Development of a Rapid, Efficient, Intelligent and Cost-Saving Tool to Diagnose Pasteurella Multocida by Using Whole Genome Sequence and Genotypes of Pasteurella Multocida From Different Hosts

Zhong Peng
Huazhong Agriculture University
Junyang Liu
Huazhong Agricultural University
Wan Liang
Huazhong Agriculture University
Fei Wang
Huazhong Agriculture University
Li Wang
Huazhong Agriculture University
Lin Hua
Huazhong Agriculture University
Xiangru Wang
Huazhong Agriculture University
Chen Tan
Huazhong Agriculture University
Rui Zhou
Huazhong Agriculture University
Huanchun Chen
Huazhong Agriculture University
Brenda A. Wilson
University of Illinois at Urbana-Champaign
Jia Wang
Huazhong Agriculture University
Bin Wu (wub@mail.hzau.edu.cn)
Huazhong Agriculture University https://orcid.org/0000-0001-9078-386X

Methodology

Keywords: Pasteurella multocida, genotyping, host tropism prediction, whole genome sequence, machine learning

Posted Date: November 13th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-104569/v1

Abstract

Background: Different typing systems including capsular genotyping, lipopolysaccharide (LPS) genotyping, multilocus sequence typing (MLST), and virulence genotyping based on the detection of different virulence factor-encoding gene (VFG) profiles have been applied to characterize *Pasteurella multocida* strains from different host species. However, these methods require much time and effort in laboratories. Particularly, relying on one of these methods is difficult to address the biology of *P. multocida* from host species. Recently, we found that assigning *P. multocida* strains according to the combination of their capsular, LPS, and MLST genotypes (marked as capsular genotype: LPS genotype: MLST genotype) could help address the biological characteristics of *P. multocida* circulation in multiple hosts. However, it is still lack of a rapid, efficient, intelligent and cost-saving tool to diagnose *P. multocida* according to this system.

Results: We have developed an intelligent genotyping and host tropism prediction tool PmGT for *P. multocida* strains according to their whole genome sequences by using machine learning and web 2.0 technologies. By using this tool, the capsular genotypes, LPS genotypes, and MLST genotypes as well as the main VFGs of *P. multocida* isolates in different host species were determined based on whole genome sequences. The results revealed a closer association between the genotypes and pasteurellosis rather than between genotypes and host species. Finally, we also used PmGT to predict the host species of *P. multocida* strains with the same capsular: lipopolysaccharide: MLST genotypes.

Conclusions: With the advent of high-quality, inexpensive DNA sequencing, this platform represents a more efficient and cost-saving tool for *P. multocida* diagnosis in both epidemiological studies and clinical settings.

Background

Rapid and accurate diagnosis of sources of infections is critical for both medical and veterinary activities, and it is important for improved understanding of disease mechanisms and measures to control the illness [1]. Microbial typing is an important link for the diagnosis of pathogens associated with diseases. The most wildly used typing methods consist of serological typing systems and PCR-based molecular typing methods [2–4]. The establishment of discriminatory typing systems help in the understanding and control of pathogens, especially those with multiple serotypes and/or genotypes from different environmental or host sources. The whole genome sequencing combined with the high-end computational technology is such an emerging tool for microbial diagnosis. Using the whole genome sequencing technology, the causative agent of specific diseases, even from new infectious diseases, can be rapidly and accurately characterized. However, diagnosis based on whole genome sequencing requires technical experts with computational and bioinformatics skills. Therefore, a practical automated, intelligent platform in combination of whole genome sequencing and computational technology will be beneficial.

Pasteurella multocida is an important zoonotic pathogen and it is able to colonize and causes infections in a wide range of domestic and wild animals including food producing animals (e.g. poultry, pigs, beef, sheep, etc.) and companion animals (e.g. cats and dogs) as well as in humans [5–7]. Animal diseases associated with *P. multocida* such as fowl cholera in poultry and other birds, progressive atrophic rhinitis and pneumonic pasteurellosis in pigs, bovine haemorrhagic septicaemia and respiratory diseases, leporine atrophic rhinitis and pneumonic pasteurellosis, are of great economic significance in agriculture [7]. In humans, opportunistic infections of soft tissue, including wound dermonecrosis, respiratory disease with chronic pulmonary, urinary tract infection and bacteremic meningitis have also been reported [7]. Although there are no reports of human infections via food chain, *P. multocida* infections in humans due to pets' biting, scratching, kissing, and/or licking have been documented in many literatures [8–10]. In this regard, *P. multocida* represents a risk to public health. Serologically, *P. multocida* strains from different hosts are serologically classified into five serogroups (A, B, D, E, F) and/or 16 serovars (serovars 1 to 16), according to their capsular and lipopolysaccharide (LPS) antigens, respectively [11, 12]. However, these two traditional serological typing methods require high-quantity antisera that are challenging to prepare, particularly for clinical use, such that those methods are no longer widely used for large-scale epidemiological studies [5, 13].

In 2001, a multiplex PCR-based method was established to type the five serogroups into five capsular genotypes (A, B, D, E, F) [14], and in 2015, another multiplex PCR-based method was also developed to classified the 16 serovars into eight LPS genotypes (L1 ~ L8) [15]. In 2004 and 2010, two multilocus sequencing typing systems were also developed to genotype *P. multocida* strains (https://pubmlst.org/pmultocida/) from multiple mammalian hosts and birds, respectively [16, 17]. In 2017, a virulence genotyping system based on the detection of different virulence factor gene (VFG) profiles was also reported for distinguishing *P. multocida* strains from different hosts [18]. Compared to the traditional serological typing methods, these molecular DNA-based typing systems are indeed highly effective and accurate, and they are now wildly used to determine the epidemiological and genetic characteristics of clinical isolates [19–23].

Despite of more than 135 years of research, differences on the molecular biological characteristics of *P. multocida* prevalence in different host species remain to be addressed. For example, *P. multocida* type A strains have been recovered from avian species, pigs, bovine species, and many other host species [6, 7], but little is known about differences on those type A isolates from different hosts. Recently, we developed a system to assign *P. multocida* strains from different host species by combining their capsular, LPS, and MLST genotypes (marked as capsular genotype: LPS genotype: MLST genotype), as well as determine the VFG profiles, which contributes to address the molecular biological characteristics of *P. multocida* prevalence in different host species [23, 19, 5]. However, this strategy by using the whole genome sequence based on local BLAST programs and require bioinformatics experts for data analysis and interpretation. Here, we report the development of an automated and intelligent platform to type *P. multocida* strains from multiple hosts that combines the use of whole genome sequencing and machine learning technologies.

Results

Development and Implementation of PmGT

Previous studies shown primers specific for *P. multocida* (*KMT1*) and its five capsular genotypes (A, B, D, E, F) [14], eight LPS genotypes (L1 ~ L8) [15], as well as 23 kinds of virulence factors-encoding genes (VFGs) commonly detected in epidemiological studies (*ptfA*, *fimA*, *hsf-1*, *hsf-2*, *pfhA*, *tadD*, *toxA*, *exbB*, *exbD*, *tonB*, *hgbA*, *hgbB*, *fur*, *tbpA*, *nanB*, *nanH*, *pmHAS*, *ompA*, *ompH*, *oma87*, *plpB*, *sodA* and *sodC*) [5]. Therefore, we extracted those primers-targeted nucleotide sequences (Supplementary Txt 1) from the complete genome sequences of our previously sequenced *P. multocida* strains, including HB01 (serogroup A) [24], HN04 (serogroup B) [23], HN06 (serogroup D, producing toxin) [25], and HN07 (serogroup F) [26]. These nucleotide sequences were then stored on a CentOS server. Afterwards, we downloaded the BLAST package from the NCBI website (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/) and installed it on this CentOS server. A PHP program was developed to call the BLAST algorithm and attain the genotyping results for capsular and LPS genotype queries. When the users select the MLST genotyping service, the PHP CURL (Client URL Library) functions were used to request and obtain the results through the RESTful service interface provided by the Public Database for Molecular Typing (PubMLST, https://pubmlst.org). The general process for genotyping is summarized as: when a query sequence is submitted via the web user interface, this sequence will be then submitted to the CentOS server via HTTP protocol. Thereafter, the sequence is evaluated by the PHP program, and the passed sequence will be BLASTed against the genotype database to yield a result, which will be returned to the webpage through the PHP program (Figs. 1A and 1B). Through the above procedures, the genotyping module of PmGT (http://liulab.hzau.edu.cn/PM/) was developed (Fig. 1).

Because *P. multocida* strains colonized wide spectrums of host species, we therefore intend to develop a model to predict the host tropism of emerging *P. multocida* strains based on their genome sequences. Many epidemiological and genomic studies on *P. multocida* have revealed that VFGs show a correlation with host species [5, 19, 27]. These findings make it possible to determine the host tropism of a *P. multocida* strain by analyzing the VFGs carried by its whole genome sequence. As of 31 May 2020, 262 sequences of *P. multocida* strains from different host species are publicly available through the NCBI genome database. These sequences are of *P. multocida* strains from different host species, including pigs (n = 66), poultry and wild birds (n = 39), cattle and other bovine species (n = 106), canis lupus (n = 3), cats (n = 2), humans (n = 13), horses (n = 2), rabbits and other leporine species (n = 20), rodents (n = 2), sheep and other ovine species (n = 6), and vicugna pacos (n = 2) (Supplementary Table S1). In addition, there is also one synthetic DNA sequence. Since there are few genome sequences of *P. multocida* isolates from other hosts except porcine, bovine and avian species publicly available in NCBI, we used the genome sequences of porcine, bovine, and avian isolates to develop and test the host tropism prediction model. Approximately 70% of the genome sequences from *P. multocida* of porcine (n = 16), bovine (n = 65), and avian origin (n = 19) were aligned against the nucleotide sequences of the 23 kinds of VFGs described above by using the BLAST tool. BLAST scores were then used as features and input into six

different machine learning models (Entropy Decision Tree [eDTA], Gini DTA [gDTA], Brute K-Nearest Neighbor [bKNN], Ball Tree KNN [btKNN], Gaussian Naive Bayes [GNB], and Complement Naive Bayes [CNB]) to calculate the precision, the recall, and F1 score, as described in the Methods section. In each of the models, three different metrics (micro-, macro-, and weighted-average of F1 score) were calculated through 10-fold cross validation. The results revealed that the Decision Tree model showed overall higher average F1 scores than the KNN and the Bayesian models (Fig. 2A and Supplementary Figures S1A, S1B, S1C). Therefore, the Decision Tree algorithm was finally chosen to construct the host tropism prediction model. Scikit-learn (Sklearn) and NumPy in Python were applied to implement the above findings into automation and intelligence machine learning model, which was available at http://liulab.hzau.edu.cn/PM/model.php.

Currently, PmGT provides the above services includes six menus: (1) the “Home” page gives a brief introduction of *P. multocida* etiological characteristics to help the users understand the bacterium; (2) the “Organisms” page displays the genotypes of *P. multocida* strains based on their whole genome sequences that are publicly available in NCBI; this page also provides the link for the users to download the genomes of these *P. multocida* strains from NCBI; (3) the “Genotyping” page enables the users to determine whether a putative isolate is a *P. multocida* and genotype *P. multocida* strains by using the whole genome sequence assembled from the sequencing reads (Fig. 1C); (4) the “Host Prediction” page enables users to predict the host tropism of *P. multocida* isolates by submission of the whole genome sequences (currently only prediction of porcine, bovine, avian, and/or human isolates is available due to the limited number of genome sequences of *P. multocida* from the other hosts in NCBI) (Fig. 1D); (5) the “About” page summarizes the guidelines for the use of this web tool; (6) the “Contact” page provides the contact information of the developers.

PmGT shows the same accuracy with PCR methods in genotyping P. multocida strains

To test the accuracy of PmGT, we used two methods to type 52 *P. multocida* isolates (HB01, HB02, HB03, HN04, HN05, HN06, HN07, HNA01 ~ HNA22, HND01 ~ HND21, HNF01, and HNF02) from our laboratory collection [23]. First, we submitted their whole genome sequences to PmGT for genotyping. As a comparison, we also determined the capsular genotypes, LPS genotypes, sequence types, as well as the profile of the abovementioned 23-kinds of virulence genes by using PCR assays. All these 52 strains were genotyped by PmGT and through this online genotyping platform (Table 1). Genotyping by PCR assays confirmed these capsular, LPS, and MLST genotypes. PCR results of capsular and LPS genotypes are provided in Supplementary Figures S2 and S3.

