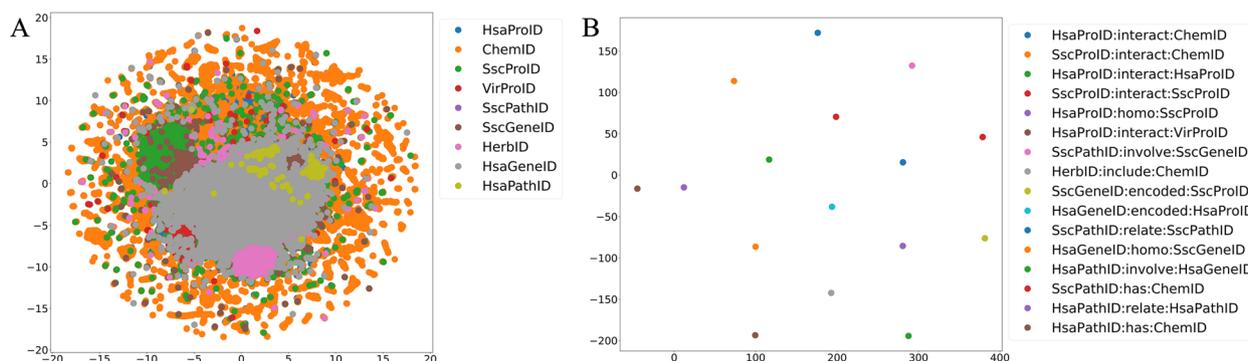


Fig. 1 The embedding vector visualization of the TransE model based on the t-SNE algorithm. **A** The embedding visualization of entities in *Sus Scrofa* KG; **B** The embedding visualization of relations in *Sus Scrofa* KG

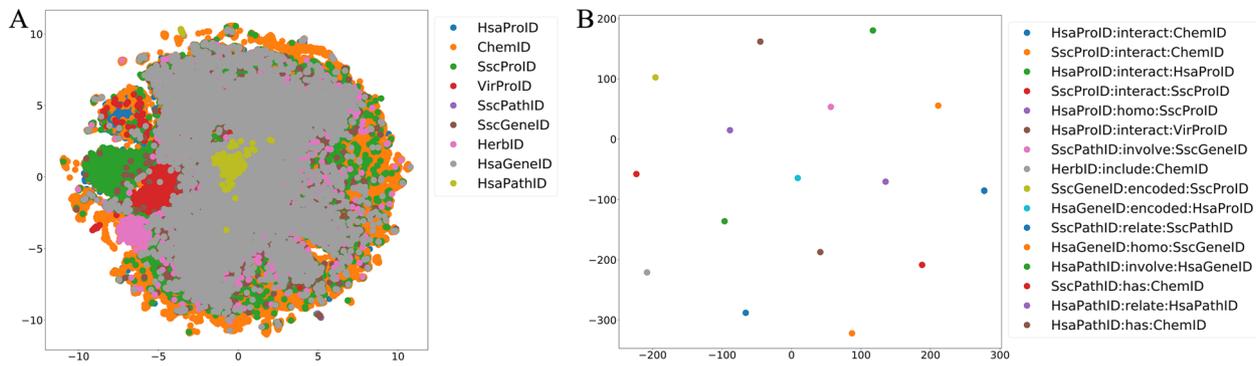


Fig. 2 The embedding vector visualization of the TransR model based on the t-SNE algorithm. **A** The embedding visualization of entities in *Sus Scrofa* KG; **B** The embedding visualization of relations in *Sus Scrofa* KG

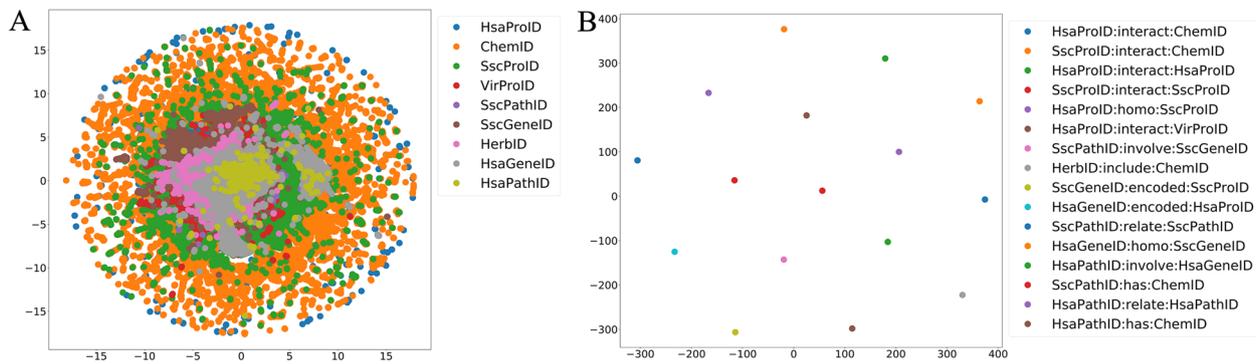


Fig. 3 The embedding vector visualization of the DistMult model based on the t-SNE algorithm. **A** The embedding visualization of entities in *Sus Scrofa* KG; **B** The embedding visualization of relations in *Sus Scrofa* KG

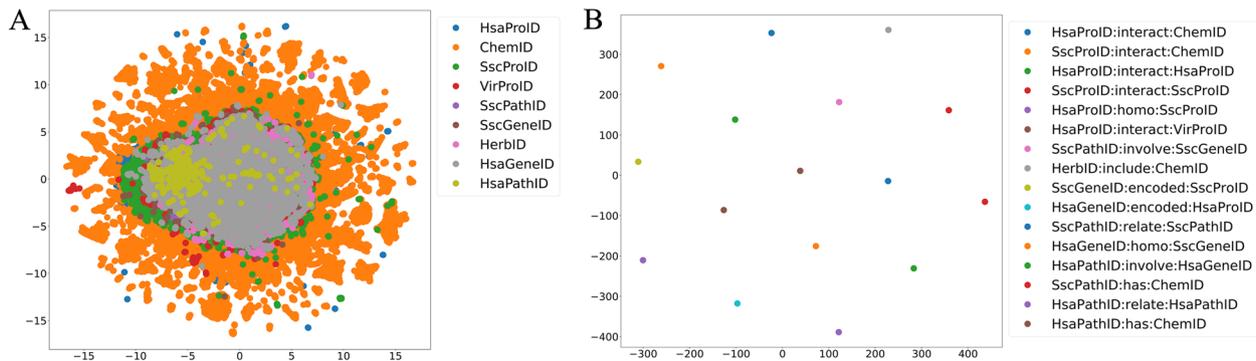


Fig. 4 The embedding vector visualization of the RESCAL model based on the t-SNE algorithm. **A** The embedding visualization of entities in *Sus Scrofa* KG; **B** The embedding visualization of relations in *Sus Scrofa* KG

Selection of the optimal model and its embeddings

For six classical knowledge graph embedding models, the embedding dimension of entities and relations were all set at 400. In addition, their batch size, learning rate (lr) and round were set at 2048, 0.1 and 49, respectively, to train the

models. The optimal model was chosen from the above six models for predictions of anti-PRRSV TCMs by embedding, which was generated by the selected model. Five indicators, including MRR, MR, HITS@1, HITS@3 and HITS@10, were used to measure the model's training effects.

