AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Clinical Microbiology®

CLINICAL VETERINARY MICROBIOLOGY

"Pathotyping" Multiplex PCR Assay for *Haemophilus parasuis*: a Tool for Prediction of Virulence

Kate J. Howell,[a]* Lucy A. Weinert,[a] Sarah E. Peters,[a] Jinhong Wang,[a] Juan Hernandez-Garcia,[a] Roy R. Chaudhuri,[b] Shi-Lu Luan,[a] Øystein Angen,[c] Virginia Aragon,[d] Susanna M. Williamson,[e] Paul R. Langford,[f] Andrew N. Rycroft,[g] Brendan W. Wren,[h] Duncan J. Maskell,[a] Alexander W. Tucker,[a] on behalf of the BRaDP1T Consortium

Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom[a]; Department of Molecular Biology and Biotechnology, University of Sheffield, Western Bank, Sheffield, United Kingdom[b]; Norwegian Veterinary Institute, Oslo, Norway[c]; Centre de Recerca en Sanitat Animal (CReSA), UAB-IRTA, Campus de la Universitat Autònoma de Barcelona, 08193, Institut de Recerca i Tecnologia Agroalimentàries (IRTA), Barcelona, Spain[d]; Animal and Plant Health Agency (APHA), Bury St. Edmunds, Suffolk, United Kingdom[e]; Section of Paediatrics, Department of Medicine, Imperial College London, St. Mary's Campus, London, United Kingdom[f]; The Royal Veterinary College, Hawkshead Campus, Hatfield, Hertfordshire, United Kingdom[g]; Faculty of Infectious & Tropical Diseases, London School of Hygiene & Tropical Medicine, London, United Kingdom[h]

ABSTRACT *Haemophilus parasuis* is a diverse bacterial species that is found in the upper respiratory tracts of pigs and can also cause Glässer's disease and pneumonia. A previous pangenome study of *H. parasuis* identified 48 genes that were associated with clinical disease. Here, we describe the development of a generalized linear model (termed a pathotyping model) to predict the potential virulence of isolates of *H. parasuis* based on a subset of 10 genes from the pangenome. A multiplex PCR (mPCR) was constructed based on these genes, the results of which were entered into the pathotyping model to yield a prediction of virulence. This new diagnostic mPCR was tested on 143 field isolates of *H. parasuis* that had previously been whole-genome sequenced and a further 84 isolates from the United Kingdom from cases of *H. parasuis*-related disease in pigs collected between 2013 and 2014. The combination of the mPCR and the pathotyping model predicted the virulence of an isolate with 78% accuracy for the original isolate collection and 90% for the additional isolate collection, providing an overall accuracy of 83% (81% sensitivity and 93% specificity) compared with that of the "current standard" of detailed clinical metadata. This new pathotyping assay has the potential to aid surveillance and disease control in addition to serotyping data.

KEYWORDS *Haemophilus parasuis*, virulence factors, pathotyping, molecular diagnostics

Received 16 December 2016 **Returned for modification** 17 January 2017 **Accepted** 28 May 2017

Accepted manuscript posted online 14 June 2017

Citation Howell KJ, Weinert LA, Peters SE, Wang J, Hernandez-Garcia J, Chaudhuri RR, Luan S-L, Angen Ø, Aragon V, Williamson SM, Langford PR, Rycroft AN, Wren BW, Maskell DJ, Tucker AW, on behalf of the BRaDP1T Consortium. 2017. "Pathotyping" multiplex PCR assay for *Haemophilus parasuis*: a tool for prediction of virulence. J Clin Microbiol 55:2617–2628. https://doi.org/10.1128/JCM.02464-16.

Editor Brad Fenwick, University of Tennessee at Knoxville

Copyright © 2017 American Society for Microbiology. All Rights Reserved.

Address correspondence to Kate J. Howell, kjh52@cam.ac.uk.