Table 1
Genotypes of 52 *Pasteurella multocida* strains determined via the PmGT Platform

Strain	Capsular genotype	LPS genotype	MLST genotype (Sequence type)	GenBank accession numbers
HB01	A	L3	ST1	CP006976
HB02	A	L1	ST128	LYOX00000000
HB03	A	L3	ST3	CP003328
HN04	B	L2	ST44	PPVE00000000
HN05	D	L6	ST11	PPVF00000000
HN06	D	L6	ST11	CP003313
HN07	F	L3	ST12	CP007040
HNA01	A	L3	ST133	PPVG00000000
HNA02	A	L6	ST10	PPVH00000000
HNA03	A	L3	ST3	PPVI00000000
HNA04	A	L6	ST10	PPVJ00000000
HNA05	A	L6	ST10	PPVK00000000
HNA06	A	L6	ST10	PPVL00000000
HNA07	A	L6	ST10	PPVM00000000
HNA08	A	L3	ST3	PPVN00000000
HNA09	A	L3	ST3	PPVO00000000
HNA10	A	L6	ST10	PPVP00000000
HNA11	A	L6	ST10	PPVQ00000000
HNA12	A	L6	ST10	PPVR00000000
HNA13	A	L3	ST3	PPVS00000000
HNA14	A	L3	ST3	PPVT00000000
HNA15	A	L3	ST3	PPVU00000000
HNA16	A	L6	ST10	PPVV00000000
HNA17	A	L3	ST3	PPVW00000000
HNA18	A	L3	ST3	PPVX00000000
HNA19	A	L3	ST3	PPVY00000000

Strain	Capsular genotype	LPS genotype	MLST genotype (Sequence type)	GenBank accession numbers
HNA20	A	L3	ST3	PPVZ00000000
HNA21	A	L6	ST10	PPWA00000000
HNA22	A	L6	ST10	PPWB00000000
HND01	D	L6	ST11	PPWC00000000
HND02	D	L6	ST134	PPWD00000000
HND03	D	L6	ST11	PPWE00000000
HND04	D	L6	ST11	PPWF00000000
HND05	D	L6	ST11	PPWG00000000
HND06	D	L6	ST11	PPWH00000000
HND07	D	L6	ST11	PPWI00000000
HND08	D	L6	ST11	PPWJ00000000
HND09	D	L6	ST11	PPWK00000000
HND10	D	L6	ST11	PPWL00000000
HND11	D	L6	ST11	PPWN00000000
HND12	D	L6	ST134	PPWM00000000
HND13	D	L6	ST134	PPWO00000000
HND14	D	L6	ST11	PPWP00000000
HND15	D	L6	ST11	PPWQ00000000
HND16	D	L6	ST11	PPWR00000000
HND17	D	L6	ST11	PPWS00000000
HND18	D	L6	ST11	PPWT00000000
HND19	D	L6	ST11	PPWU00000000
HND20	D	L6	ST11	PPWV00000000
HND21	D	L6	ST11	PPWW00000000
HNF01	F	L3	ST12	PPWX00000000
HNF02	F	L3	ST12	PPWY00000000

Determination of the 23 types of virulence genes for each of the 52 strains by using this online system revealed that several genes (*ptfA*, *fimA*, *oma87*, and *sodC)* were broadly presented in the genome sequences genotyped (Fig. 3). However, several genes (*hsf-1*, *hsf-2*, *pfhA*, and *tadD)* were heterogeneously distributed, and in particularly, none of the 52 sequences genotyped carried the *toxA* or *tbpA* genes (Fig. 3). These results were also confirmed by PCR assays (Supplementary Table S1).

Genotypes of P. multocida from different hosts

To understand the genotypes of *P. multocida* strains circulation in different host species, the 262 whole genome sequences of *P. multocida* strains that are publicly available through the NCBI genome database as of 31 May 2020 were downloaded and were genotyped by PmGT. The results revealed that *P. multocida* strains isolated from different host species were preference to several specific capsular genotypes, LPS genotypes, and/or sequence types (Fig. 4). For example, most of the porcine strains were determined as capsular genotypes A (52%) and D (39%), LPS genotypes L3 (36%) and L6 (61%), sequence types ST3 (29%), ST11 (22%), and ST10 (34%), respectively; while most of the genotyped bovine strains were determined as capsular genotypes A (72%) and B (28%), LPS genotypes L3 (67%) and L2 (27%), and sequence types ST1 (59%) and ST44 (25%), respectively (Fig. 4). When combining the capsular genotypes and the LPS genotypes, it revealed that most of the genotyped avian *P. multocida* were typed as A:L1 and A:L3, while most of the genotyped bovine *P. multocida* were typed as A:L3 and B:L2; the genotyped porcine *P. multocida* mainly belonged to D:L6, A:L3, and A:L6; while the genotyped leporine *P. multocida* mainly belonged to A:L3; most of the genotyped human *P. multocida* were typed as A:L3 and A:L1 (Fig. 5A). If the capsular genotypes, LPS genotypes, and MLST genotypes were combined, most of the genotyped avian *P. multocida* were typed as A:L1:ST128 (Fig. 5B), while most of the genotyped bovine *P. multocida* were typed as A:L3:ST1 and B:L2:ST44 (Fig. 5C); the genotyped porcine *P. multocida* mainly belonged to D:L6:ST11, A:L3:ST3, and A:L6:ST10 (Fig. 5I); while the genotyped leporine *P. multocida* mainly belonged to A:L3:ST12 (Fig. 5H).

Virulence genotyping using the system developed herein revealed that the presence of multiple VFGs, including *ptfA*, *fimA*, *hsf-2*, *exbB*, *exbD*, *tonB*, *hgbA*, *hgbB*, *fur*, *nanB*, *nanH*, *ompA*, *ompH*, *oma87*, *plpB*, *sodA*, and *sodC*, was a broad characteristic of *P. multocida* strains from multiple host species (Fig. 6). However, several VFGs were only determined in the genome sequences of *P. multocida* from certain hosts. For example, *toxA*, a gene encoding a dermonecrotic toxin, was found only in strains from pig, sheep, and alpacas, while *tbpA*, a transferrin binding protein coding gene, was found only in strains from cattle, sheep, and alpacas (Fig. 6).

PmGT is able to predict the host tropism of P. multocida

By using the Entropy Decision Tree algorithms, the correlation of *P. multocida* VFGs and host species was revealed (Fig. 2B). We then used the remaining 30% of the genome sequences of *P. multocida* from porcine ($n = 3$), bovine ($n = 31$), and avian origin ($n = 10$) to test the host tropism prediction model developed herein. The average micro-F1 score reached 0.898, revealing that the model could predict the host species of the tested strains. In particularly, it could determine host species of *P. multocida* strains

possessing the same genotypes and close relatedness correctly (compare the result of an avian F:L3:ST25 type isolate Pm70 vs. the result of a porcine F:L3:ST12 isolate HN07 [Figs. 7A vs. 7B]; as well as compare the result of a porcine B:L2:ST44 type isolate HN04 vs. the result of a bovine B:L2:ST44 type isolate ATTK [Figs. 7C vs. 7D]).

Because *P. multocida* strains are also frequently recovered in clinical settings of human medicine [7]. To facilitate a rapid for help diagnosis, we also implement a way to predict the hosts of putative *P. multocida* strains from humans by using the same principles, even though the current publicly available genome sequences for *P. multocida* of human origin are still limited (only 13 sequences as of 31 May 2020). We used 9 sequences to develop the model and used the additional 4 sequences to test. However, the results showed this model was still appliable for *P. multocida* of humans (Fig. 7E).

Discussion

P. multocida is the causative agent of multiple diseases with a wide spectrum of host species, including humans and other primates [5–7]. In addition, *P. multocida* isolates recovered from different hosts with different diseases can be classified in many different serotypes/genotypes according to different typing systems [7, 5]. Relying on only one or two typing systems is difficult to address the characteristics of *P. multocida* isolates from different host species and/or their association with different diseases. For example, *P. multocida* isolates from different host species might have the same capsular genotypes but possess different LPS genotypes and/or MLST genotypes; even those from different host species that share the same capsular, LPS, and MLST genotypes might carry different VFGs [27, 23]. Therefore, we have proposed a combined “capsular: LPS: MLST” genotyping system that includes virulence genotyping to discriminate *P. multocida* isolates from different host species and/or those associated with different diseases [5]. However, this combined genotyping system is multiplex PCR-based and is laborious and time-consuming.

Advances in bioinformatics and machine learning tools enable the application of whole genome sequence data for inclusion of various demographic information for bacterial characterization, such as capsular and LPS genotyping; the presence of adhesins, toxins, or other virulence factors [28]. In the present study, we reported the development of a machine-learning genotyping platform for distinguishing *P. multocida* strains from different host species according to the bacterial whole genome sequences (Fig. 1). Validation of the PmGT platform was performed on a collection of *P. multocida* isolates from our laboratory. Results revealed that this genotyping system provides consistent results of determining the capsular-, LPS-, MLST genotypes, and VFGs, as well as host species prediction, as compared with that obtained using multiplex PCR-based typing systems (Fig. 3, Table 1, and Supplementary Table S1). Compared to the multiplex PCR-based typing systems [14, 15, 17, 18] and traditional serological typing systems [11, 12], this machine-learning system takes less time to yield results and does not require high-quality antisera, which represents a more efficient and cost-saving tool for characterizing *P. multocida* isolates in both epidemiological studies and clinical settings.

By using PmGT, the capsular-, LPS-, MLST genotypes, and VFGs of *P. multocida* strains from different hosts were determined according to the whole genome sequences (Figs. 4, 5, Supplementary Table S1). These results are in agreement in those of the epidemiological studies [19, 20, 29, 30]. For example, *P. multocida* serotypes B: 2 and A: 3 strains are frequently associated with bovine haemorrhagic septicaemia and respiratory diseases, respectively [31, 32]. It is known that *P. multocida* serogroups A and B are assigned to capsular genotypes A and B by multiplex PCR, respectively [14]; while *P. multocida* Heddleston serotypes 2 and 3 are assigned to LPS genotypes L2 and L3 by multiplex PCR, respectively [15]. That is why the capsular: LPS genotypes of most of the bovine strains were determined as A: L3 and B: L2, respectively (Fig. 5A). In addition, *P. multocida* strains isolated from bovine haemorrhagic septicaemia are commonly determined as ST122 [33], this sequence type can be reassigned to ST44 by using the multihost MLST database [23]. These findings could explain why *P. multocida* strains associated with bovine haemorrhagic septicaemia were typed as capsular: LPS: MLST genotype B: L2: ST44 (Fig. 5C). Similar findings were also observed in *P. multocida* strains from the other host species (Fig. 5). In particularly, most of the *P. multocida* strains from pigs were determined as capsular: LPS: MLST genotypes D: L6: ST11, A: L3: ST3, and A: L6: ST10 (Fig. 5I). These results are also in agreement with the results of our previously epidemiological study [19], suggesting that these three genotypes, particularly genotype D/L6/ST11, are likely to be strongly associated with swine respiratory diseases. However, during our test we also found the capsular-, LPS-, and/or MLST-genotypes of several strains could not be determined by PmGT according to the whole genome sequences (Figs. 4, 5, Supplementary Table S1). After check the data we put forward several reasons to explain this result: 1.) most of these nontypeable genomes are sequenced and assembled using the second-generation sequencing technologies and the quality of these genomes are not high, some of the genes used for capsular/LPS/MLST genotyping fell within the gaps between genome contigs in the assemblies [5]; 2.) the genome sequences might be those of the capsular nontypeable strains reported in clinic [19, 34]; 3.) several strains belong to novel sequence types and the current *Pasteurella multocida* MLST database do not include these sequence types.

Our PmGT platform also uses the machine-learning technology to predict the host species of *P. multocida* strains according to the whole genome sequence. Although many epidemiological and genomic studies have shown that *P. multocida* isolates from different hosts have preferences for different capsular/LPS/MLST genotypes and/or VFGs [18, 19, 5, 27, 6], not all *P. multocida* isolates could be typed through the capsular genotyping system and/or MLST genotyping system [19, 18, 5] (Supplementary Table S1), and in addition, we have determined that *P. multocida* strains with the same capsular: LPS: MLST genotypes can be isolated from different host species [23]. For example, capsular: LPS: MLST genotype B: L2: ST44 strains have been isolated from hemorrhagic septicemia cases in pigs (e.g. strain HN04, GenBank accession no. PPVE00000000) and bovine species (e.g. strain ATTK, GenBank accession no. JQEA00000000). Likewise, capsular: LPS: MLST genotype A: L3: ST3 strains have been isolated from pneumonic pasteurellosis cases in pigs (e.g. strain HB03, GenBank accession no. CP003328) and bovine species (e.g. strain 2125PM, GenBank accession no. LQCZ00000000) (Supplementary Table S1). These

findings suggest a closer association between the genotypes and diseases rather than between genotypes and host species.

In contrast to the finding that *P. multocida* isolates from different hosts often share the same capsular: LPS: MLST genotypes, *P. multocida* strains from different hosts usually show different VFG profiles [27, 5, 18, 35], which makes it possible to develop a host-prediction model based on VFGs. We applied different machine learning algorithms to determine the association between *P. multocida* VFGs and the host species (Fig. 2), and our results revealed that different VFGs displayed different presence-scores in the genome sequences of strains from different hosts (Fig. 2). These results are in agreement with the results from previous epidemiological and molecular evolutionary studies [35, 18, 5, 27]. Our test results also revealed that the algorithm developed based upon the VFGs was able to determine the host species of *P. multocida* strains possessing the same genotypes and close relatedness (Pm70 vs. HN07; HN04 vs. ATTK) (compare Figs. 7A vs. 7B; Figs. 7C vs. 7D). It is noteworthy that using the same principle the host tropism of *P. multocida* strains from humans are predicted by using the whole genome sequence (Fig. 7E), even though the publicly available genome sequences for human *P. multocida* strains are currently limited (Supplementary Table S1). However, these findings are still suggestive of a possibility of developing a host-prediction model based on VFGs by using the machine learning technologies. With the availability of more genome sequences of *P. multocida* from different host species we will be able to expand this algorithm to predict *P. multocida* from the other hosts.