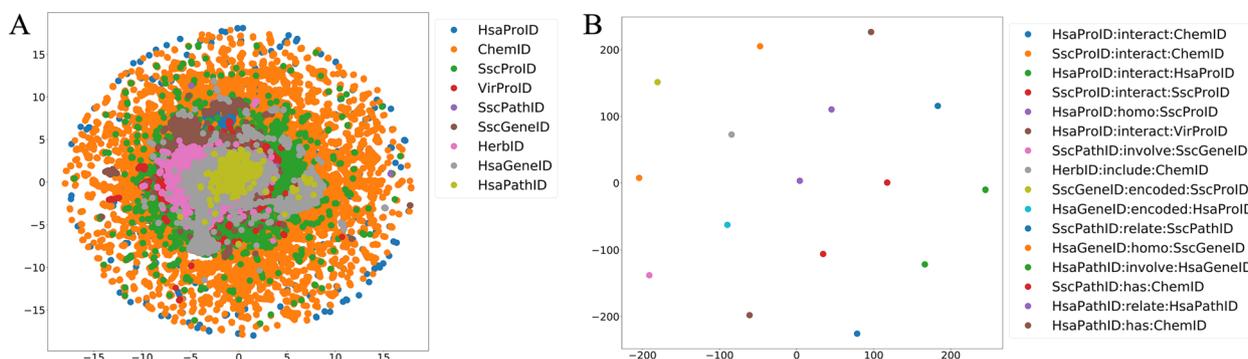


Fig. 5 The embedding vector visualization of the ComplEx model based on the t-SNE algorithm. **A** The embedding visualization of entities in *Sus Scrofa* KG; **B** The embedding visualization of relations in *Sus Scrofa* KG

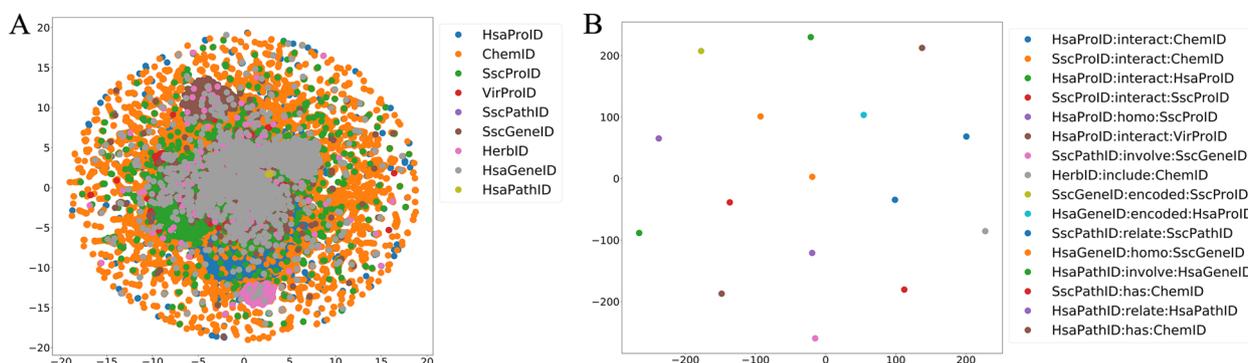


Fig. 6 The embedding vector visualization of the RotatE model based on the t-SNE algorithm. **A** The embedding visualization of entities in *Sus Scrofa* KG; **B** The embedding visualization of relations in *Sus Scrofa* KG

Table 1 The evaluation indicators of models generating embedding vectors. The maximal of every indicator is highlighted in bold

| | MRR | MR | HITS@1 | HITS@3 | HITS@10 |
|----------|-------------|--------------|-------------|-------------|-------------|
| TransE | 0.67 | 74.76 | 0.58 | 0.74 | 0.85 |
| TransR | 0.73 | 55.07 | 0.66 | 0.77 | 0.86 |
| DistMult | 0.57 | 244.70 | 0.48 | 0.61 | 0.74 |
| RESKAL | 0.70 | 135.79 | 0.64 | 0.73 | 0.83 |
| ComplEx | 0.59 | 270.98 | 0.52 | 0.62 | 0.72 |
| RotatE | 0.65 | 191.69 | 0.57 | 0.71 | 0.80 |

Table 1 shows that the MRR of transR is 0.73, which is the largest value in the six models, and the MR of transR is 55.07, which is smaller than the other values. For the HITS@n indicators, transR achieves the best values on HITS@1, HITS@3 and HITS@10, which are 0.66, 0.77 and 0.86, respectively. TransR can therefore be regarded as an optimal model, and its embeddings of entities and relations were employed to predict anti-PRRSV TCMs.

Predictions of anti-PRRSV TCMs

The predictions of anti-PRRSV TCMs were obtained by using embedding vectors of entities and relations in the *Sus Scrofa* knowledge graph, and the top 20 TCMs in the two paths of VHPC-Herb and VHC-Herb were selected to find their associations with PRRS and identify potential anti-PRRSV TCMs (Tables S3 and S4).

PRRS is also known as “blue ear disease” due to color changes in the ears of sick pigs from light pink to blue. After infection with PRRSV, sick swine are susceptible to secondary viral and bacterial diseases, and the clinical symptoms of PRRSV therefore vary from swine to swine. Specifically, the manifestations in swines of most ages and breeds cover pyrexia, diarrhea, tachypnea, dyspnea, cough, lethargy, anorexia, and reduced growth performance (Cho and Dee 2006; Karniychuk et al. 2010), finally resulting in reproductive disorders or failure in sows, lowered semen quality in boars, and respiratory disease in piglets (Wang et al. 2022).

The HERB database (referred to “BenCaoZujian” as Chinese name, <http://herb.ac.cn/>) is a high-throughput experiment- and reference-guided database of TCM that

was jointly established by Beijing University of Chinese Medicine, Institute of Computing Technology of Chinese Academy of Sciences and Institute of Nephrology, West China Hospital of Sichuan University. Information and knowledge of the most common TCMs, such as their functions, indications, clinical manifestations and therapeutic class, can be retrieved from the HERB database (Fang et al. 2021). By reviewing the HERB database, our research determined potential anti-PRRSV TCMs based on whether their pharmacological activity corresponded to the symptoms of PRRS.

In the path of VPHC-Herb, most of the top 20 TCMs were common TCMs, of which 14 TCMs were regarded as potential anti-PRRSV TCMs (Table 2). In the path of VHC-Herb, 11 TCMs can be considered as possibly effective herbs because of their pharmacological activity targeting clinical manifestations of PRRS (Table 3).

Comparing two paths, in terms of 14 TCMs in VHPC-Herb, 11 TCMs in VHC-Herb and their indications, VHPC-Herb can be perceived as a more useful path to some degree because it recognizes more quantities of promising anti-PRRSV TCMs and aims at more symptoms of PRRS. Specifically, VHPC-Herb includes homologous associations between human proteins and swine proteins from which this approach can cover more information and knowledge of swine and make connections between TCMs and swine. In addition, five TCMs, namely, *Citrus reticulata*, *Raphanus sativus*, *Lonicera japonica*, *Elsholtzia splendens* and *Bupleurum chinense*, were predicted both in VHPC-Herb and VHC-Herb, and all of them were identified by the HERB database as potential anti-PRRSV. It is generally known in the biological field that homologous proteins in different species can sometimes interact with the same compound and

Table 2 Potential anti-PRRSV effects in the VHPC-Herb path

| Rank | Pinyin name of TCM | Latin name of TCM | Indications of TCMs |
|------|--------------------|--------------------------------|--|
| 1 | Ju Pi | <i>Citrus reticulata</i> | Cough with profuse phlegm; indigestion, anorexia; diarrhea; lung qi block |
| 2 | Dan Shen | <i>Salvia miltiorrhiza</i> | Mastitis; postpartum stasis stagnation abdominal pain; heat impediment swelling and pain |
| 4 | Chai Hu | <i>Bupleurum chinense</i> | Fever and cold; cough; inflammatory |
| 5 | Jin Yin Hua | <i>Lonicera japonica</i> | Fever; throat impediment; inflammatory (viral conjunctivitis, pneumonia, mastitis); infection of upper respiratory tract |
| 6 | Lai Fu | <i>Raphanus sativus</i> | Diarrhea; indigestion, anorexia; cough |
| 7 | Ba Dou | <i>Croton tiglium</i> | Diarrhea; throat impediment |
| 8 | Hou Po | <i>Magnolia officinalis</i> | Diarrhea; food accumulation and qi stagnation; phlegm-rheum cough asthma |
| 9 | Ku Xing Ren | <i>Semen armeniacae amarum</i> | Cough, asthma |
| 12 | Yin Chen Hao | <i>Artemisia capillaris</i> | Fever; cough |
| 14 | Gou Shu Bai Pi | <i>Broussonetia papyrifera</i> | Metrorrhagia and metrostaxis |
| 15 | Xiang Ru | <i>Elsholtzia splendens</i> | Fever; abdominal pain; diarrhea |
| 16 | Bai Lan Hua | <i>Michelia alba</i> | Cough |
| 17 | Sheng Jiang | <i>Zingiber officinale</i> | Fever; diarrhea; phlegm-rheum cough asthma |
| 20 | Guang Huo Xiang | <i>Pogostemon cablin</i> | Diarrhea with abdominal pain |