* Present address: Kate J. Howell, University Department of Paediatrics, University of Cambridge, Cambridge, United Kinddom.

Haemophilus parasuis is a diverse Gram-negative bacterial species commonly found as a commensal in the upper respiratory tract (URT) of the pig. Some isolates of this bacterium can cause pneumonia as well as a systemic disease of pigs commonly known as Glässer's disease, with more severe presentations including arthritis, meningitis, polyserositis, and septicemia (1–5). The U.K. Animal and Plant Health Agency (APHA) reported an increase in *H. parasuis*-related outbreaks over the last few years (6), with 15 outbreaks involving Glässer's disease and seven from pneumonia due to *H. parasuis* in 2015. *H. parasuis* can cause disease throughout the life cycle of the pig, affecting nursery herds (commonly alongside porcine reproductive and respiratory syndrome

Downloaded from http://jcm.asm.org/ on December 14, 2017 by LONDON SCHOOL OF HYGIENE & TROPICAL MED

virus [7]) as well as contributing to multifactorial porcine respiratory disease complex (PRDC) in grower-finisher pigs (8).

Species-level identification of *H. parasuis* can be useful diagnostically but is not sufficient for all situations. For example, the presence of *H. parasuis* in a swab from the nose or from oral fluid is not very informative, as this bacterium is a common member of the flora of the URT. Therefore, additional information of this bacterium would be very useful to allow for more thorough surveillance. Several subtyping methods or epidemiological tools are available, with serotyping (9–11) being the most commonly used for this bacterium. However, serovar is regarded as a poor proxy for virulence, with the exception of serovars 4 and 5 (12–14). The *vtaA* multiplex PCR (mPCR) (15–17) is the only method that has shown an association between the test result for a given isolate and its virulence, but its wide-scale use has not been reported.

The link between virulence and serovar for *H. parasuis* is predominantly based on early experimental reproduction of disease in specific-pathogen free (SPF) pigs using the reference strains (with somewhat inconsistent results) and the globally prevalent disease-causing isolates (1, 9, 13, 18). Serovars 4 and 5 are currently the most prevalent disease-causing serovars of *H. parasuis* globally (12–14). However, generalization of the virulence of a serovar is often made for all isolates based on SPF challenge studies performed with small numbers of isolates and animals (1, 9, 18–21). Even the serovar 3 reference strain, widely considered to be avirulent, resulted in some clinical signs in a recent challenge study using colostrum-deprived piglets (22). Therefore, a better marker for virulence or understanding of virulence potential is required for this bacterium.

Our previous analysis of more than 200 isolates of *H. parasuis* (23) revealed a diverse pangenome of over 7,000 genes, with 1,049 genes classified as part of the core genome (i.e., present in every isolate). However, no links between previously suggested virulence factors and clinical metadata were found, i.e., no previously suggested virulence factors were present in a high proportion of isolates from one category while absent from the other and, consequently, were not useful as indicators of the virulence of an isolate from the pangenome. Therefore, we used a genome-wide association study (GWAS) approach (23) using discriminant analysis of principal components (24), which identified 48 genes associated with disease, which could be found in either disease-associated or carriage isolates. While the relative importance in virulence and the functional roles of these newly designated virulence-associated genes have yet to be investigated, they have the potential to aid in the prediction of the virulence of an isolate based purely on their correlation with the virulent phenotype.

The definition of virulence for our studies has been carefully considered. The isolates used in our previous pangenome analysis (23) were predominantly field isolates with detailed clinical metadata based on postmortem results (e.g., tissue of origin, serovar, age of pig, antemortem signs, postmortem diagnosis, known welfare issues, and known coinfections). This information was used to classify isolates of *H. parasuis* obtained from systemic or respiratory sites as disease-associated isolates and all isolates from the URT of healthy pigs as nondisease-associated or "carriage" isolates. In reality, these clinical metadata are not a guaranteed means for determining disease-causing potential, as many other factors may be involved in the outcome of an infection (additional host factors and variation in the microbiota), and so we call the clinical metadata a silver standard data set for comparison.