Conclusions

In conclusion, we developed an automated, intelligent genotyping and host tropism prediction system for *P. multocida* strains (PmGT platform), which combines whole genome sequence analysis tools with machine learning technologies. By using this system, we determined the genotypes of *P. multocida* isolates from different host species. In addition, this tool can help to predict the possible host species for *P. multocida* by using the whole genome sequence. Overall, this system represents a more efficient and cost-saving tool for *P. multocida* diagnosis in both epidemiological studies and clinical settings. More importantly, our study provides an example to develop rapid, efficient, intelligent and cost-saving tools for bacterial diagnosis by using their whole genome sequences in the coming age of artificial intelligence.

Methods

Bacterial strains and nucleotide sequences

P.multocida strains used in this study include one isolate of bovine origin (strain HB01), one isolate of avian origin (strain HB02), and 50 isolates of porcine origin (strains HB03, HN04, HN05, HN06, HN07, HNA01~HNA22, HND01~HND21, HNF01 and HNF02) (Supplementary Table S1). All of these strains are from our laboratory collection, for which we have previously characterized their whole genome sequences [24,36,23,26,25].

Nucleotide sequences specific for the determination of *P. multocida* strains (*KMT1*, 460 bp), and their the five capsular genotypes (A , 1044 bp; B, 760 bp; D, 657 bp; E, 511 bp; F, 851 bp; as well as their eight LPS genotypes (L1, 1307 bp; L2, 810 bp; L3, 474 bp; L4, 550 bp; L5, 1175 bp; L6, 668 bp; L7, 931 bp; L8, 255 bp) were extracted from the genome sequences of the different *P. multocida* strains according to the positions documented in previous publications [14,15] and were deposited in GenBank under accession numbers MT570166, MN938443~MN938455 (Supplementary Text 1).

The nucleotide sequences of 23 types of virulence genes commonly detected in *P. multocida* epidemiological studies, including those encoding fimbriae and other adhesins (*ptfA*, *fimA*, *hsf-1*, *hsf-2*, *pfhA*, and *tadD*), toxin (*toxA*), iron acquisition proteins (*exbB*, *exbD*, *tonB*, *hgbA*, *hgbB*, *fur*, and *tbpA*), sialidases (*nanB* and *nanH*), hyaluronidase (*pmHAS*), outer membrane proteins (OMPs) (*ompA*, *ompH*, *oma87*, and *plpB*), and superoxide dismutase (*sodA* and *sodC*), were amplified from the genomic DNA of *P. multocida* HN06 and HB01 by PCR assays using the protocols documented elsewhere [19,37]. These nucleotide sequences were deposited in GenBank under accession numbers MT570167~ MT570166 (Supplementary Text 1).

The publicly available whole genome sequences of 262 *P. multocida* strains from bovine species (*n* = 106), avian species (*n* = 39), porcine species (*n* = 66), leporine species (*n* = 20), ovine species (*n* = 6), humans (*n* = 13), canines (*n* = 3), murine species (*n* = 2), horses (*n* = 2), cats (*n* = 2), alpacas (*n* = 2) and 1 synthetic DNA sequence in NCBI genome database (https://www.ncbi.nlm.nih.gov/genome/browse/#!/prokaryotes/912/) as of 31 May 2020 were downloaded for use (Supplementary Table S1).

System implementation

The PmGT platform was integrated on a CentOS server, mainly providing three kinds of online services: genotyping tool, host tropism prediction and data query and display. To establish the genotyping online service, we first used Apache (https://www.apache.org) as the web container. Then, we downloaded the BLAST package (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/LATEST/) from NCBI, which was thereafter installed and configured on the web container. PHP was used as the server-side language and the browser-side script used jQuery, which is a fast, small, and feature-rich JavaScript library. The view pages were constructed with markup language technologies, such as Hypertext Markup Language (HTML) and Cascading Style Sheets (CSS). For the target strain, the format of the sequence was first verified by the web user interface and then the sequence data was uploaded to the server through the PHP program which subsequently called the localized BLAST to align the uploaded sequence with the reference database. The nucleotide sequences specific for the determination of *P. multocida* strains, capsular genotypes, LPS genotypes, and the 23 types of virulence factor genes (VFGs) were packaged and used as the reference database for sequence alignment. Finally, the result was returned and displayed in the web page. In addition, if the user selected the option of “MLST genotyping”, the http request function “curl_setopt” in PHP was used to request PubMLST's RESTful interface (http://rest. pubmlst.

org/db/pubmlst_Pmultocida_seqdef/sequence) and the function “curl_exec” was used to catch the response which thereafter was parsed to the final result and displayed in the genotyping page.

The host tropism prediction service integrated the trained decision tree model with scikit-learn library on the PmGT platform (The details of model training can be found in the next section). Similarly, the strain sequence was submitted through the web page and the server-side PHP program called the genotyping tool to get the scoring results of the virulence genes which were subsequently input into the trained prediction model. Next, the “model.predict” function was used to export the prediction result, and the PHP program parsed the result and finally displayed it on the web page.

To provide the data query and display service, we used the Bootstrap framework which is a popular front-end toolkit used to create interactive applications on the client side. The data query and display page of this study used the table display module of Bootstrap, providing functions such as search, column selection, full-screen display, switching, refreshing data, and paging. Meanwhile, because of the relatively small volume, the strain data set was saved in JSON format.

Development of the host tropism prediction tool for *Pasteurella multocida*

In order to construct an effective host tropism prediction tool for *P. multocida*, appropriate features were first selected. Previous studies have shown that VFGs of *P. multocida* are related to host tropism [5,19,27]. Therefore, this study selected 23 VFGs as input features of the training model for machine learning. Furthermore, it is necessary to convert VFGs into numerical vectors. The feature extraction method of a single nucleotide is relatively simple, but it cannot reflect the characteristics of the entire gene. To make the model simplified and effective, whole genome sequences of *P. multocida* porcine, bovine, and avian isolates were first aligned against the nucleotide sequences of the 23 types of VFGs using the BLAST tool and then these BLAST scores were used as input features so that each strain was transformed into a numerical vector of 23 scores and subsequently the numerical matrix of training samples was input to perform model training.

Compared with Support Vector Machine (SVM) and deep Neural Network (NN), which are generally regarded as black box models, Decision Tree Algorithm (DTA), K Nearest Neighbor (KNN), and Naive Bayes (NB) models have higher interpretability. Therefore, this study first chose the DTA, KNN, and NB algorithms to construct the host tropism model. Next, Scikit-learn (https://scikit-learn.org/), which is a widely-used machine learning library of Python language, was used to construct DT, KNN and NB prediction models. To evaluate and compare these three kinds of models, this study used 10-fold cross-validation. The data set was divided into 10 subsets of similar size, each time the union of k-1 subsets was used as the training set and the remaining subset was used as the test set. Finally, the average of the 10 test results was returned for comparison. Precision, recall and F1 score (PRF) were calculated as indicators of the model performance. The higher PRF values are, the better model performance is. The definition of PRF were as follows:

$$P(Precision) = \frac{TP}{TP + FP}$$

$$R(Recall) = \frac{TP}{TP + FN}$$

$$F(F_1\ \ Score) = \frac{2PR}{P + R}$$

where TP is the number of true positives, FP is the number of false positives, and FN is the number of false negatives. Furthermore, because the prediction model of this study is a multi-classification problem involving multiple hosts, we calculated the micro-average, macro-average and weighted-average of the PRF. The micro-average was derived by first calculating the total number of TP, FP and FN of all categories, and then calculating results according to the above PRF formulas. For the macro-average, PRF values of each category were first calculated and then averaged. In addition, the weighted-average was averaged from the PRF value in the macro-average, according to the proportion of each category of sample.

PCR detection of capsular genotypes, LPS genotypes, MLST genotypes, and virulence genes of *P. multocida* strains from pigs

Capsular genotypes and LPS genotypes of *P. multocida* strains from our laboratory collection were determined using multiplex PCR-based assays, as documented elsewhere [15,14]. Profiles of 23 types of virulence genes mentioned above were determined by PCR assays, as described previously [19]. Sequence types (STs) were determined according to the protocols described in *Pasteurella multocida* MLST database (https://pubmlst.org/pmultocida/).

Data Availability

Nucleotide sequences specific for *P. multocida* and its capsular genotypes, LPS genotypes, as well as VFGs were publicly available in GenBank under accession numbers MN938443-MN938455 and MT570166~MT570166. The typing system developed in the present study is available at: http://liulab.hzau.edu.cn/PM/.

Abbreviations

BLAST: Basic Local Alignment Search Tool; LPS: lipopolysaccharide; MLST: Multilocus sequence typing; VFGs: Virulence factors-encoding genes; PRF: precision, recall, and false-positive rate; eDTA: Entropy Decision Tree Algorithm; gDTA: Gini Decision Tree Algorithm; vKNN: Violence K-Nearest Neighbor (KNN) algorithm; btKNN: Ball Tree K-Nearest Neighbor (KNN) algorithm; GNB: Gaussian Naive Bayes algorithm; CNB: Complement Naive Bayes algorithm.

Declarations

Acknowledgements

We sincerely acknowledge Huazhong Agricultural University College of Informatics for providing the CentOS server and the PubMLST database for the RESTful port for connection.

Authors' contributions

Conceptualization: ZP, BAW, JW, and BW; methodology: ZP, JL, WL, FW, LW, LH, XW, and CT; writing—original draft preparation: ZP and JL; writing—review and editing: ZP, RZ, HC, BAW, and BW; supervision: ZP, JW, and BW; project administration: ZP, JW, and BW; funding acquisition: ZP, JW, and BW.

Funding

This work was supported in part by China Postdoctoral Science Foundation (grant numbers: 2020T130232 and 2018M640719), the Fundamental Research Funds for the Central Universities of China (grant number 2662018JC034), Guangdong Provincial Key Laboratory of Livestock Disease Prevention (grant number YDWS1901), the Agricultural Science and Technology Innovation Program of Hubei Province (grant number 2018skjcx05), and the earmarked fund for China Agriculture Research System (grant number CARS-35).

Availability of data and materials

Nucleotide sequences specific for *P. multocida* and its capsular genotypes, LPS genotypes, as well as VFGs were publicly available in GenBank under accession numbers MN938443-MN938455 and MT570166~MT570166. The typing system developed in the present study is available at: http://liulab.hzau.edu.cn/PM/.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

References

1. Kessel M (2015) Why microbial diagnostics need more than money. Nat Biotechnol 33 (9):898-900. doi:10.1038/nbt.3328

2. Peng Z, Ling L, Stratton CW, Li C, Polage CR, Wu B, Tang YW (2018) Advances in the diagnosis and treatment of Clostridium difficile infections. Emerg Microbes Infect 7 (1):15. doi:10.1038/s41426-017-0019-4
3. Schmitz JE, Tang YW (2018) The GenMark ePlex((R)): another weapon in the syndromic arsenal for infection diagnosis. Future Microbiol 13:1697-1708. doi:10.2217/fmb-2018-0258
4. Sandler SG, Chen LN, Flegel WA (2017) Serological weak D phenotypes: a review and guidance for interpreting the RhD blood type using the RHD genotype. Br J Haematol 179 (1):10-19. doi:10.1111/bjh.14757
5. Peng Z, Wang X, Zhou R, Chen H, Wilson BA, Wu B (2019) Pasteurella multocida: Genotypes and Genomics. Microbiol Mol Biol Rev 83 (4). doi:10.1128/mmbr.00014-19
6. Wilkie IW, Harper M, Boyce JD, Adler B (2012) Pasteurella multocida: diseases and pathogenesis. Curr Top Microbiol Immunol 361:1-22. doi:10.1007/82_2012_216
7. Wilson BA, Ho M (2013) Pasteurella multocida: from zoonosis to cellular microbiology. Clin Microbiol Rev 26 (3):631-655. doi:10.1128/cmr.00024-13
8. Ryan JM, Feder HM, Jr. (2019) Dog licks baby. Baby gets Pasteurella multocida meningitis. Lancet 393 (10186):e41. doi:10.1016/s0140-6736(19)30953-5
9. Dryden MS, Dalgliesh D (1996) Pasteurella multocida from a dog causing Ludwig's angina. Lancet 347 (8994):123. doi:10.1016/s0140-6736(96)90250-0
10. Godey B, Morandi X, Bourdinière J, Heurtin C (1999) Beware of dogs licking ears. Lancet 354 (9186):1267-1268. doi:10.1016/s0140-6736(99)04197-5
11. Carter GR (1955) Studies on Pasteurella multocida. I. A hemagglutination test for the identification of serological types. Am J Vet Res 16 (60):481-484
12. Heddleston KL, Gallagher JE, Rebers PA (1972) Fowl cholera: gel diffusion precipitin test for serotyping Pasteruella multocida from avian species. Avian Dis 16 (4):925-936
13. Peng Z, Liang W, Wu B (2016) [Molecular typing methods for Pasteurella multocida-A review]. Wei Sheng Wu Xue Bao 56 (10):1521-1529
14. Townsend KM, Boyce JD, Chung JY, Frost AJ, Adler B (2001) Genetic organization of Pasteurella multocida cap Loci and development of a multiplex capsular PCR typing system. J Clin Microbiol 39 (3):924-929. doi:10.1128/jcm.39.3.924-929.2001
15. Harper M, John M, Turni C, Edmunds M, St Michael F, Adler B, Blackall PJ, Cox AD, Boyce JD (2015) Development of a rapid multiplex PCR assay to genotype Pasteurella multocida strains by use of the lipopolysaccharide outer core biosynthesis locus. J Clin Microbiol 53 (2):477-485. doi:10.1128/jcm.02824-14
16. Davies RL, MacCorquodale R, Reilly S (2004) Characterisation of bovine strains of Pasteurella multocida and comparison with isolates of avian, ovine and porcine origin. Vet Microbiol 99 (2):145-158. doi:10.1016/j.vetmic.2003.11.013