Table 3 Potential anti-PRRSV effects in the VHC-Herb path

| Rank | Pinyin name of TCM | Latin name of TCM | Indications/functions of TCMs |
|------|--------------------|---------------------------------|--|
| 4 | Ju Pi | <i>Citrus reticulata</i> | Cough with profuse phlegm; indigestion, anorexia; diarrhea; lung qi block |
| 8 | Gua Lou | <i>Trichosanthes kirilowii</i> | Lung heat cough |
| 9 | Lai Fu | <i>Raphanus sativus</i> | Diarrhea; indigestion, anorexia; cough |
| 10 | Ku Shen Shi | <i>Sophora flavescens</i> | Ability to clear heat and resolve toxin |
| 12 | Ku Shen | <i>Radix Sophora flavescens</i> | Diarrhea; conjunctivitis; ability to remove heat and damp |
| 13 | Jin Yin Hua | <i>Lonicera japonica</i> | Fever; throat impediment; inflammatory (viral conjunctivitis, pneumonia, mastitis); infection of upper respiratory tract |
| 14 | Xiang Ru | <i>Elsholtzia splendens</i> | Fever; abdominal pain; diarrhea |
| 15 | Gou Qi Zi | <i>Fructus Lycii</i> | Lung disease with cough |
| 16 | Kong Shi Chun | <i>Ulva lactuca</i> | Ability to clear heat, resolve toxin and transform phlegm |
| 18 | Mu Xu | <i>Medicago sativa</i> | Febrile diseases |
| 19 | Chai Hu | <i>Bupleurum chinense</i> | Fever and cold; cough; inflammatory |

play identical roles, so such prediction results are normal phenomena.

Validation of the anti-PRRSV pharmacological activity of Dan Shen

Dan Shen, also called *Salvia miltiorrhiza Bunge* (SM) as a scientific name, belongs to the Lamiaceae family, which was first stated in “The Herbal Classic of Shen Nong” during the Eastern Han Dynasty approximately 2000 years ago. It is a popular, valuable, widely used medicinal herb in traditional Chinese medicine whose main medicinal parts are dry roots and rhizomes. In traditional use, Dan Shen possesses diverse functions, including promoting blood flow to remove blood stasis and stimulate menstrual discharge, cooling blood to ease the mind and relieve dysphoria, clearing heat to remove carbuncles and relieving pain. To the best of our knowledge, Dan Shen contains different kinds of chemical compounds, such as diterpenoid quinones, hydrophilic phenolic acids, and essential oil constituents, in which the two former compounds principally play bioactive roles. The various pharmacological activities of Dan Shen are important to point out, which is either used independently or combined with other TCMS. These pharmacological activities include anti-inflammatory, antioxidant, anti-hepatocyte injury, anti-neuropathic pain and other therapeutic effects, making Dan Shen an herbal therapy for various indications (Su et al. 2015; Jia et al. 2019). Therefore, our research chose Dan Shen to experimentally validate its therapeutic effect on PRRS and accelerate the development of drug discovery in the swine health field.

The Marc-145 cell line is vulnerable to PRRSV. Our experiment first conducted a toxicity test of Dan Shen extract in which different concentrations of Dan Shen extract were applied to the Marc-145 cell line to

determine the concentration range of Dan Shen extract under the circumstance that the cell survival rate was more than 90%. As shown in Fig. 7A and Table S5, the survival rate of the Marc-145 cell line is up to 90% at concentrations ranging from 0.004 mg/mL to 0.032 mg/mL, which suggests that Dan Shen extract has low toxicity to the Marc-145 cell line in such a concentration range and is therefore employed for follow-up study.

Subsequently, our research detected the PRRSV inhibition rate resulting from Dan Shen extract in the Marc-145 cell line at survivable concentrations. Figure 7B illustrates that the inhibition rate of PRRSV exceeded 90% when the concentrations of Dan Shen extract were 0.004, 0.008, 0.016 and 0.032 mg/mL. These results suggested that Dan Shen can take an inhibitory action to control the life cycle of PRRSV, further having the capability to resist PRRSV infection within cells.

Discussion

With the cross-integration of BT and IT in the background, multiple kinds of technology are applied to the agricultural biological production domain, and a new data management paradigm represented by KG adopts a graph data structure to model and record biological data (Hogan et al. 2021). For the swine industry, the application of KGs is beneficial for aggregating a large amount of useful knowledge and achieving more accurate object-level searches and predictions, which accelerates the mechanization and intellectualization of swine breeding and facilitates the development of swine production. However, there is currently no swine KG with knowledge of TCMS. Here, we established a *Sus Scrofa* knowledge graph that contained varieties of biomedical data in humans and swine, homologous information between

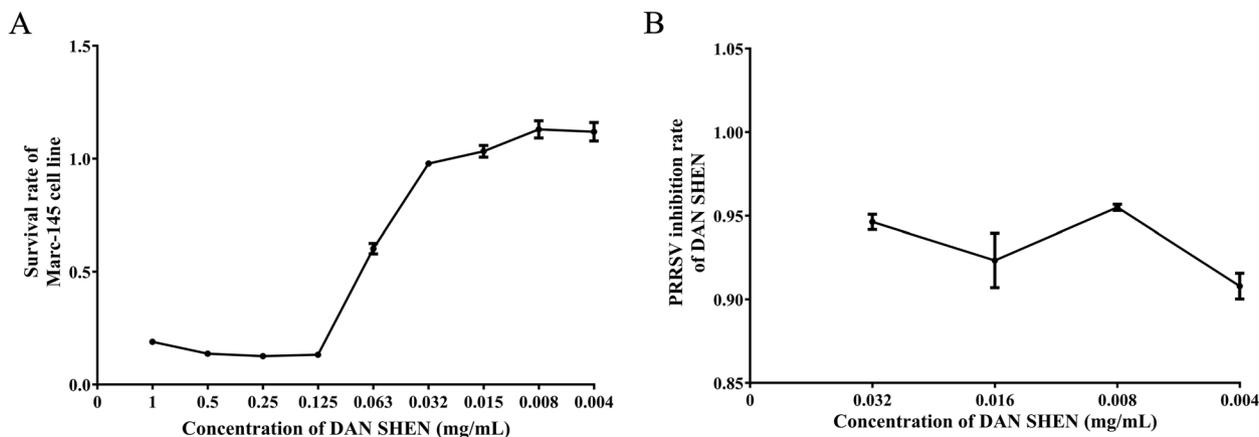


Fig. 7 Experimental results of the anti-PRRSV pharmacological activity of Dan Shen. **A** Toxicity test of Dan Shen; **B** PRRSV inhibition rate of Dan Shen

humans and swine and information on thousands of TCMs to fill this vacancy for follow-up research.

In the swine production field, PRRS, also known as 'blue ears' disease, is one of the most important diseases affecting the growth of the global swine industry, leading to great economic losses (Cui et al. 2021). PRRS is caused by PRRSV, whose high frequency rate of mutation and recombination makes it difficult for researchers to control it (Sha et al. 2022). Therefore, it is a high priority to look for a useful approach to relieve and even cure PRRS, and TCMs are generally conceived as helpful methods.

TCM develops a unique theoretical system, and it has a long history in clinical practice and experience to ameliorate the health of human beings and animals (Gong et al. 2014; Cui et al. 2021). On the one hand, TCMs are characterized by their low toxicity, few drug residues in livestock products and little drug resistance in pathogens (Gong et al. 2014; Abdallah et al. 2019). On the other hand, they have comparably low applied cost and great convenience (Hsu and Chung 2012). Therefore, TCM is a valuable drug resource library with enormous potential and good development prospects.

Based on embedding vectors of the constructed *Sus Scrofa* knowledge graph trained by the optimal model, namely, transR selected from six classical models (transE, transR, DistMult, ComplEx, RESCAL, RotatE), anti-PRRSV TCMs were predicted by the VHPC-Herb path and VHC-Herb path. Most predicted herbs are common TCMs in China, Japan, Korea and other places. By retrieving the HERB database, possibly effective anti-PRRSV TCMs were recognized according to pharmacological activity corresponding to clinical manifestations of PRRS. Fourteen potential TCMs were identified in the VHPC-Herb path, and 11 TCMs were identified in the VHC-Herb path. We also compared the predicted results of the VHPC-Herb and VHC-Herb paths and analyzed and explained the reasons. Finally, Dan Shen (rank 2 in VHPC-Herb path) was selected to validate its efficacy by experiment because of its multiple pharmacological properties, chemical constituents and traditional use, and the experiment demonstrated that Dan Shen possessed anti-PRRSV bioactive roles.