This paper describes the design and validation of an mPCR, termed a "pathotyping" tool, which can be used to predict the virulence of an isolate based on a subset of 10 genes from the pangenome of *H. parasuis*.

RESULTS

Design of the pathotyping mPCR. The final list of genes included in the model is shown in Table 1 alongside their primers and information regarding their importance in the model. This list included genes that were either positively or negatively associated with virulence and a species-specific marker that was designed based on one of

Downloaded from http://jcm.asm.org/ on December 14, 2017 by LONDON SCHOOL OF HYGIENE & TROPICAL MED

TABLE 1 Summary of isolates used in the validation of the mPCR based on serovar and country of origin

Country	No. of isolates																					
	Disease-associated serovar											Carriage serovar										
	1	2	4	5 or 12	6	7	9	13	14	15	Total	1	2	3	4	5 or 12	6	7	8	10	13	14	Total
Argentina				1							1												
Denmark	1	1	6	5	2	3	1	6	1	1	27												
Germany				2							2												
Greece					1						1												
Italy				1							1												
Japan				1							1		2	2	1								5
Spain	1			5	1	1			1		9	1					1	1	2			2	7
Sweden											0												
Switzerland											0						1						1
United Kingdom	2	7	23	11	2	9	2	8	1	3	68	1	2			3	1	2			3		12
Unknown	2										2	1		1						1			3
United States				1					1	1	3												
Total	6	8	29	27	6	13	3	14	4	5	115	3	4	3	1	3	3	3	2	1	3	2	28

the most conserved genes in the pangenome of *H. parasuis* (25). Two genes (HPS_23879 and HPS_22976c) that were included in the model were found in high proportions (85% and 78%, respectively) in the disease-associated Bayesian analysis of population structure (BAPS) populations (BAPS 4 and BAPS 5) (23). The model mostly included genes of unknown function but also included genes encoding an inosine-5-monophosphate dehydrogenase, an aspartate kinase monofunctional class protein, a helix-turn-helix family protein, and a glycosyltransferase. This pathotyping model had an Akaike information criterion (estimate of the quality of the model) of 156.5, and the area under the curve (AUC) was 0.88 (Fig. 1). The sensitivity and specificity of the model were both estimated at greater than 80% using the receiver operating characteristic (ROC) curve (Fig. 1) and led to the choice of a cutoff 0.72 for the model.

Validation of the pathotyping multiplex. The mPCR was validated using 143 of the original *H. parasuis* isolates, with 115 being disease associated (80%) and 28 being carriage isolates (20%). An example of a gel from the mPCR on 16 different isolates from the additional isolate collection can be found in Fig. 2. A summary of the results of the mPCR for these isolates can be seen in Table 2, and detailed information is provided in Table S1 in the supplemental material. Overall, 88% of these isolates had the amplicon pattern plus or minus one band compared with the expected pattern from the presence or absence of each gene from the draft genome sequences of the isolates. For those isolates where more than one amplicon difference was identified by the mPCR ($n = 34$), the model classified 85% as the same virulence category as the clinical

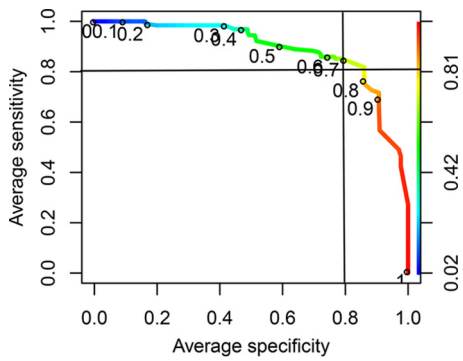

FIG 1 ROC curve for the final model based on the 10 genes chosen for the pathotyping PCR. The curve is labeled with the different cutoffs for the model to achieve the sensitivity and specificity at each point on the curve. The lines represent the desired boundaries for 80% specificity and sensitivity to aid the choice of model cutoff.