17. Subaaharan S, Blackall LL, Blackall PJ (2010) Development of a multi-locus sequence typing scheme for avian isolates of Pasteurella multocida. Vet Microbiol 141 (3-4):354-361. doi:10.1016/j.vetmic.2010.01.017
18. Garcia-Alvarez A, Vela AI, San Martin E, Chaves F, Fernandez-Garayzabal JF, Lucas D, Cid D (2017) Characterization of Pasteurella multocida associated with ovine pneumonia using multi-locus sequence typing (MLST) and virulence-associated gene profile analysis and comparison with porcine isolates. Vet Microbiol 204:180-187. doi:10.1016/j.vetmic.2017.04.015
19. Peng Z, Wang H, Liang W, Chen Y, Tang X, Chen H, Wu B (2018) A capsule/lipopolysaccharide/MLST genotype D/L6/ST11 of Pasteurella multocida is likely to be strongly associated with swine respiratory disease in China. Arch Microbiol 200 (1):107-118. doi:10.1007/s00203-017-1421-y
20. Li Z, Cheng F, Lan S, Guo J, Liu W, Li X, Luo Z, Zhang M, Wu J, Shi Y (2018) Investigation of genetic diversity and epidemiological characteristics of Pasteurella multocida isolates from poultry in southwest China by population structure, multi-locus sequence typing and virulence-associated gene profile analysis. J Vet Med Sci 80 (6):921-929. doi:10.1292/jvms.18-0049
21. Devi LB, Bora DP, Das SK, Sharma RK, Mukherjee S, Hazarika RA (2018) Virulence gene profiling of porcine Pasteurella multocida isolates of Assam. Vet World 11 (3):348-354. doi:10.14202/vetworld.2018.348-354
22. Massacci FR, Magistrali CF, Cucco L, Curcio L, Bano L, Mangili P, Scoccia E, Bisgaard M, Aalbaek B, Christensen H (2018) Characterization of Pasteurella multocida involved in rabbit infections. Vet Microbiol 213:66-72. doi:10.1016/j.vetmic.2017.11.023
23. Peng Z, Liang W, Wang F, Xu Z, Xie Z, Lian Z, Hua L, Zhou R, Chen H, Wu B (2018) Genetic and Phylogenetic Characteristics of Pasteurella multocida Isolates From Different Host Species. Front Microbiol 9:1408. doi:10.3389/fmicb.2018.01408
24. Peng Z, Liang W, Liu W, Wu B, Tang B, Tan C, Zhou R, Chen H (2016) Genomic characterization of Pasteurella multocida HB01, a serotype A bovine isolate from China. Gene 581 (1):85-93. doi:10.1016/j.gene.2016.01.041
25. Liu W, Yang M, Xu Z, Zheng H, Liang W, Zhou R, Wu B, Chen H (2012) Complete genome sequence of Pasteurella multocida HN06, a toxigenic strain of serogroup D. J Bacteriol 194 (12):3292-3293. doi:10.1128/jb.00215-12
26. Peng Z, Liang W, Wang Y, Liu W, Zhang H, Yu T, Zhang A, Chen H, Wu B (2017) Experimental pathogenicity and complete genome characterization of a pig origin Pasteurella multocida serogroup F isolate HN07. Vet Microbiol 198:23-33. doi:10.1016/j.vetmic.2016.11.028
27. Ujvári B, Makrai L, Magyar T (2019) Virulence gene profiling and ompA sequence analysis of Pasteurella multocida and their correlation with host species. Vet Microbiol 233:190-195. doi:10.1016/j.vetmic.2019.05.005
28. Stoesser N, Sheppard AE, Pankhurst L, De Maio N, Moore CE, Sebra R, Turner P, Anson LW, Kasarskis A, Batty EM, Kos V, Wilson DJ, Phetsouvanh R, Wyllie D, Sokurenko E, Manges AR, Johnson TJ, Price LB, Peto TE, Johnson JR, Didelot X, Walker AS, Crook DW (2016) Evolutionary History of the Global

Emergence of the Escherichia coli Epidemic Clone ST131. mBio 7 (2):e02162. doi:10.1128/mBio.02162-15

29. Massacci FR, Magistrali CF, Cucco L, Curcio L, Bano L, Mangili P, Scoccia E, Bisgaard M, Aalbæk B, Christensen H (2018) Characterization of Pasteurella multocida involved in rabbit infections. Vet Microbiol 213:66-72. doi:10.1016/j.vetmic.2017.11.023
30. Ewers C, Lübke-Becker A, Bethe A, Kiebling S, Filter M, Wieler LH (2006) Virulence genotype of Pasteurella multocida strains isolated from different hosts with various disease status. Vet Microbiol 114 (3-4):304-317. doi:10.1016/j.vetmic.2005.12.012
31. Shivachandra SB, Viswas KN, Kumar AA (2011) A review of hemorrhagic septicemia in cattle and buffalo. Anim Health Res Rev 12 (1):67-82. doi:10.1017/s146625231100003x
32. Welsh RD, Dye LB, Payton ME, Confer AW (2004) Isolation and antimicrobial susceptibilities of bacterial pathogens from bovine pneumonia: 1994–2002. J Vet Diagn Invest 16 (5):426-431. doi:10.1177/104063870401600510
33. Hotchkiss EJ, Hodgson JC, Lainson FA, Zadoks RN (2011) Multilocus sequence typing of a global collection of Pasteurella multocida isolates from cattle and other host species demonstrates niche association. BMC Microbiol 11:115. doi:10.1186/1471-2180-11-115
34. Tang X, Zhao Z, Hu J, Wu B, Cai X, He Q, Chen H (2009) Isolation, antimicrobial resistance, and virulence genes of Pasteurella multocida strains from swine in China. J Clin Microbiol 47 (4):951-958. doi:10.1128/jcm.02029-08
35. Hurtado R, Maturrano L, Azevedo V, Aburjaile F (2020) Pathogenomics insights for understanding Pasteurella multocida adaptation. Int J Med Microbiol 310 (4):151417. doi:10.1016/j.ijmm.2020.151417
36. Peng Z, Liang W, Liu W, Chen H, Wu B (2017) Genome characterization of Pasteurella multocida subspecies septica and comparison with Pasteurella multocida subspecies multocida and gallicida. Arch Microbiol 199 (4):635-640. doi:10.1007/s00203-017-1341-x
37. Khamesipour F, Momtaz H, Azhdary Mamoreh M (2014) Occurrence of virulence factors and antimicrobial resistance in Pasteurella multocida strains isolated from slaughter cattle in Iran. Front Microbiol 5:536. doi:10.3389/fmicb.2014.00536

Figures

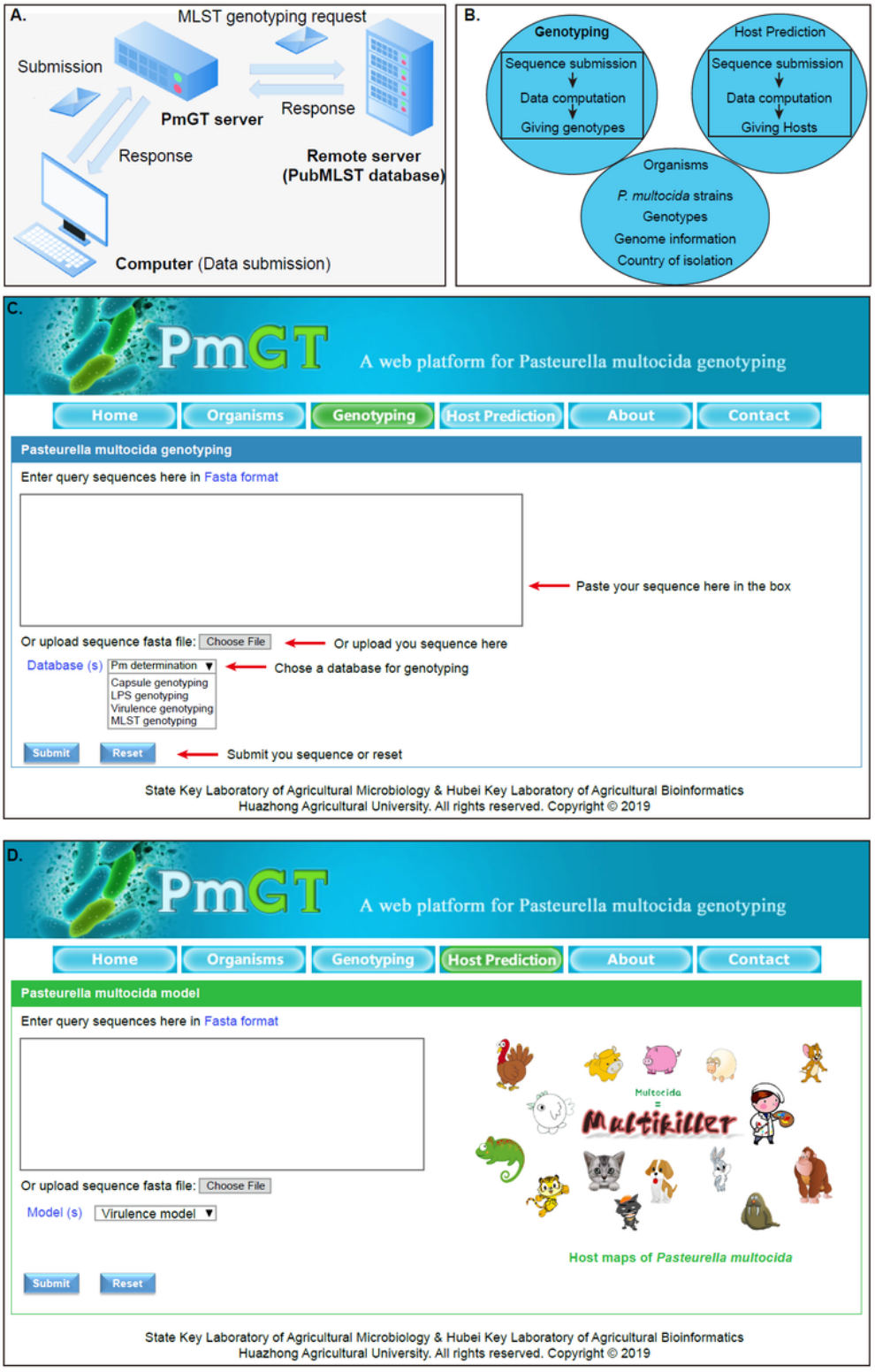

Figure 1

Development of the P. multocida genotyping and host prediction platform. (A.) Flowchart showing the system design; (B.) Main functions of the web platform; (C.) Overview of the genotyping system of P. multocida; (D.) Overview of the host prediction system of P. multocida.