The *Sus Scrofa* knowledge graph includes biological data about humans, swine and varieties of viruses, such as genes, proteins, pathways, chemicals and their associations, as well as information on thousands of TCMs. It collates multiple types of data by graph structure and is applied to machine learning algorithms to obtain new insights into swine health research, such as drug discovery aimed at swine diseases. The two paths for predictions, namely, VHC-Herb and VHPC-Herb, maintain a balance between more information and a shorter path

length, which represent more knowledge and higher accuracy, respectively.

One limitation is that the *Sus Scrofa* knowledge graph merely integrates diverse types of TCMs and their ingredients and that it lacks dosage information. It is generally known that the effects of TCMs are not only related to their categories but also to dose and compatibility, so we suggest that researchers make use of TCMs according to the actual situation (Zha et al. 2015). Another limitation of this study is that the edges of the *Sus scrofa* knowledge graph describe rough information of associations between two entities, so the predictions of TCMs represent a certain connection with PRRSV, and we need to determine whether it is a positive or negative effect that one certain TCM exerts on the treatment of PRRS. For example, MAI JIAO, a predicted common TCM ranking 19 in the VHPC-Herb path, has the function of contracting the uterus, which may aggravate the condition of miscarriage when used in farrowing sows with PRRS. Another case in our study is QIAN NIU ZI (rank 1 in VHC-Herb path), whose therapeutic class is Drastic Purgatives, which has adverse effects on swine with diarrhea infected by PRRSV. Both MAI JIAO and QIAN NIU ZI are annotated by the HERB database. In addition, except for Dan Shen, whether a TCM can be considered a potential anti-PRRSV TCM is only based on its pharmacological property, without performing experimental verification. The efficacy of these TCMs should be explored experimentally in the future.

Conclusion

In summary, we constructed a *Sus Scrofa* knowledge graph, and based on it, we successfully predicted and determined potential anti-PRRSV TCMs where the efficacy of Dan Shen was validated experimentally. The *Sus Scrofa* knowledge graph is the first reported swine health KG, including information on thousands of TCMs. Through the application of the *Sus scrofa* knowledge graph for the prediction of TCMs, our study shows the enormous utilization potential of deep learning on graphs in the global swine industry and provides significant medicinal resources for swine diseases, which greatly promotes the development of traditional Chinese veterinary medicine (TCVM).

Methods

Figure 8 shows the workflow for construction of the *Sus Scrofa* knowledge graph and prediction of anti-PRRSV TCMs. Our workflow consists of five main steps: (1) data gathering; (2) construction of the *Sus Scrofa* knowledge

graph; (3) optimal model selection from six classical models; (4) anti-PRRSV TCM prediction; and (5) identification of potential anti-PRRSV TCMs. All steps are depicted in detail as follows.

Data sources

We obtained four types of associations from the Ensembl database (<https://www.ensembl.org>) using a powerful tool called BioMart, which performed cross-database annotations. The four types of associations included gene–protein associations in humans and swine as well as gene and protein homologous relations between humans and swine. We collected six types of associations from the Kyoto Encyclopedia of Genes and Genomes database (KEGG, <https://www.genome.jp/kegg/>) by API. These six types of associations were involved in biological pathways in humans and swine, including pathway–gene associations, pathway–compound associations and pathway–pathway associations. The interactions between proteins in both humans and swine were downloaded from the STRING database (<https://string-db.org/>), which integrates all known and predicted associations between proteins in more than 1,400 organisms, including both physical interactions and functional associations (Szklarczyk et al. 2021). Protein–chemical associations in both humans and swine came from the STITCH database (<http://stitch.embl.de/>), which contains data on drug (chemical)–target (protein) relationships and binding affinities (Kuhn et al. 2010). The Human-Virus Interaction Database (HVIDB, <http://zzdlab.com/hvidb>) is a comprehensive database for human–virus protein–protein interactions and provides online PPI prediction tools (Yang et al. 2021), and we used information on interactions between human proteins and virus proteins from HVIDB as human protein–virus protein associations in our knowledge graph. The Traditional Chinese Medicine Integrated Database (TCMID, <http://www.megabionet.org/tcmid/>) is a comprehensive database designed for TCM standardization and modernization, and it has been highly acknowledged among scholars and pharmacologists in the TCM domain (Huang et al. 2018). From TCMID, we gathered information about the chemical ingredients of TCMs as herb–compound associations in the *Sus scrofa* knowledge graph.

Data processing and construction of *Sus Scrofa* knowledge graph

Due to the limitation of BioMart, we could not obtain homologous gene information named by NCBI gene (formerly Entrezgene) ID (e.g., 4535). Therefore, we first obtained homologous gene information named Gene stable ID (e.g., ENSG00000198888) and gene ID mapping between NCBI gene ID and Gene stable ID in human and swine, and then we merged the above three information files to obtain homologous gene information named NCBI gene ID. The compound of pathway–compound associations, herb–compound associations and protein–chemical associations are represented by KEGG compound ID (e.g., C00001), compound chemical name (e.g., parthenolide) and STITCH compound ID (e.g., CIDm00010457), respectively. To align the entities of compounds from different sources, we conducted ID mapping by aliases files from the STITCH database and finally transformed the other two representations into STITCH compound ID. The human proteins are described by protein stable ID, and we used the ID mapping tool provided by the UniProt database (<https://www.uniprot.org/>) to convert UniProt protein ID (e.g., P62277) in human protein–virus protein associations to protein stable ID (e.g., ENSP00000435777). In addition, according to binding affinities, the interaction between proteins and protein–chemical associations in human and swine were further filtered, and their combined scores over 400 were selected and used to ensure the effectiveness of interactions.

Other entity types without processing have unique representations when gathered from the database, whose nomenclature in the *Sus Scrofa* knowledge graph used ID or name from the corresponding databases. Entity and relation representations are shown in Tables S1 and S2, respectively.

Then, we organized the above data into "head–entity–relation–tail entity" triplets and stored them in a 'tsv' format file, completing the construction of the *Sus scrofa* knowledge graph.

Embedding generation of KGs and evaluation of training models

A deep graph library–knowledge graph (DGL-KE) is an open-source package developed by Amazon Web

(See figure on next page.)

Fig. 8 Workflow of the construction of the *Sus scrofa* knowledge graph and the prediction of anti-PRRSV TCM. In this study, we collected and integrated varieties of biomedical data from humans and swine as well as TCM information to construct a *Sus scrofa* knowledge graph. Then, six models were adopted to generate embedding vectors of the *Sus scrofa* knowledge graph from which the optimal embedding model was selected according to evaluation indicators, and its embedding vectors were used to predict anti-PRRSV TCMs. By reviewing the HERB database, potential anti-PRRSV TCMs were identified based on their pharmacological activity corresponding to symptoms of PRRS