Downloaded from http://jcm.asm.org/ on December 14, 2017 by LONDON SCHOOL OF HYGIENE & TROPICAL MED

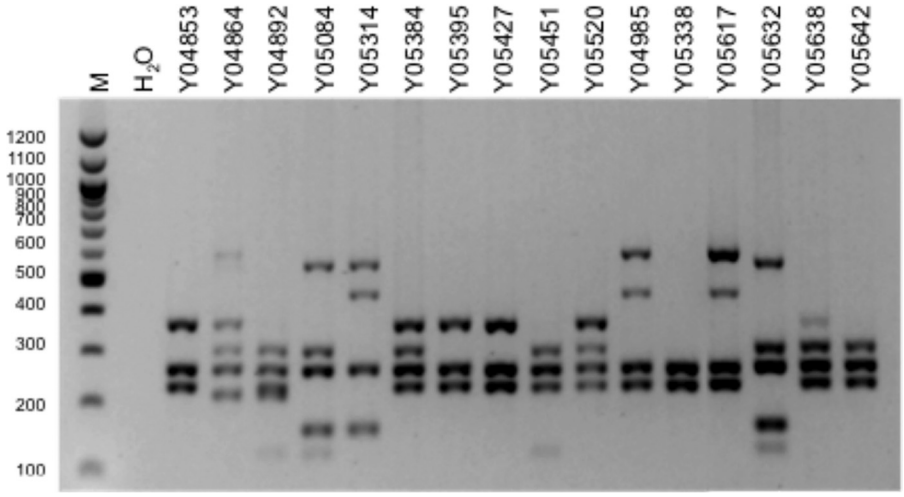

FIG 2 Results of the pathotyping multiplex on a subset of *H. parasuis* isolates from the additional isolate collection. M, Quick-load 100 bp DNA ladder.

metadata. Therefore, only 5 isolates varied by more than one band and had a different virulence prediction from the clinical metadata. Taking the clinical metadata as the "current or silver standard," the sensitivity and specificity of the model and mPCR were calculated at 75% and 93%, respectively, with an overall accuracy of 78%. For the additional clinical isolate collection (Table 2), the amplicon patterns were entered into the model, and 90% were predicted as being disease-associated isolates.

The amplicon patterns were identical whether genomic DNA (gDNA) or colony PCR was used. Of the 11 genes in the mPCR, two were amplified from the negative-control panel (Fig. 3), including HPS_21058 from one of the two isolates of *Actinobacillus porcinus* (HS206) and HPS_23879 from the single isolate of *Actinobacillus indolicus* (HS213), but no species-specific band was produced for any these isolates. Six reference strains of *H. parasuis* were used to test the limit of detection of the mPCR. The average minimum concentration of DNA detectable by the mPCR was determined to be 1.02 ng/μl for an individual pure gDNA preparation or 3.4×10^5 genomes/μl across the isolates (Fig. 4).

DISCUSSION

We have described the development of a pathotyping mPCR assay exploiting the recent discovery of putative virulence-associated genes based on a whole-genome analysis of a large population of *H. parasuis* with well-characterized clinical origins. An analysis of the pangenome of the original isolate collection showed that the majority

TABLE 2 mPCR results for the original and additional isolate collections using the model cutoff of 0.72[a]

PCR results	Metadata results (no. [type, %])		
	Disease associated	Carriage	Total
Original isolate collection			
Disease associated	85 (TP, 75)	2 (FP, 7)	87
Carriage	30 (FN, 25)	26 (TN, 93)	56
Additional isolate collection			
Disease associated	76 (TP, 90)	0	76
Carriage	8 (FN, 10)	0	8
Total	199	28	227

[a]The table shows the categorization of the isolates according to available disease-associated metadata and mPCR. The corresponding rates for true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) are given.