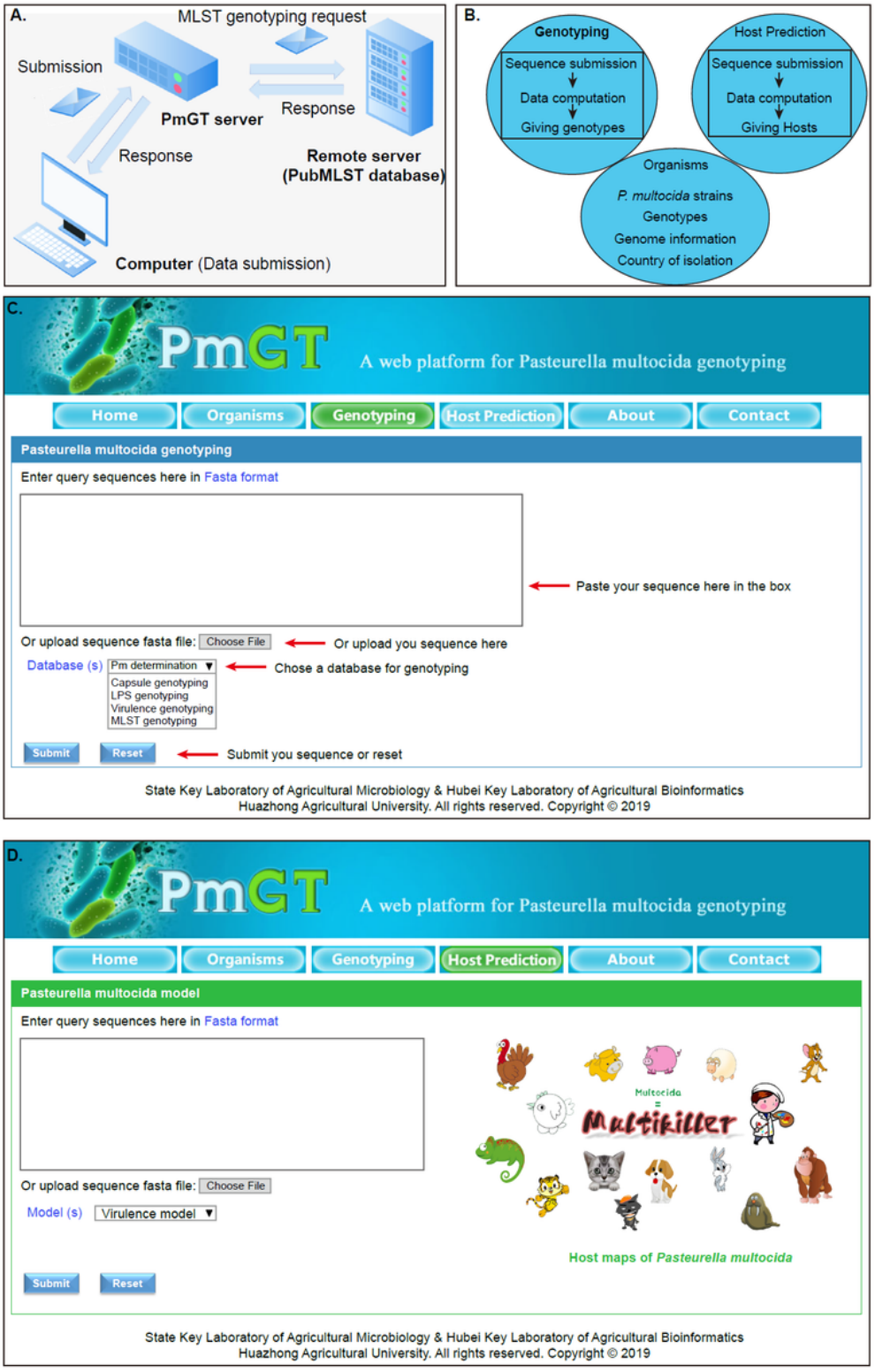

Figure 1

Development of the P. multocida genotyping and host prediction platform. (A.) Flowchart showing the system design; (B.) Main functions of the web platform; (C.) Overview of the genotyping system of P. multocida; (D.) Overview of the host prediction system of P. multocida.

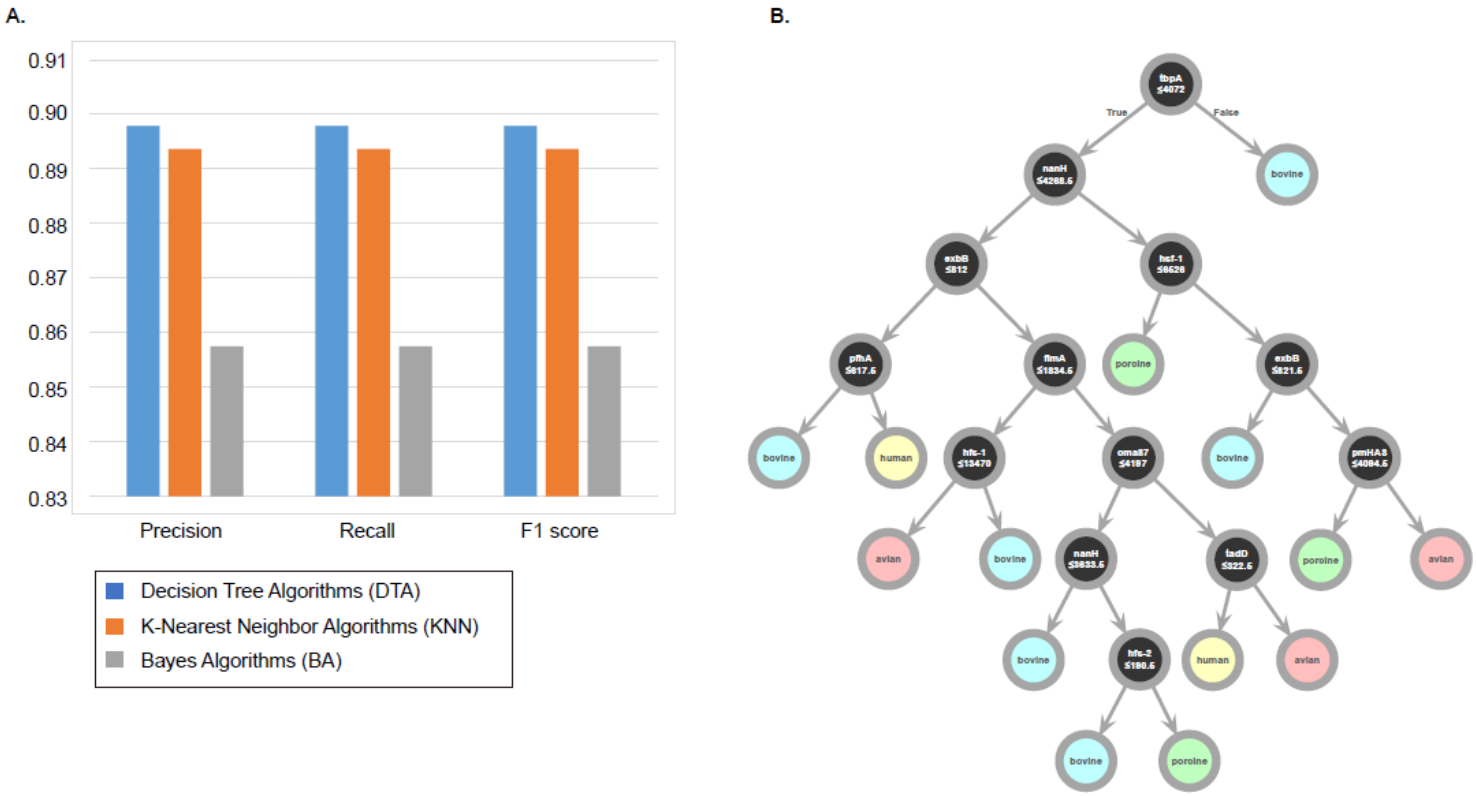

Figure 2

Development of a host tropism prediction system for P. multocida by different machine learning algorithms. (A.) Column chart revealing the average values of precision, the recall, and F1 score (PRF) determined by micro-eDTA/gDTA, macro-eDTA/gDTA, and weighted-eDTA/gDTA; (B.) A decision tree generated by the entroy Decision Tree Algorithms to reveal the association of P. multocida VFGs and host species. Green, red, blue, and yellow nodes refer to genome sequences of porcine, avian, bovine, and human isolates, respectively. Black nodes refer to the determination rules in which the left side is the blast score for the association VFGs and the right side is the thersholds.

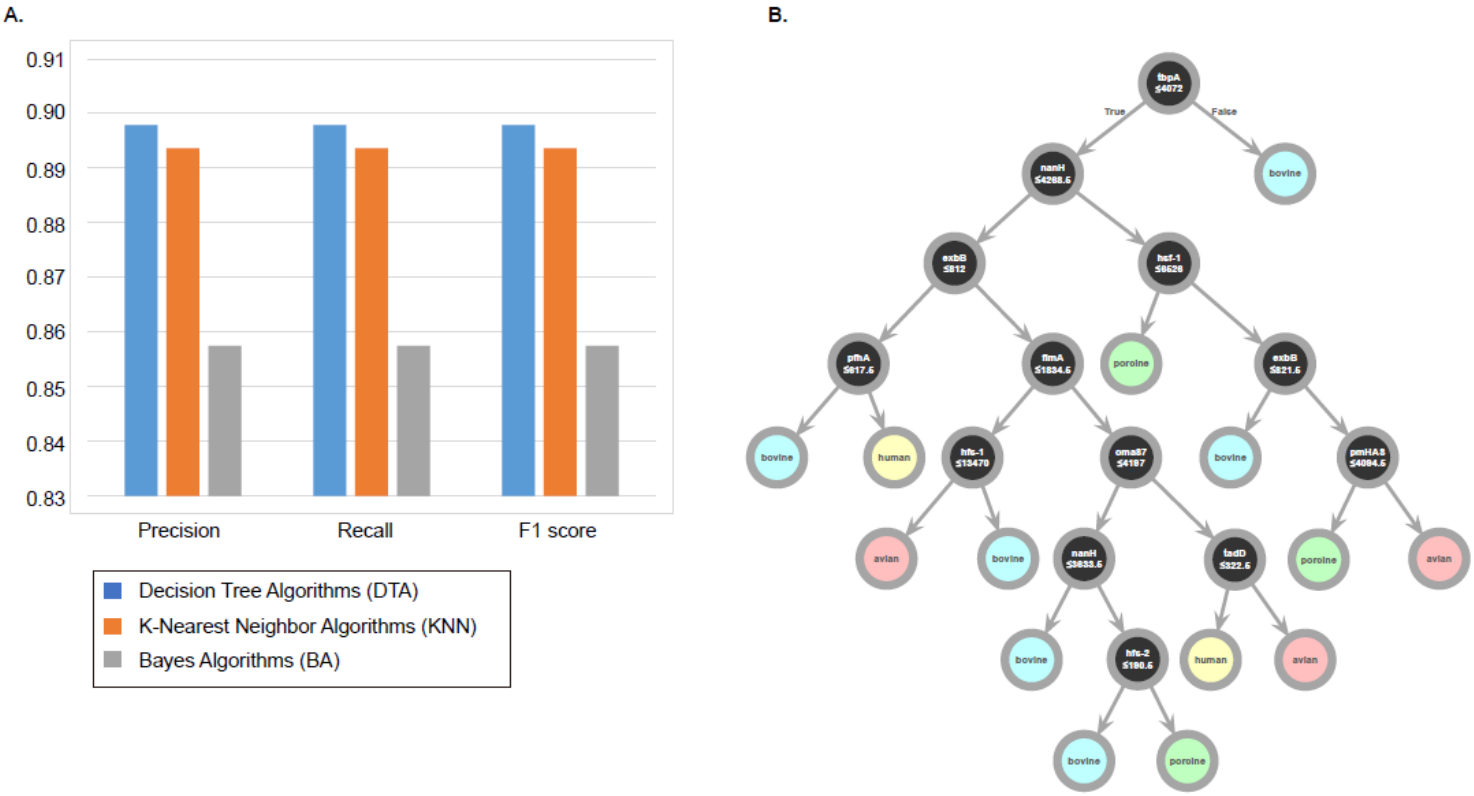

Figure 2

Development of a host tropism prediction system for P. multocida by different machine learning algorithms. (A.) Column chart revealing the average values of precision, the recall, and F1 score (PRF) determined by micro-eDTA/gDTA, macro-eDTA/gDTA, and weighted-eDTA/gDTA; (B.) A decision tree generated by the entroy Decision Tree Algorithms to reveal the association of P. multocida VFGs and host species. Green, red, blue, and yellow nodes refer to genome sequences of porcine, avian, bovine, and human isolates, respectively. Black nodes refer to the determination rules in which the left side is the blast score for the association VFGs and the right side is the thersholds.

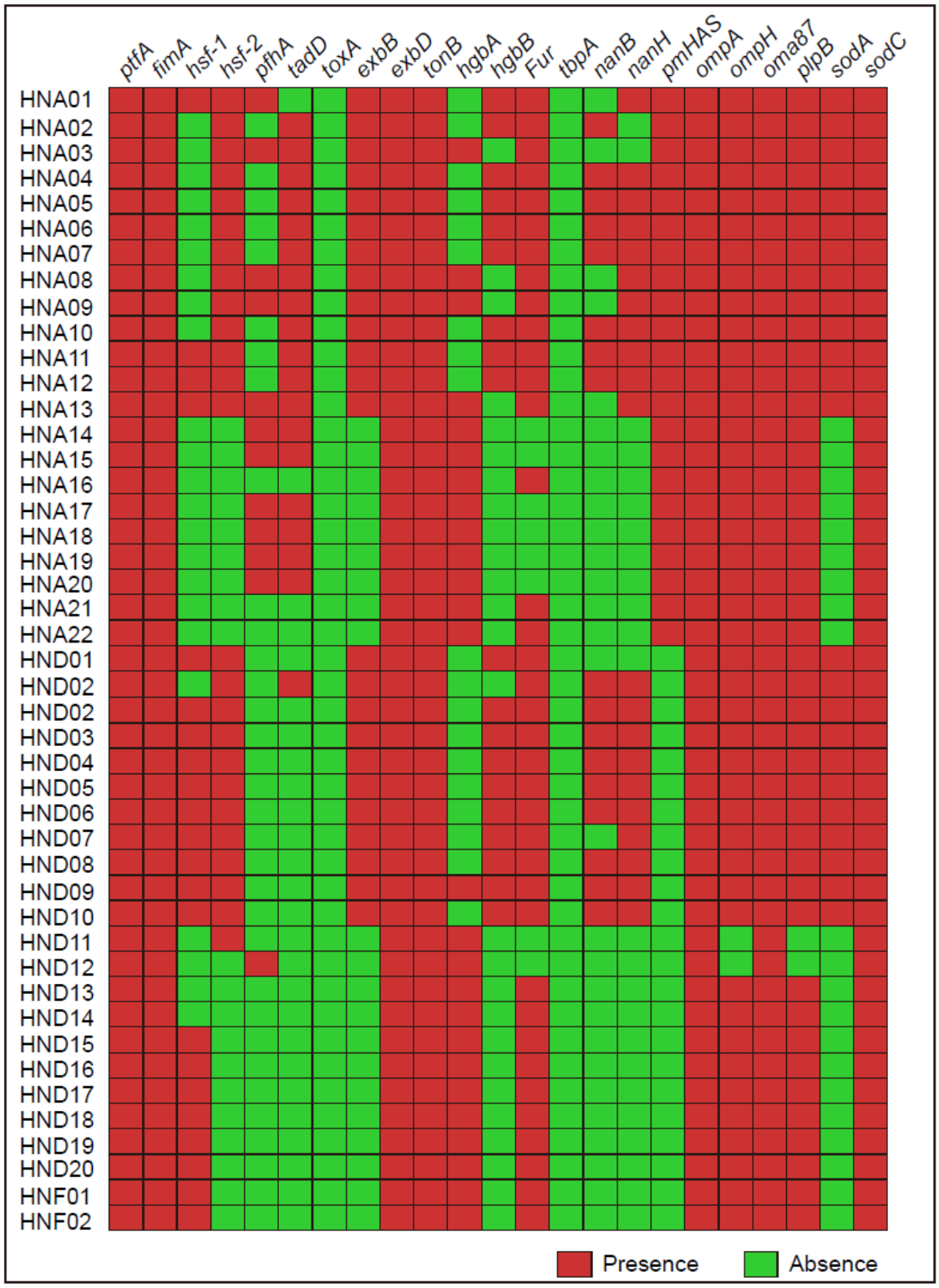

Figure 3

Heatmap showing the distribution of the 23 types of virulence genes (VFGs) among the 45 P. multocida strains from pigs. Boxes in red indicate a VFG is presence in the strain while boxes in green represent a VFG is missing in the strain.