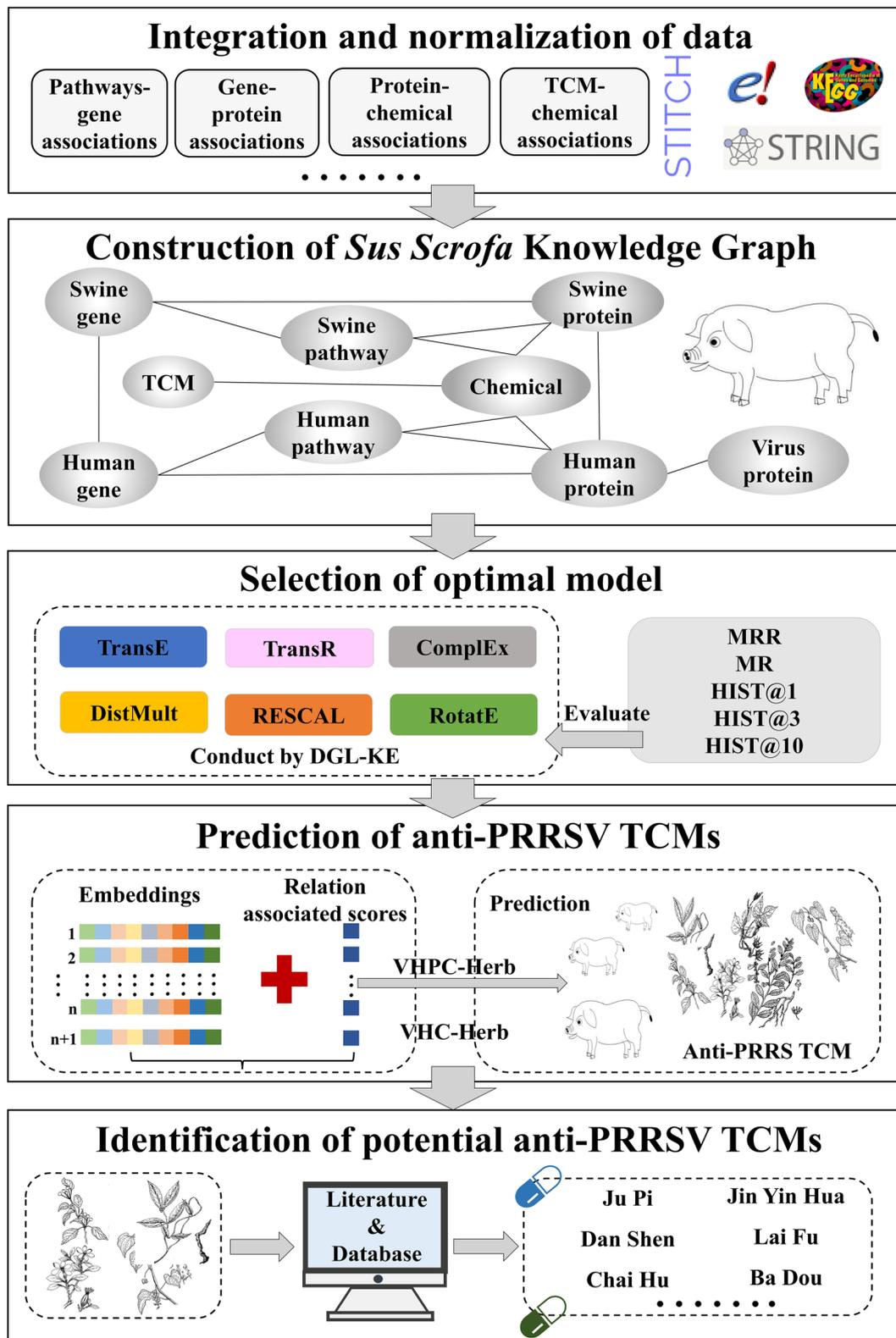


Fig. 8 (See legend on previous page.)

Services AI Shanghai Lablet for efficient computation of knowledge graph embeddings, which can accelerate training on large-scale knowledge graphs by using multiprocessing, multi-GPU, and distributed parallelism and ultimately achieve high operation efficiency. Furthermore, DGL-KE is extensible and uncomplicated to use, and it covers a series of classical models, including TransE, TransR, DistMult, ComplEx, RESCAL and RotatE (Zheng et al. 2020).

TransE: TransE is a well-known knowledge graph representation learning method used to map entities and relations to low-dimensional vector representations. It is based on the fundamental assumption that relations in a knowledge graph can be interpreted as translations from one entity vector to another. TransE learns vector representations by minimizing the distance between the translated entity vectors and the actual relation vectors. Specifically, given a triplet (h, r, t) where h is the head entity, r is the relation, and t is the tail entity, TransE represents this triplet in vector form. It predicts the tail entity vector by adding the head entity vector and the relation vector, i.e., $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$. For each triple (h, r, t) , TransE defines the scoring function as follows:

$$f_r(h, t) = - \|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_{1/2} \quad (1)$$

where $\|\cdot\|_{1/2}$ represents the L_1 or L_2 norm, $\mathbf{h} \in \mathbb{R}^d$ and $\mathbf{t} \in \mathbb{R}^d$ represent the vector representations of the head and tail entities, respectively, and $\mathbf{r} \in \mathbb{R}^d$ represents the vector representation of the relation (Bordes et al. 2013).

TransR: In TransE, entities and relations are represented as vectors in the same embedding space. It does not explicitly handle hierarchical relations because it assumes that all relations are treated equally, and the model focuses on capturing the translation patterns between entities. In contrast, TransR introduces a separate relation-specific projection matrix for each relation. Entities are represented in an entity space, while relations are represented in a relation space. The projection matrix maps the entity embeddings from the entity space to the relation space. This allows TransR to capture the semantic properties of hierarchical relations more effectively. Given a triplet (h, r, t) , the scoring function for TransR is defined as follows:

$$f_r(h, t) = - \|\mathbf{h}_\perp + \mathbf{r} - \mathbf{t}_\perp\|_2^2 \quad (2)$$

where $\mathbf{h}_\perp = \mathbf{M}_r \mathbf{h}$, $\mathbf{t}_\perp = \mathbf{M}_r \mathbf{t}$, and $\mathbf{M}_r \in \mathbb{R}^{k \times d}$ is the projection matrix from the entity space to the relation space (Lin et al. 2015).

DisMult: DistMult is a popular knowledge graph representation learning model. It models the relations between entities using a symmetric bilinear function. In DistMult, each entity and relation are represented as embedding

vectors in a continuous vector space. The model predicts the likelihood of a triple by computing the score as the dot product of the embeddings:

$$f_r(h, t) = \mathbf{h}^T \cdot \text{diag}(\mathbf{r}) \cdot \mathbf{t} \quad (3)$$

Here, $\text{diag}(\mathbf{r})$ represents a diagonal matrix with the elements of the relation vector \mathbf{r} on its diagonal (Yang et al. 2015).

ComplEx: ComplEx is a model based on complex-valued vectors that extends DistMult by introducing complex operations. It represents entities and relationships using complex-valued vectors and predicts the score of a triple by computing the Hermitian product (complex conjugate multiplication) between the vectors. More specifically, ComplEx performs multiplication operations on the complex-valued vectors of the head entity, relation, and tail entity and takes the real part as the predicted score:

$$f_r(h, t) = \text{Re}(\mathbf{h}^T \cdot \text{diag}(\mathbf{r}) \cdot \bar{\mathbf{t}}) \quad (4)$$

where $\mathbf{h} \in \mathbb{C}^d$ and $\mathbf{t} \in \mathbb{C}^d$ represent the vector representations of the head and tail entities, respectively, while $\mathbf{r} \in \mathbb{C}^d$ represents the vector representation of the relation. $\bar{\mathbf{t}}$ is the conjugate of \mathbf{t} , and $\text{Re}(\cdot)$ represents taking the real part of a complex number.

Therefore, ComplEx has the ability to better capture the symmetric and anti-symmetric relationships in knowledge graphs compared to DistMult. It is also advantageous in modeling multiple types of relationships accurately due to the multidimensional representation power of complex-valued vectors. However, ComplEx requires more computational resources and training time than DistMult (Trouillon et al. 2016).

RESCAL: RESCAL is based on tensor decomposition and aims to capture complex interactions between entities and relationships. Specifically, RESCAL represents the knowledge graph as a three-dimensional tensor, where each element corresponds to the strength of the relationship between entity pairs. By decomposing this tensor, embeddings of entities and relationships can be obtained. In RESCAL, the knowledge graph (KG) is formed by a large tensor X , where if a triplet exists in the KG, X_{htr} is denoted as 1; otherwise, it is 0. The score function is defined as:

$$f_r(h, t) = \mathbf{h}^T \mathbf{M}_r \mathbf{t} = \sum_{i=0}^{d-1} \sum_{j=0}^{d-1} [\mathbf{M}_r]_{ij} \cdot [\mathbf{h}]_i \cdot [\mathbf{t}]_j \quad (5)$$

where $\mathbf{M}_r \in \mathbb{R}^{d \times d}$, $[\mathbf{M}_r]_{ij}$ represents the (i, j) -th element of matrix \mathbf{M}_r , and $[\mathbf{h}]_i$ and $[\mathbf{t}]_j$ represent the i -th and j -th components of vectors \mathbf{h} and \mathbf{t} , respectively. The objective is to decompose X into entity embeddings and relationship embeddings to make X_{htr} close to $\mathbf{h}^T \mathbf{M}_r \mathbf{t}$ (Nickel et al. 2011).