Downloaded from http://jcm.asm.org/ on December 14, 2017 by LONDON SCHOOL OF HYGIENE & TROPICAL MED

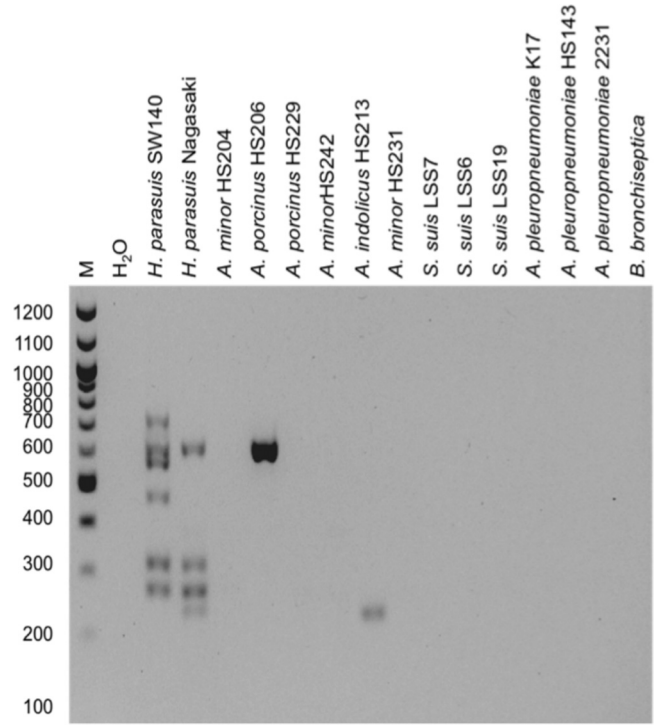

FIG 3 Negative-control panel showing specificity of the primer sets across a range of commensal and pathogenic bacteria of the pig respiratory tract (*Haemophilus parasuis* positive controls serovar 2 (SW140) and serovar 5 (Nagasaki), *Actinobacillus minor*, *A. porcinus*, *A. indolicus*, *Streptococcus suis*, *A. pleuropneumoniae*, and *Bordetella bronchiseptica*). M, Quick-load 100 bp DNA ladder.

of previously published putative "virulence factors" for *H. parasuis* did not show a strong relationship with virulence, as they were core genes (i.e., present in all isolates). Furthermore, no single marker of virulence could be identified, emphasizing the necessity for an mPCR approach for any future pathotyping tool. These findings meant that the design of a diagnostic test to predict the virulence of isolates would not be straightforward. Therefore, the decision was made to develop a pathotyping model for *H. parasuis* based on a combination of genes found in higher proportions of disease-associated isolates and in predominantly disease-associated BAPS populations together with a set of virulence-associated genes from the GWAS analysis of the original isolate collection (23). The resulting mPCR generated a pattern of 11 amplicons, including a species-specific marker (25) which, when entered into the model, correctly predicted the allocation of 78% of the original isolate collection to their observed virulence category from our silver-standard clinical metadata. The majority of the genes targeted by this mPCR have not previously been linked to virulence in *H. parasuis* or other bacterial species, an exception being inosine-5′-monophosphate (IMP) dehydrogenase, which has been implicated in the adhesion of *Streptococcus suis* serotype 2 (26). Testing of the additional collection of disease-associated isolates resulted in a 90% accuracy. Combining the results from both collections gave an overall accuracy of 83%.