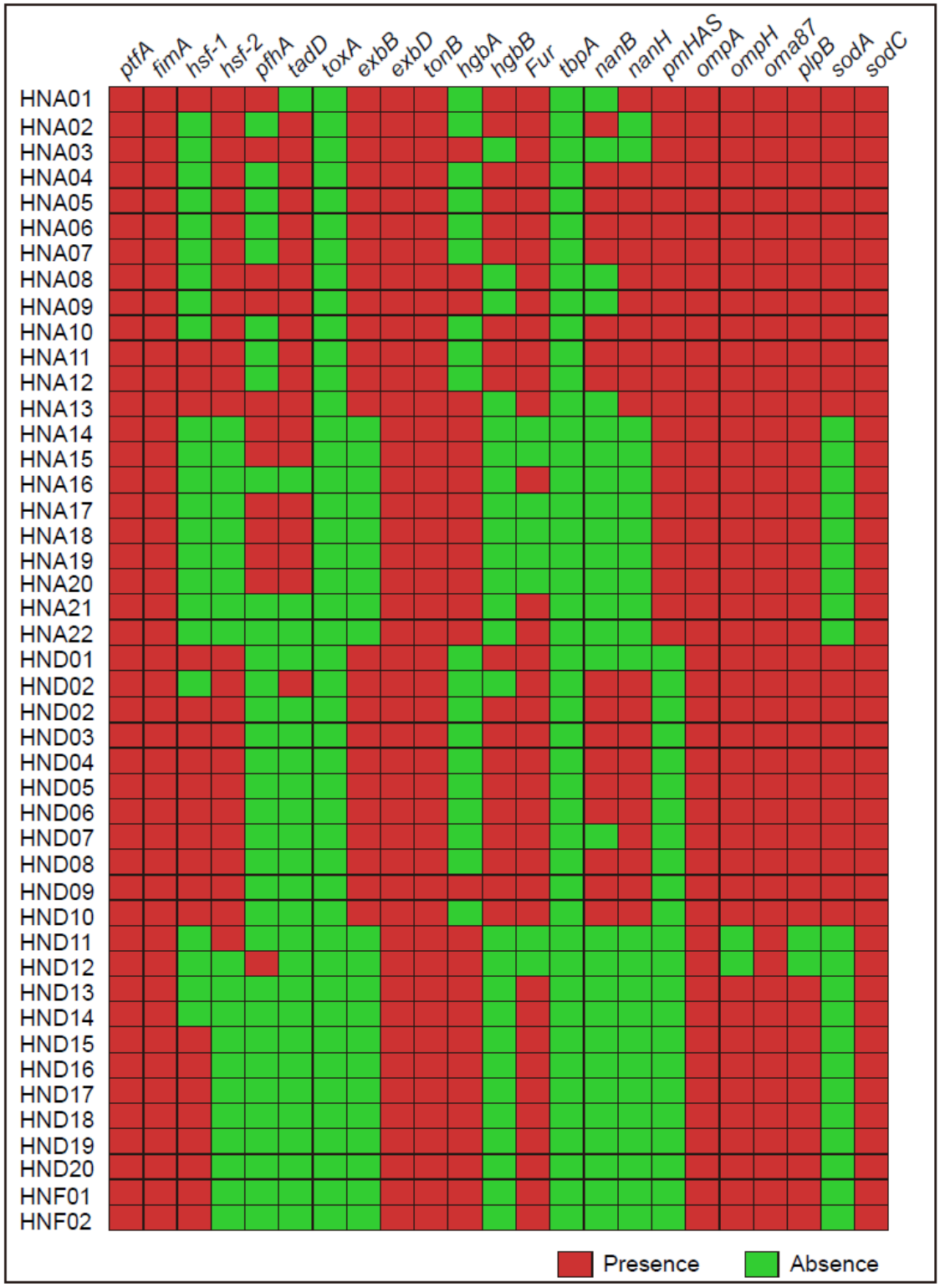

Figure 3

Heatmap showing the distribution of the 23 types of virulence genes (VFGs) among the 45 P. multocida strains from pigs. Boxes in red indicate a VFG is presence in the strain while boxes in green represent a VFG is missing in the strain.

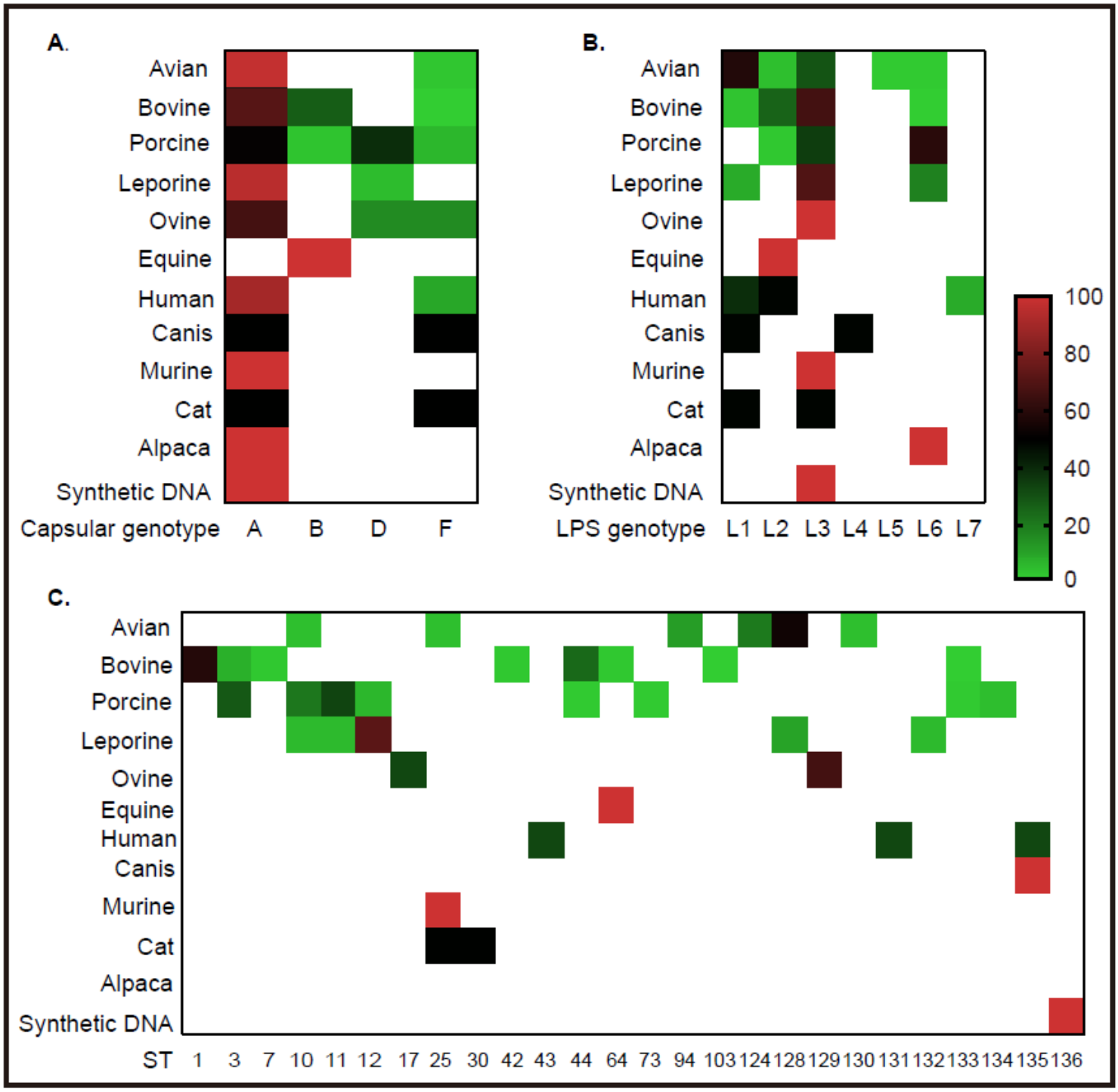

Figure 4

Heatmap revealing the association between capsular/LPS/MLST genotypes and P. multocida strains from different host species determined by PmGT. (A.) Heatmap revealing the association between capsular genotypes and P. multocida strains from different host species; (B.) Heatmap revealing the association between LPS genotypes and P. multocida strains from different host species; (C.) Heatmap revealing the association between MLST genotypes and P. multocida strains from different host species. Percentages of sequences typed are shown with different colors displayed at right corner.

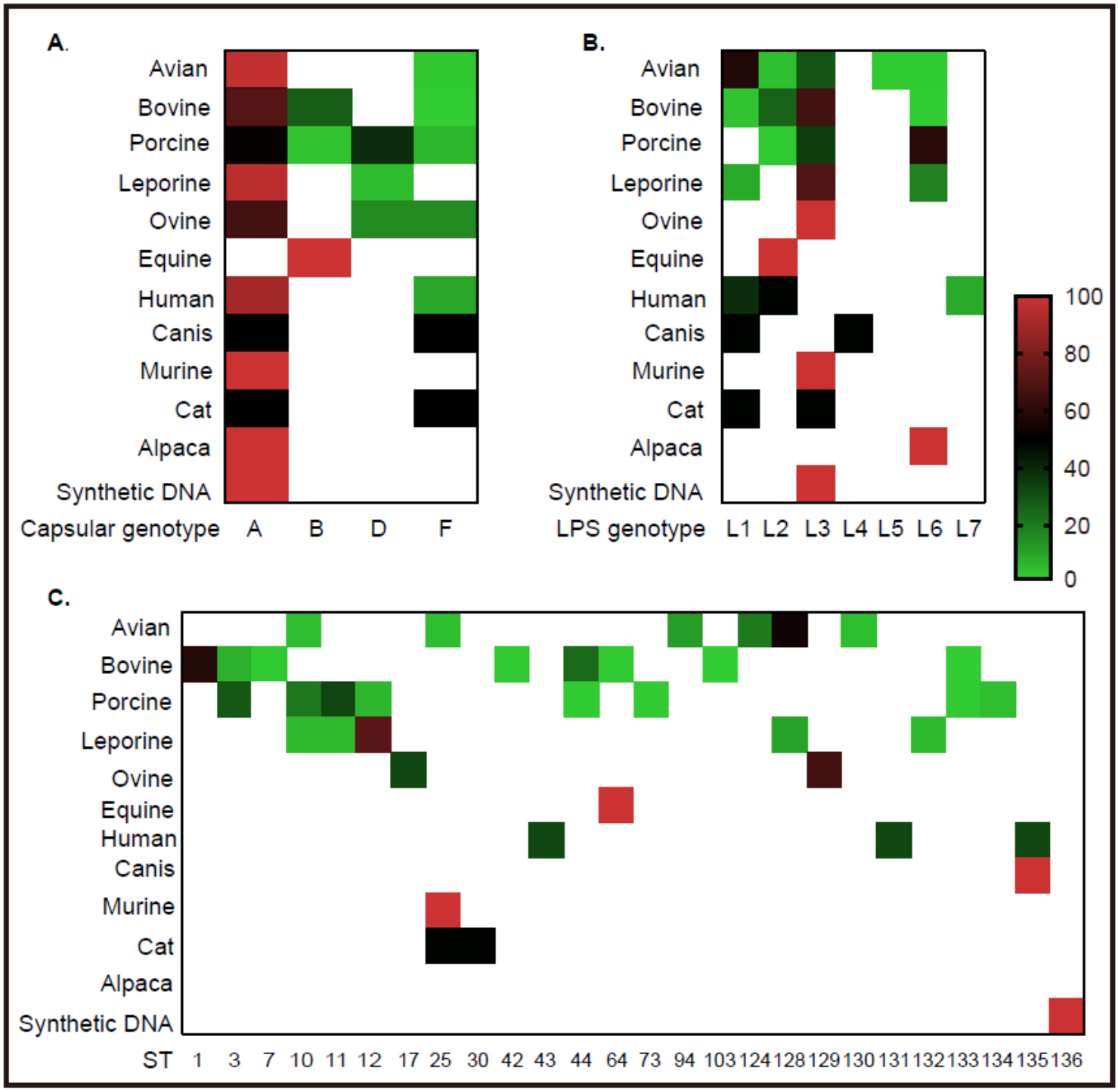

Figure 4

Heatmap revealing the association between capsular/LPS/MLST genotypes and P. multocida strains from different host species determined by PmGT. (A.) Heatmap revealing the association between capsular genotypes and P. multocida strains from different host species; (B.) Heatmap revealing the association between LPS genotypes and P. multocida strains from different host species; (C.) Heatmap revealing the association between MLST genotypes and P. multocida strains from different host species. Percentages of sequences typed are shown with different colors displayed at right corner.