RotatE: In RotatE, each relation is represented as a complex-valued rotation in a multidimensional space. The head and tail entities are also represented as vectors. By applying elementwise rotating operations on the head entity vector, the relationship vector, and the complex conjugate of the tail entity vector, RotatE measures the plausibility of a triplet in the knowledge graph. Given a triplet (h, r, t) , where we expect $t = h \circ r$ and the magnitude $|r_i| = 1$. Therefore, for each dimension in the complex space, the expectation is that $t_i = h_i r_i$, where t_i, h_i , and $r_i \in \mathbb{C}$. Its score function is defined as:

$$f_r(h, t) = \| \mathbf{h} \circ \mathbf{r} - \mathbf{t} \| \tag{6}$$

where $\mathbf{t} \in \mathbb{C}^d$, $\mathbf{h} \in \mathbb{C}^d$ and $\mathbf{r} \in \mathbb{C}^d$ are the embeddings (Sun et al. 2019).

In this study, the DGL-KE package was adopted to generate low-dimensional embedding vector representations of entities and relations in the *Sus Scrofa* knowledge graph. Before training, the triples were split into training/valid/test sets at a 9:0.5:0.5 ratio. Then, we conducted the abovementioned six models over the *Sus Scrofa* knowledge graph to obtain embedding vectors by using DGL-KE.

General indicators to evaluate the performance of knowledge graph embedding models included MRR, MR, HITS@1, HITS@3 and HITS@10. Based on these indicators, six models, namely, transE, transR, DistMult, ComplEx and RotatE, were evaluated, and the optimal model was selected to predict anti-PRRSV TCMs by its generated embedding vectors.

Anti-PRRSV TCM prediction

Subsequently, our research predicted anti-PRRSV TCMs, which was achieved by the prediction of the interactions between the proteins of PRRSV and TCMs. Due to the absence of the above interactions in the *Sus Scrofa* knowledge graph, two paths, called VHC-Herb and VHPC-Herb, were used to handle this problem and

perform prediction. Specifically, VHC-Herb started from the proteins of PRRSV, passed through human proteins and chemicals successively, and finally reached TCMs. The routine included viral protein-human protein associations, human protein-chemical interactions, and TCM-chemical relationships (Fig. 9A). VHPC-Herb started from the proteins of PRRSV, passed through human proteins, swine proteins and chemicals successively, and eventually arrived at TCMs. The routine included viral protein-human protein associations, human protein-swine protein homologous relationships, swine protein-chemical interactions, and TCM-chemical relationships (Fig. 9B). In fact, the prediction of anti-PRRSV TCMs can be perceived as a knowledge graph completion problem. This problem can be represented as a ranking task, which is essentially the task of learning a prediction function that scores high on true triplets and low on false triplets. In this study, transR was considered the optimal model according to evaluation indicators, and the prediction method of transR was therefore introduced in this section. The edge scores of transR were calculated using the following algorithm:

$$\mathbf{d} = \gamma - \| \mathbf{h}_\perp + \mathbf{r} - \mathbf{t}_\perp \|_2^2 \tag{7}$$

$$\text{score} = \text{logSigmoid}(\mathbf{d}) = \text{log} \left(\frac{1}{1 + \exp(-\mathbf{d})} \right) \tag{8}$$

where γ is the margin separating the positive and negative triplets and $\| \cdot \|_2^2$ represents the L2 norm. Note that here, LogSigmoid is used to make all scores less than 0, so the larger a score is, the stronger the corresponding association between entities is.

In the process of prediction in both the VHC-Herb and VHPC-Herb paths, we sorted the edge scores in descending order and chose the top 20 entities for the next step. Finally, the top 20 TCMs in the two paths were chosen to determine whether they were possibly effective anti-PRRSV TCMs. By reviewing the HERB database, a

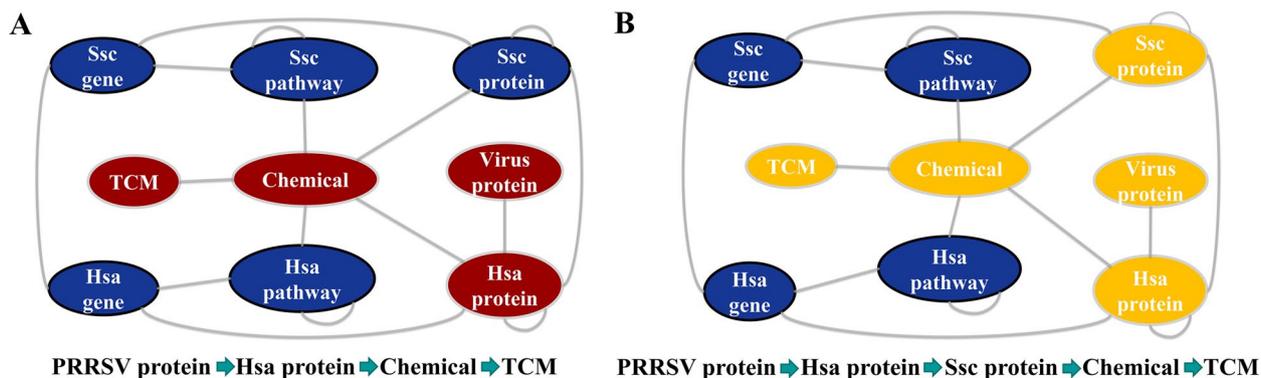


Fig. 9 Two paths predicting anti-PRRSV TCMs. **A** VHC-Herb path; **B** VHPC-Herb path

high-throughput experiment- and reference-guided database of TCM, potential TCMs were identified by their pharmacological activity corresponding with symptoms of PRRS.

Preparation of Dan Shen extract

Dan Shen was purchased from JinYeZi Pharmaceutical Co. Ltd. (Hebei, China).

First, we weighed 20 g of Dan Shen and crushed it into a dry powder, which was then soaked for 12 h in 10 volumes of absolute ethanol. Next, Dan Shen and its soaking solution, whose total volume was 200 mL, were heated and refluxed for 2 h, and we recovered the filtrate. After that, we added 200 mL of absolute ethanol to the Dan Shen residue for secondary heating and reflux and recovered the filtrate again. Finally, after heating and refluxing twice, the two filtrates were mixed, and rotary evaporation was conducted until powder of the Dan Shen extract was obtained.

Toxicity test of Dan Shen extract on the Marc-145 cell line

Marc-145 cells, provided by the School of Animal Science and Technology, School of Animal Medicine, Huazhong Agricultural University, were maintained in DMEM supplemented with 10% fetal bovine serum (FBS) and 1% penicillin–streptomycin. PRRSV strain CH-1a was obtained from the School of Animal Science and Technology, School of Animal Medicine, Huazhong Agricultural University, and amplified in Marc-145 cells. The viral titer was measured by a plaque assay. The steps of the toxicity test are as follows:

(1) Seeding of Marc-145 cells We seeded Marc-145 cells into cell suspensions using DMEM containing 10% FBS and seeded 10,000 Marc-145 cells per well into a 96-well cell culture plate, with a seeding volume of 100 μ L per well. **(2) Marc-145 cells** cultured after cell seeding were cultivated in a cell culture incubator with 37°C and 5% carbon dioxide for 24 h. **(3) The Dan Shen extract** was dissolved in DMSO and diluted gradiently in DMEM containing 2% FBS to final concentrations of 1, 0.5, 0.25, 0.125, 0.0625, 0.032, 0.015, 0.008 and 0.004 mg/mL, and then we added prepared media with different concentrations of Dan Shen extract into the 96-well cell culture plate. Simultaneously, we regarded wells with DMEM containing 2% FBS as the control group and wells without cells and Dan Shen extract as the blank group. Every group was repeated in three wells, and they were put in a cell culture incubator with 37°C and 5% carbon dioxide for 48 h. **(4) Measurement of absorbance** We added 10 μ L of CCK-8 solution to every well (be careful not to generate air bubbles) and set the 96-well cell culture plate in an incubator for 1–4 h, and then we measured the absorbance

at 450 nm by a microplate reader. **(5) Calculation of the cell survival rate** According to absorbance, the Marc-145 cell survival rate in different concentrations of Dan Shen extract was determined by the following formula:

$$CSR = \frac{E - B}{C - B} \times 100\% \quad (9)$$

where CSR represents the Marc-145 cell survival rate, and E, C, and B denote the optical density (OD) values of the experimental groups, control groups and blank groups, respectively.

Detection of the PRRSV inhibition rate of Dan Shen extract Steps **(1)**, **(2)** and **(4)** are identical to steps **(1)**, **(2)** and **(4)** in the toxicity test.