There was no obvious pattern based on this mPCR for the prediction of virulence, as each gene has its own importance in the model and may have a positive or negative effect. In addition, the amplicon-based outcome was not cumulative, i.e., the number of bands did not relate to the virulence potential, and so it is necessary to input the amplicon pattern into the model to interpret the results of the mPCR and predict the virulence potential. Therefore, we have developed a user-friendly online tool (https://hps-pathotyping.shinyapps.io/Patho-app) to implement the model, built using R and the Shiny package (v0.10.2.1; RStudio, Inc., Boston, MA), which requires only the input of a table of the mPCR amplicon pattern. It then provides a fitted value for the amplicon

Downloaded from http://jcm.asm.org/ on December 14, 2017 by LONDON SCHOOL OF HYGIENE & TROPICAL MED

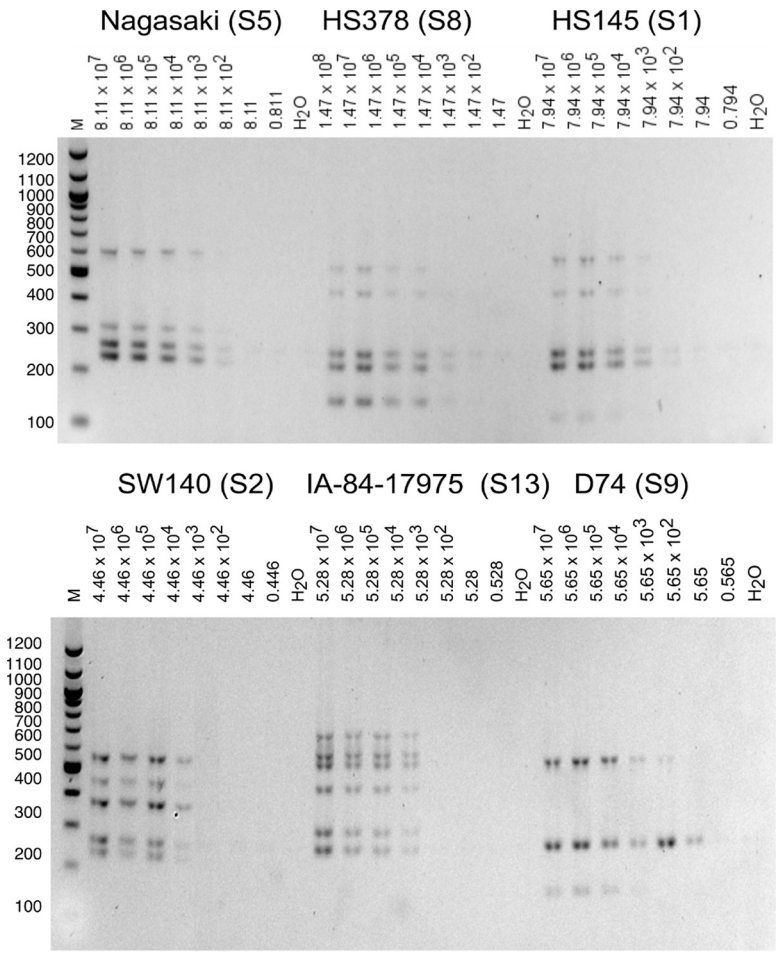

FIG 4 Limit of detection for the pathotyping multiplex PCR using pure genomic DNA (genomes/μl) for seven isolates representing all of the bands present in the pathotyping multiplex PCR. M, Quick-load 100 bp DNA ladder.

pattern using the model and an associated interpretation of this value in simple virulence categories: virulent, potentially virulent, and carriage. The additional category of potentially virulent (based on a buffer zone of $\pm$0.1 on the fitted model value) allows the user to interpret the risk of a particular isolate on a case-by-case basis. The threshold for this buffer zone was chosen to minimize the false-positive and false-negative rates for our model; but, in clinical practice, there are many additional factors that contribute to the outcome of disease beyond bacterial genotype, such as the interplay between host immunity and coinfections with additional pathogens present at the time of challenge. For any approach that attempts only to determine the genotype of the bacterial pathogen, it is unrealistic to be able to predict the outcome of disease with 100% accuracy.