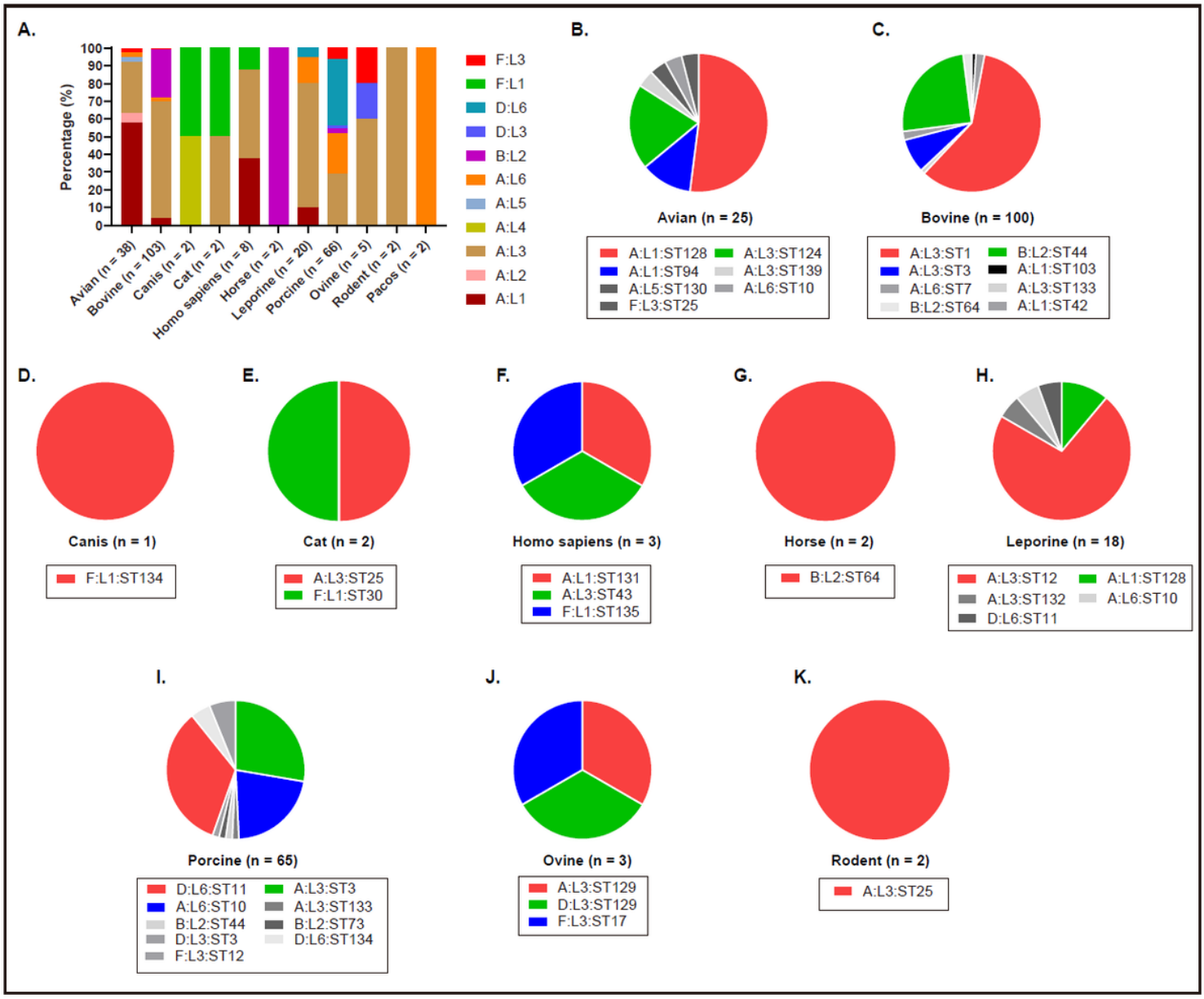

Figure 5

Column and pie charts showing the distribution of capsular: LPS genotypes and/or the capsular: LPS: MLST genotypes of P. multocida strains from different host species determined by PmGT by using the whole genome sequences. (A.) Column chart showing the distribution of capsular: LPS genotypes of P. multocida strains from different host species; (B.)~(K.) Pie charts showing the distribution of capsular: LPS: MLST genotypes of P. multocida strains from avian species, bovine species, canis, cats, humans, horses, leporine species, pigs, ovine species, and rodents, respectively.

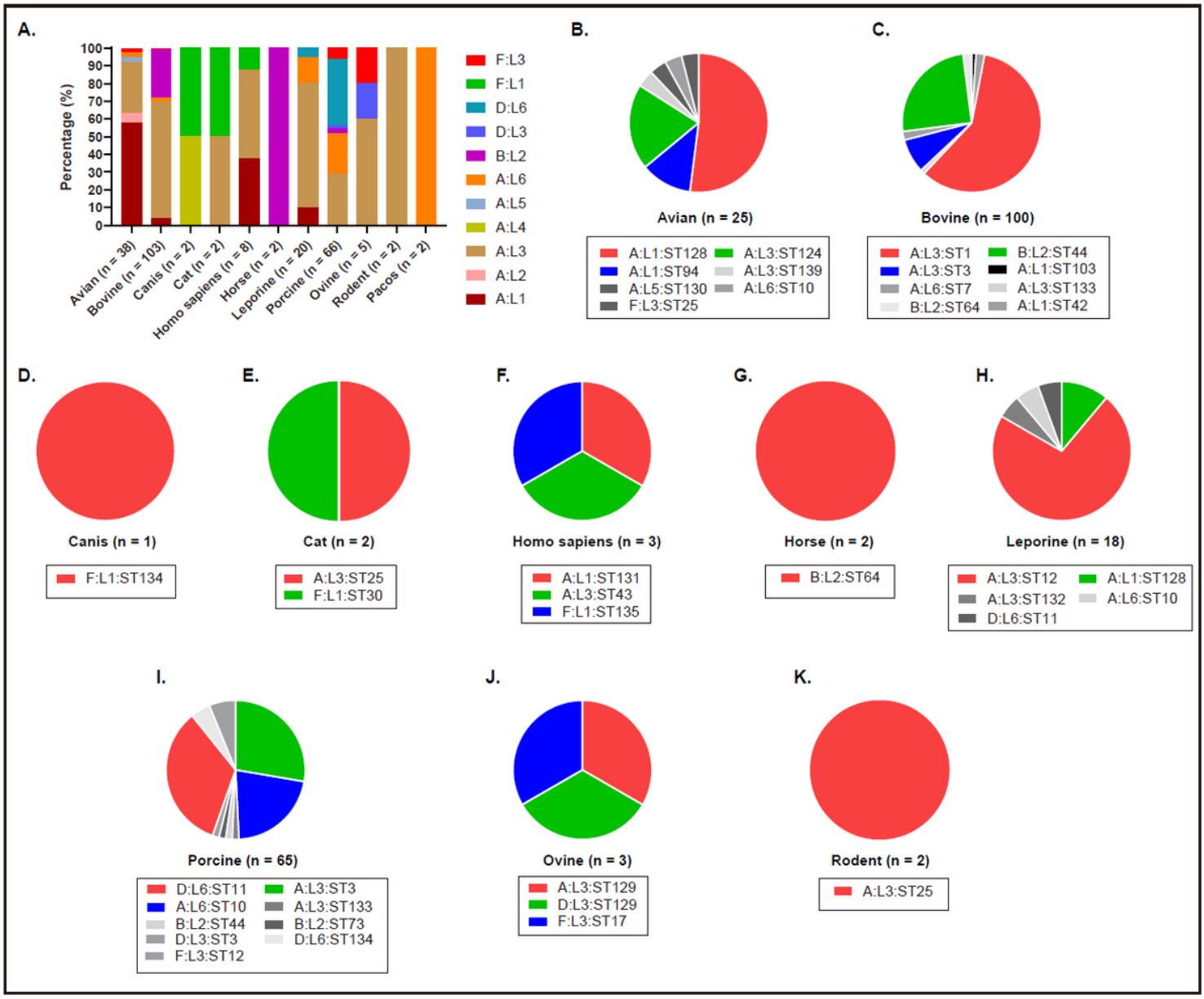

Figure 5

Column and pie charts showing the distribution of capsular: LPS genotypes and/or the capsular: LPS: MLST genotypes of P. multocida strains from different host species determined by PmGT by using the whole genome sequences. (A.) Column chart showing the distribution of capsular: LPS genotypes of P. multocida strains from different host species; (B.)~(K.) Pie charts showing the distribution of capsular: LPS: MLST genotypes of P. multocida strains from avian species, bovine species, canis, cats, humans, horses, leporine species, pigs, ovine species, and rodents, respectively.

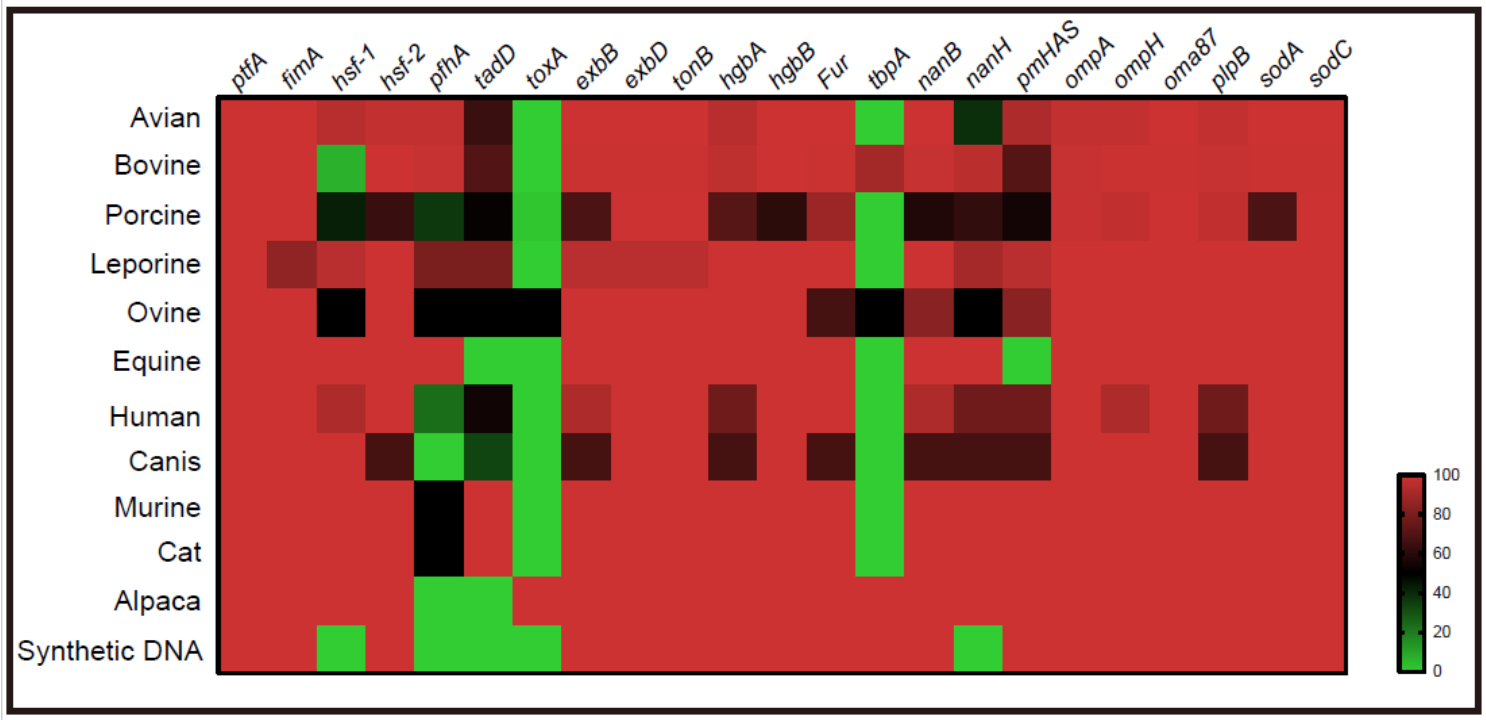

Figure 6

Heatmap revealing the association between virulence genes and P. multocida strains from different host species.