(3) Addition of Dan Shen extract and inoculation of PRRSV Dan Shen extract was dissolved in DMSO and diluted gradiently in DMEM containing 2% FBS and 100 \times TCID₅₀ PRRSV to final concentrations of 0.032, 0.015, 0.008 and 0.004 mg/mL, respectively, and then we added the mentioned prepared media into the 96-well cell culture plate. Simultaneously, we regarded wells with DMEM containing 2% FBS and PRRSV but no Dan Shen extract as the virus control group and wells with DMEM containing only 2% FBS as the cell control group. Every group was repeated in three wells, and they were cultivated in a cell culture incubator with 37°C and 5% carbon dioxide for 48 h. **(5) Calculation of the PRRSV inhibition rate of Dan Shen** According to absorbance, the PRRSV inhibition rate in different concentrations of Dan Shen extract was determined by the following formula:

$$IR = \frac{T - V}{C - V} \times 100\% \quad (10)$$

where IR represents the PRRSV inhibition rate, and T, C, and V denote the optical density (OD) values of the Dan Shen treatment groups, cell control groups and virus control groups, respectively.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s44149-023-00106-7>.

Additional file 1.

Acknowledgments

Not applicable.

Authors' contributions

M.C conceived and designed research, collected and analyzed data, and wrote the manuscript, Z.H analyzed data and wrote the manuscript, Y.L collected data, B.L performed the experiments, H.Z designed research, Y.Q designed research and participated as a project administrator, L.Q designed research, provided funding support for this paper and participated as a project administrator. All authors have read and agreed to the published version of the manuscript.

Funding

This study was supported by the China Fundamental Research Funds for the Central Universities (No. 2662022XXYJ001, 2662022JC004, 2662023XXPY005).

Availability of data and materials

The data used to support the findings of this study are included within the article.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The author declares that they have no competing interests.

Author details

¹Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, Huazhong Agricultural University, Wuhan 430070, P. R. China. ²Hubei Engineering Technology Research Center of Agricultural Big Data, College of Informatics, Huazhong Agricultural University, Wuhan 430070, P. R. China.

Received: 15 September 2023 Accepted: 22 November 2023

Published online: 12 January 2024

Reference

- Abdallah, A., P. Zhang, Q.Z. Zhong, and Z.W. Sun. 2019. Application of traditional Chinese herbal medicine byproducts as dietary feed supplements and antibiotic replacements in animal production. *Current Drug Metabolism* 20 (1): 54–64. <https://doi.org/10.2174/1389200219666180523102920>.
- Baron, T., E. Albina, Y. Leforban, F. Madec, H. Guilamoto, J. Plana Duran, and P. Vannier. 1992. Report on the first outbreaks of the porcine reproductive and respiratory syndrome (PRRS) in France. Diagnosis and viral isolation. *Annales de recherches vétérinaires* 23 (2): 161–166.
- Bonner, S., I.P. Barrett, C. Ye, R. Swiers, O. Engkvist, A. Bender, C.T. Hoyt, and W.L. Hamilton. 2022. A review of biomedical datasets relating to drug discovery: a knowledge graph perspective. *Briefings in Bioinformatics* 23 (6): bbac404. <https://doi.org/10.5555/2999792.2999923>.
- Bordes A, N. Usunier, A. García-Durán, J. Weston, O. Yakhnenko. 2013. Translating embeddings for modeling multirelational data. In: Proc. Of the 26th Int'l Conf. on Neural Information Processing Systems (NIPS). Lake Tahoe. 2:2787–2795. <https://doi.org/10.5555/2999792.2999923>.
- Chang, C.C., W.B. Chung, M.W. Lin, C.N. Weng, P.C. Yang, Y.T. Chiu, and R.M. Chu. 1993. Porcine reproductive and respiratory syndrome (PRRS) in Taiwan. I. Viral isolation. *Journal of the Chinese Society of Veterinary Science* 19: 268–276.
- Chen, X.J., S.B. Jia, and Y. Xiang. 2020. A review: Knowledge reasoning over knowledge graph. *Expert Systems with Applications* 141: 112948. <https://doi.org/10.1016/j.eswa.2019.112948>.
- Cho, J.G., and S.A. Dee. 2006. Porcine reproductive and respiratory syndrome virus. *Theriogenology* 66 (3): 655–662. <https://doi.org/10.1016/j.theriogenology.2006.04.024>.
- Cui, W.Q., F. Yu, Y.F. Zhang, X. Han, R.F. Zou, Y.D. Tang, L.G. Wang, N. Eliphaz, J. Wang, S.G. Yuan, X.H. Cai, and Y.H. Li. 2021. Discovery of traditional Chinese medicines against porcine reproductive and respiratory syndrome virus. *Pharmacological Research - Modern Chinese Medicine* 1: 100003. <https://doi.org/10.1016/j.prmcm.2021.100003>.
- Dokland, T. 2010. The structural biology of PRRSV. *Virus Research* 154 (1–2): 86–97. <https://doi.org/10.1016/j.virusres.2010.07.029>.
- Fang, S.S., L. Dong, L. Liu, J.C. Guo, L.H. Zhao, J.Y. Zhang, D.C. Bu, X.K. Liu, P.P. Huo, W.C. Cao, et al. 2021. HERB: a high-throughput experiment- and reference-guided database of traditional Chinese medicine. *Nucleic Acids Research* 49 (D1): D1197–D1206. <https://doi.org/10.1093/nar/gkaa1063>.
- Gong, J., F. Yin, Y. Hou, and Y. Yin. 2014. Review: Chinese herbs as alternatives to antibiotics in feed for swine and poultry production: potential and challenges in application. *Canadian Journal of Animal Science* 94: 223–241. <https://doi.org/10.4141/CJAS2013-144>.
- Hogan, A., E. Blomqvist, M. Cochez, C. d'Amato, G. de Melo, C. Gutierrez, J.E.L. Gayo, S. Kirrane, S. Neumaier, A. Polleres, et al. 2021. Knowledge Graphs. *ACM Computing Surveys*. 54(4): 71:1–71:37. <https://doi.org/10.1145/3447772>.
- Holman, D.B., and M.R. Chénier. 2015. Antimicrobial use in swine production and its effect on the swine gut microbiota and antimicrobial resistance. *Canadian Journal of Microbiology* 61 (11): 785–798. <https://doi.org/10.1139/cjm-2015-0239>.
- Hsu, S.C., and J.G. Chung. 2012. *Anticancer potential of emodin. Biomedicine (Taipei)* 2 (3): 108–116. <https://doi.org/10.1016/j.biomed.2012.03.003>.
- Huang, L., D.L. Xie, Y.R. Yu, H.L. Liu, Y. Shi, T.L. Shi, and C.P. Wen. 2018. TCMID comprehensive resource for TCM. *Nucleic Acids Research* 46 (D1): D1117–D1120. <https://doi.org/10.1093/nar/gkx1028>.
- Jia, Q., R. Zhu, Y. Tian, B. Chen, R. Li, L. Li, L. Wang, Y. Che, D. Zhao, F. Mo, S. Gao, and D. Zhang. 2019. *Salvia miltiorrhiza* in diabetes: a review of its pharmacology, phytochemistry, and safety. *Phytomedicine* 58: 152871. <https://doi.org/10.1016/j.phymed.2019.152871>.
- Karimi, A., M. Majlesi, and M. Rafeian-Kopaei. 2015. Herbal versus synthetic drugs; beliefs and facts. *Journal of Nephrotherapeutics* 4 (1): 27–30.
- Karniyuchuk, U.U., M. Geldhof, M. Vanhee, J. Van Doorselaere, T.A. Saveleva, and H.J. Nauwynck. 2010. Pathogenesis and antigenic characterization of a new East European subtype 3 porcine reproductive and respiratory syndrome virus isolate. *BMC Veterinary Research* 6: 30. <https://doi.org/10.1186/1746-6148-6-30>.
- Keffaber, K.K. 1989. Reproductive failure of unknown etiology. *American Association of Swinem Practice News* 1: 1–9.
- Kuhn, M., D. Szklarczyk, A. Franceschini, M. Campillos, C. von Mering, L.J. Jensen, A. Beyer, and P. Bork. 2010. STITCH 2: an interaction network database for small molecules and proteins. *Nucleic Acids Research* 38 (Database issue): D552–D556. <https://doi.org/10.1093/nar/gkp937>.
- Kuwahara, H., T. Nunoya, M. Tajima, A. Kato, and T. Samejima. 1994. An outbreak of porcine reproductive and respiratory syndrome in Japan. *Journal of Veterinary Medical Science* 56 (5): 901–909. <https://doi.org/10.1292/jvms.56.901>.
- Larmande, P., G. Tagny Ngompe, A. Venkatesan, and M. Ruiz. 2022. AgroLD: A knowledge graph database for plant functional genomics. *Methods in Molecular Biology* 2443: 527–540. https://doi.org/10.1007/978-1-0716-2067-0_28.
- Li, Y.M., Z.C. Wu, K. Liu, P.F. Qi, J.P. Xu, J.C. Wei, B.B. Li, D.H. Shao, Y.Y. Shi, Y.F. Qiu, and Z.Y. Ma. 2017. Doxycycline enhances adsorption and inhibits early-stage replication of porcine reproductive and respiratory syndrome virus in vitro. *FEMS Microbiology Letters* 364 (17): fnx170. <https://doi.org/10.1093/femsle/fnx170>.
- Lin, Y.K., Z.Y. Liu, M.S. Sun, Y. Liu, X. Zhu. 2015. Learning entity and relation embeddings for knowledge graph completion. In: Proc. of the 29th AAAI Conf. on Artificial Intelligence (AAAI). Austin: AAAI Press 2181–2187. <https://doi.org/10.5555/2886521.2886624>.
- MacLean, F. 2021. Knowledge graphs and their applications in drug discovery. *Expert Opinion on Drug Discovery* 16 (9): 1057–1069. <https://doi.org/10.1080/17460441.2021.1910673>.
- Mencia-Ares, O., H. Argüello, H. Puente, M. Gómez-García, A. Álvarez-Ordóñez, E.G. Manzanilla, A. Carvajal, and P. Rubio. 2021. Effect of antimicrobial use and production system on *Campylobacter* spp., *Staphylococcus* spp. and *Salmonella* spp. resistance in Spanish swine: A cross-sectional study. *Zoonoses Public Health* 68 (1): 54–66. <https://doi.org/10.1111/zph.12790>.
- Nickel, M., V. Tresp, H.P. Kriegel. 2011. A three-way model for collective learning on multirelational data. In: Proc. of the 28th Int'l Conf. on Machine Learning (ICML). Bellevue: Omnipress 809–816. <https://doi.org/10.5555/3104482.3104584>.
- Sha, H.Y., H. Zhang, Y. Chen, L.Z. Huang, M.M. Zhao, and N. Wang. 2022. Research progress on the NSP9 protein of porcine reproductive and respiratory syndrome virus. *Frontiers in Veterinary Science* 9: 872205. <https://doi.org/10.3389/fvets.2022.872205>.
- Singhal. 2012. Introducing the knowledge graph: things, not strings. *Official google blog*. <https://blog.google/products/search/introducing-knowledge-graph-things-not/>
- Snijder, E.J., M. Kikkert, and Y. Fang. 2013. Arterivirus molecular biology and pathogenesis. *Journal of General Virology* 94 (Pt 10): 2141–2163. <https://doi.org/10.1099/vir.0.056341-0>.