We identified a number of discrepancies between the expected and actual amplicon patterns. These could be explained through errors in sequence assemblies, as we used predominantly draft genome sequences. In addition, the assessment of the species specificity of the mPCR revealed that two of the genes could be identified in single representatives of commensal *Actinobacillus* species. This result was not predicted by the BLAST comparison of the available draft genomes and may have occurred via horizontal gene transfer in these closely related species.

The mPCR was developed from, and designed for use on, DNA extracted from individual bacterial isolates. The presence of multiple isolates of *H. parasuis*, particularly in URT samples (27–29), is one of the major challenges in designing a prospective

Downloaded from http://jcm.asm.org/ on December 14, 2017 by LONDON SCHOOL OF HYGIENE & TROPICAL MED

molecular tool for determining virulence. Current surveillance methods for other pathogens such as *Actinobacillus pleuropneumoniae* operate on an individual isolate level, where the detection of the bacterium initiates a more thorough investigation at the single-colony level to look for the presence of toxins (30, 31). We see this pathotyping mPCR as a useful tool for surveillance as a preventative measure. For example, if two herds were about to be mixed, then prospective surveillance sampling by nasal swabs from these populations might enable the isolation of *H. parasuis* colonies (32), which, in turn, could be tested by mPCR to determine whether different isolates are being carried and if any are likely to cause disease when introduced to a naive population. This could also be used in conjunction with a recently published mPCR for molecular serotyping (25). Therefore, this mPCR approach to the pathotyping of *H. parasuis* presents a major step forward for preventive health programs through its rapidity as well as its specificity.

Ideally, any new research tool or diagnostic test should be compared against an accepted "gold standard" assay. This enables meaningful comparisons of sensitivity, specificity, and overall accuracy for the new test. For *H. parasuis*, this is challenging for a number of reasons. First, the ideal standard would entail a panel of isolates for which a series of consistently controlled *in vivo* challenge experiments had been performed using pigs of identical immune status and genetics (1, 9, 13, 18). Both bacterial and host factors impact significantly the disease outcome for a given isolate, and although *in vivo* challenge data are available for a number of isolates of *H. parasuis*, most of these data come from separately published studies using small sets of isolates and under differing conditions (1, 9, 13, 18). Our approach, which has its caveats, was to take a large collection of field-derived isolates from pigs with or without clinical evidence of *H. parasuis*-related disease and to use this for comparative purposes as a silver standard. These data were obtained from detailed postmortem results, including clinical signs, the tissue of isolation, and the presence of any additional pathogens, and so a certain number of false negatives and false positives were expected based on the unique balance between host immunity and bacterial virulence potential in each individual clinical context. However, we were able to perform an *in silico* comparison of our isolates (23) to the existing *vtaA* mPCR primers (16) (data not shown) as the leading method for the prediction of virulence to date. This was performed using BLAST, which resulted in 82% sensitivity but only 56% specificity compared with our clinical metadata. Our mPCR had an accuracy (overall accuracy 83%) similar to that of the *vtaA* mPCR (accuracy 80%) when tested on our isolate collection, but, although the sensitivity was slightly lower, it had a far higher specificity and exceeds the usefulness of previous methods, such as fingerprinting methods that have limited repeatability between laboratories or usefulness in field situations (33–36).

While our mPCR performed very well, a number of false negatives were identified when testing on the original isolate collection. Of these, 59% were linked to coinfections or respiratory disease, and since pneumonia in pigs is typically a multifactorial disease (37), it is likely that these isolates were coinfections. Similarly, for the additional isolate collection, 4 of 9 (43%) false negatives were isolated from the lungs or from animals that had respiratory disease. A far lower proportion of isolates was recorded as false positive from the original isolate collection, with only 2 recorded from a total of 143 tested. These isolates were predicted, using our model, as being capable of causing disease, but were isolates from the URT of a pig with no clinical evidence of disease. This phenomenon could be explained by a preexisting passive or active immunity resulting in subclinical carriage, which has been described in the field and in controlled challenge studies using disease-associated isolates (1, 38, 39).