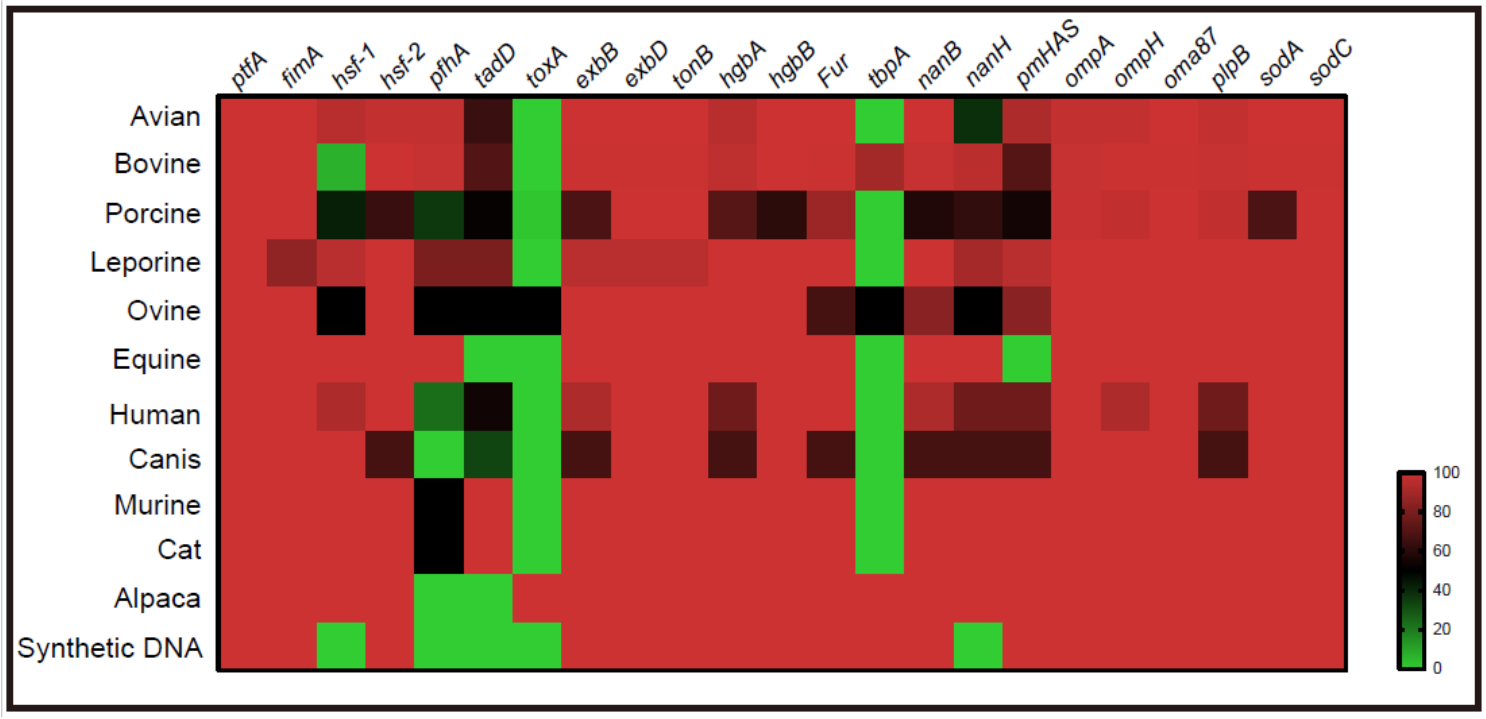

Figure 6

Heatmap revealing the association between virulence genes and P. multocida strains from different host species.

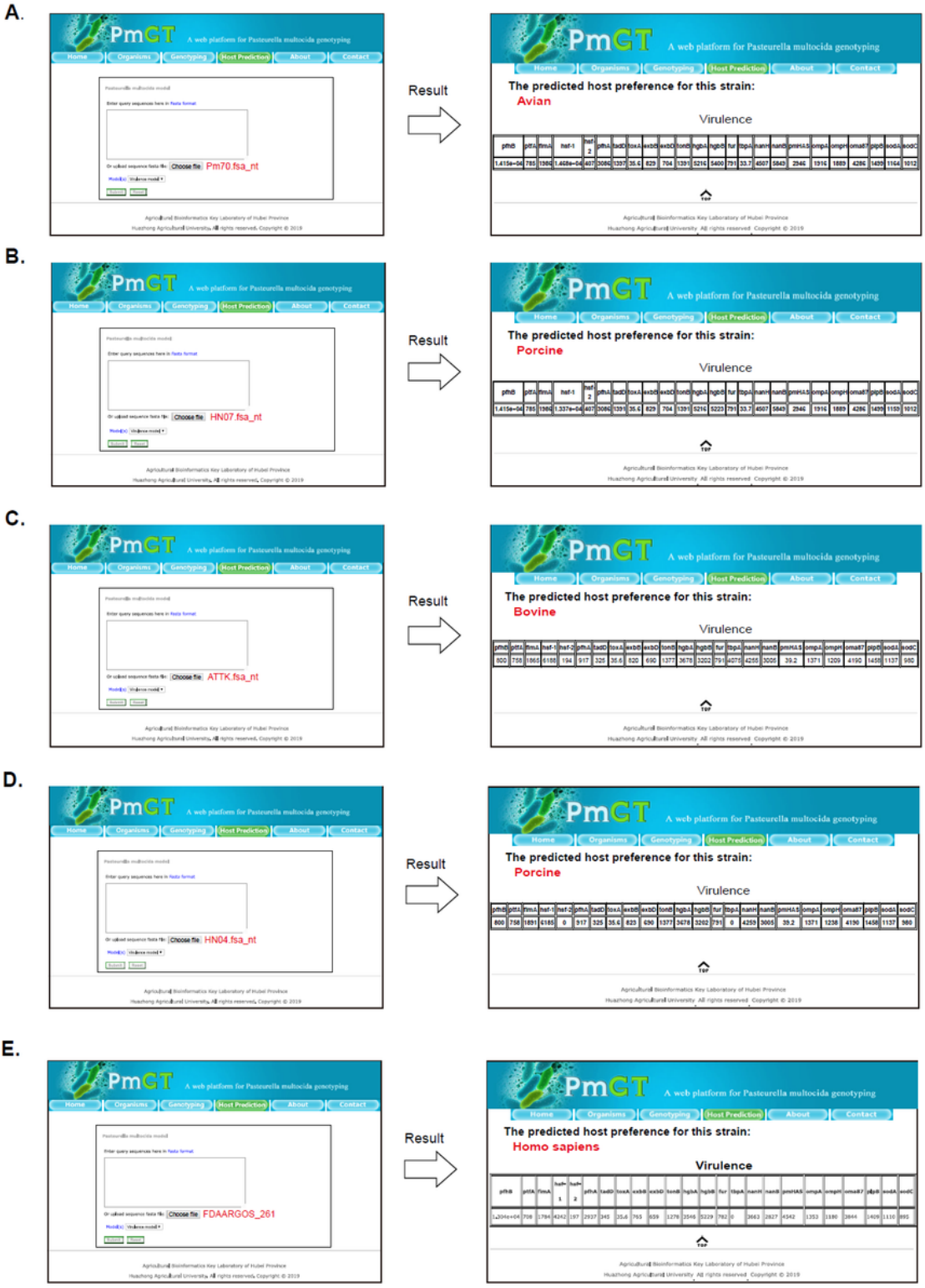

Figure 7

Testing of the developed model predicting the host species of P. multocida. (A.) Test by using the whole genome sequence of P. multocida Pm70 (avian isolate, genotype F:L3:ST25, GenBank accession no. AE004439); (B.) Test by using the whole genome sequence of P. multocida HN07 (porcine isolate, genotype F:L3:ST12, GenBank accession no. CP007040); (C.) Test by using the whole genome sequence of P. multocida ATTK (bovine isolate, genotype B:L2:ST44, GenBank accession no. JQEA00000000); (D.)

Test by using the whole genome sequence of P. multocida HN04 (porcine isolate, genotype B:L2:ST44, GenBank accession no. PPVE00000000); (E.) Test by using the whole genome sequence of P. multocida FDAARGOS_261 (human isolate, genotype A:L1:ST94, GenBank accession no. CP020403). The whole genome sequences were submitted to the PmGT platform using the Host prediction function. Results yielded based on VFG profile with host prediction

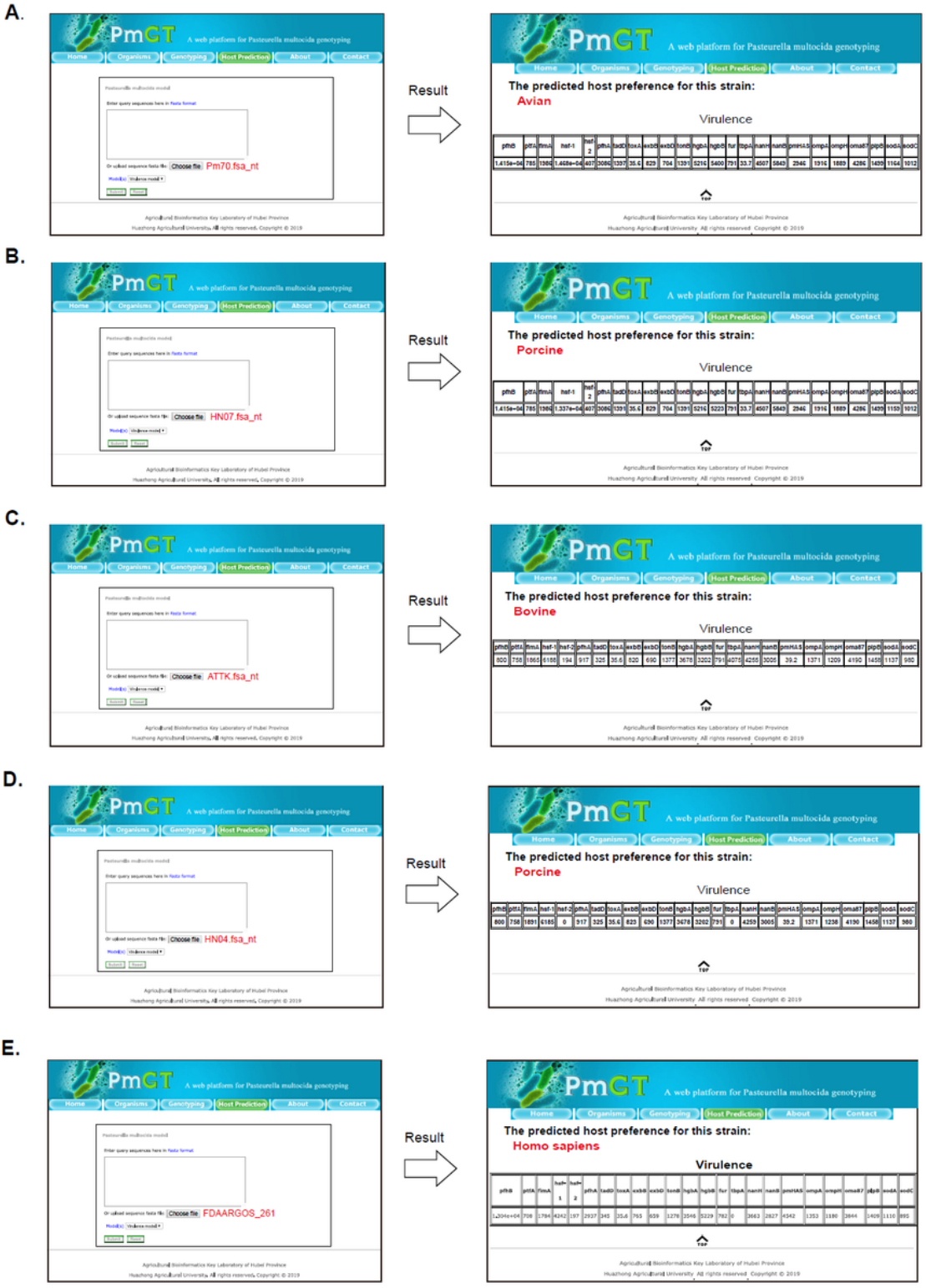

Figure 7

Testing of the developed model predicting the host species of P. multocida. (A.) Test by using the whole genome sequence of P. multocida Pm70 (avian isolate, genotype F:L3:ST25, GenBank accession no. AE004439); (B.) Test by using the whole genome sequence of P. multocida HN07 (porcine isolate, genotype F:L3:ST12, GenBank accession no. CP007040); (C.) Test by using the whole genome sequence of P. multocida ATTK (bovine isolate, genotype B:L2:ST44, GenBank accession no. JQEA00000000); (D.) Test by using the whole genome sequence of P. multocida HN04 (porcine isolate, genotype B:L2:ST44, GenBank accession no. PPVE00000000); (E.) Test by using the whole genome sequence of P. multocida FDAARGOS_261 (human isolate, genotype A:L1:ST94, GenBank accession no. CP020403). The whole genome sequences were submitted to the PmGT platform using the Host prediction function. Results yielded based on VFG profile with host prediction

Supplementary Files

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

- FigureS1.pdf
- FigureS1.pdf
- FigureS2.pdf
- FigureS2.pdf
- FigureS3.pdf
- FigureS3.pdf
- TableS1.xlsx
- TableS1.xlsx
- Txt1.txt
- Txt1.txt