- Soares, V.M., J.G. Pereira, F. Barreto, L. Jank, R.B. Rau, C.B. Dias Ribeiro, T. Dos Santos Castilhos, C.A. Tomaszewski, D.R. Hillesheim, R.G. Mondadori, et al. 2022. Residues of veterinary drugs in animal products commercialized in the Border Region of Brazil, Argentina, and Uruguay. *Journal of Food Protection* 85 (6): 980–986. <https://doi.org/10.4315/JFP-21-415>.
- Su, C.Y., Q.L. Ming, K. Rahman, T. Han, and L.P. Qin. 2015. *Salvia miltiorrhiza*: Traditional medicinal uses, chemistry, and pharmacology. *Chinese Journal of Natural Medicines* 13 (3): 163–182. [https://doi.org/10.1016/S1875-5364\(15\)30002-9](https://doi.org/10.1016/S1875-5364(15)30002-9).
- Sun Z.Q., Z.H. Deng, J.Y. Nie, J. Tang. 2019. RotatE: Knowledge graph embedding by relational rotation in complex space. [arXiv:1902.10197](https://arxiv.org/abs/1902.10197). <https://doi.org/10.48550/arXiv.1902.10197>
- Szklarczyk, D., A.L. Gable, K.C. Nastou, D. Lyon, R. Kirsch, S. Pyysalo, N.T. Doncheva, M. Legeay, T. Fang, P. Bork, L.J. Jensen, and C. von Mering. 2021. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research* 49 (D1): D605–D612. <https://doi.org/10.1093/nar/gkaa1074>.
- Tantituvanont, A., W. Yimprasert, P. Werawatganone, and D. Nilubol. 2009. Pharmacokinetics of ceftiofur hydrochloride in pigs infected with porcine reproductive and respiratory syndrome virus. *Journal of Antimicrobial Chemotherapy* 63 (2): 369–373. <https://doi.org/10.1093/jac/dkn496>.
- Trouillon T., J. Welbl, S. Riedel, É. Gaussier, G. Bouchard. 2016. Complex embeddings for simple link prediction. In: *Proc. of the 33rd Int'l Conf. on Machine Learning (ICML)*. New York: JMLR 48:2071–2080. <https://doi.org/10.5555/3045390.3045609>.
- Wang, T., K. Guan, Q. Su, X. Wang, Z. Yan, K. Kuang, Y. Wang, Q. Zhang, X. Zhou, and B. Liu. 2022. Change of gut microbiota in PRRSV-Resistant Pigs and PRRSV-susceptible pigs from tongcheng pigs and large white pigs crossed population upon PRRSV infection. *Animals (Basel)*. 12 (12): 1504. <https://doi.org/10.3390/ani12121504>.
- Wensvoort, G., C. Terpstra, J.M. Pol, E.A. ter Laak, M. Bloemraad, E.P. de Kluyver, C. Kragten, L. van Buiten, A. den Besten, F. Wagenaar, and A. Et. 1991. Mystery swine disease in The Netherlands: the isolation of Lelystad virus. *Veterinary Quarterly* 13 (3): 121–130. <https://doi.org/10.1080/01652176.1991.9694296>.
- Yang, X.D., X.Y. Lian, C. Fu, S. Wuchty, S.P. Yang, and Z.D. Zhang. 2021. HVIDB: a comprehensive database for human-virus protein-protein interactions. *Briefings in Bioinformatics* 22 (2): 832–844. <https://doi.org/10.1093/bib/bbaa425>.
- Yang, B.S., W.T. Yih, X.D. He, J.F. Gao, L. Deng. 2015. Embedding entities and relations for learning and inference in knowledge bases. [arXiv:1412.6575](https://arxiv.org/abs/1412.6575). <https://doi.org/10.48550/arXiv.1412.6575>.
- Yin, B.S., S.S. Qi, W.L. Sha, H.Y. Qin, L.M. Liu, J.Y. Yun, J.H. Zhu, G.J. Li, and D.B. Sun. 2021. Molecular characterization of the Nsp2 and ORF5 (ORF5a) genes of PRRSV strains in nine provinces of China during 2016–2018. *Frontiers in Veterinary Science* 8: 605832. <https://doi.org/10.3389/fvets.2021.605832>.
- Zha, L.H., L.S. He, F.M. Lian, Z. Zhen, H.Y. Ji, L.P. Xu, and X.L. Tong. 2015. Clinical Strategy for Optimal Traditional Chinese Medicine (TCM) Herbal Dose Selection in Disease Therapeutics. *The American Journal of Chinese Medicine* 43 (8): 1515–1524. <https://doi.org/10.1142/S0192415X1550086X>.
- Zhang, M.X., C.N. Lu, L.Z. Su, F.X. Long, X. Yang, X.F. Guo, G.P. Song, T.Q. An, W.S. Chen, and J.X. Chen. 2022. Toosendanin activates caspase-1 and induces maturation of IL-1 β to inhibit type 2 porcine reproductive and respiratory syndrome virus replication via an IFI16-dependent pathway. *Veterinary Research* 53 (1): 61. <https://doi.org/10.1186/s13567-022-01077-2>.
- Zheng, D., X. Song, C. Ma, Z.Y. Tan, Z.H. Ye, J. Dong, H. Xiong, Z. Zhang, G. Karypis. 2020. DGL-KE: Training knowledge graph embeddings at scale. [arXiv: 2004.08532](https://arxiv.org/abs/2004.08532). <https://doi.org/10.48550/arXiv.2004.08532>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