This method was designed using a U.K.-biased isolate collection (54% of the original isolate collection) from diverse serovars and will require testing against isolates from a wider range of locations. However, the method offers flexibility and can be easily updated as additional training sets are studied, whether by changing the model cutoff or adding new virulence markers or population-based markers that would give further indications of mixed isolates in a sample. Overall, the proposed mPCR is an accurate,

Downloaded from http://jcm.asm.org/ on December 14, 2017 by LONDON SCHOOL OF HYGIENE & TROPICAL MED

sensitive, and specific method for predicting the virulence of *H. parasuis*, which could be applied to a variety of samples and, with further optimization, could become part of prospective surveillance procedures.

MATERIALS AND METHODS

Isolate collection and culture. The pangenome of *H. parasuis* was previously defined (23) based on a collection of 212 diverse field isolates and reference strains. For the ease of description, we refer to the previously sequenced *H. parasuis* population as the original isolate collection, containing both the field isolates and the reference strains. In this study, we used a subset of this sequenced isolate collection with the most detailed clinical metadata (including the tissue of origin, clinical signs from the postmortem investigation, and cause of death) for the development and testing of this mPCR ($n = 143$). This was used as the training set for the mPCR. A summary of the training set based on geographic location, serovar, and virulence is included in Table 1.

An additional 84 disease-associated isolates of *H. parasuis* were subsequently collected by the APHA, and this was termed the additional isolate collection. These isolates were used as an outside test set with known virulence but no expected amplicon pattern. Culture and gDNA extraction were performed as described previously (23) for evaluation using the mPCR.

Isolates of closely related commensal *Pasteurellaceae*, including *Actinobacillus indolicus* ($n = 1$ [European Nucleotide Archive no. ERS132160]), *Actinobacillus minor* ($n = 3$ [ERS132116, ERS132158, and ERS132165]), and *Actinobacillus porcinus* ($n = 2$ [ERS132148 and ERS132163]) were identified from routine APHA diagnostic investigations and were also genome sequenced. These genome sequences were used to test the specificity of primers designed against *H. parasuis* (using BLASTn) and as part of the negative-control panel. In addition, field isolates of the pathogens *Actinobacillus pleuropneumoniae* ($n = 3$), *Bordetella bronchiseptica* ($n = 1$), and *Streptococcus suis* ($n = 3$) were used in the negative-control panel for the mPCR.

Pathotyping multiplex design. (i) Gene choice. A list of candidate "virulence-associated" genes was constructed from three sources. The first was 48 virulence-associated genes identified from a GWAS of the accessory genome (23, 24), including genes that were either positively or negatively associated with virulence. The second source was 10 previously suggested virulence factors that were identified in more than 60% of disease-associated isolates and less than 40% of carriage isolates from the pangenome analysis (23). BAPS (23, 40) of *H. parasuis* had identified five populations, two of which were predominantly of disease-associated isolates (BAPS 4 to 95% and BAPS 5 to 98%). Genes that were found in a large proportion of isolates from these BAPS populations (>75%) were used as the third list ($n = 251$). Only three genes were identified in both the BAPS and GWAS lists.

This list of 306 genes was refined for suitability as markers for a pathotyping mPCR. A gene was ruled out as a candidate if it was less than 100 bp in length, if the predicted function of the gene was a transposase, phage gene, or integrase, if there were no conserved regions in the gene to which primers could be designed, or if the gene was commonly found to have a pseudogene in the previously published pangenome (cutoff, <20% of representatives of a given gene were pseudogenes). Finally, a species-specific marker for *H. parasuis* was also included as previously described (25).

(ii) Model optimization and statistical analysis. A series of stepwise logistic generalized linear models (in both directions using the MASS R library [25, 41]) was used to build a model using disease association as the phenotype and small subsets of the candidate genes (<20) as independent variables based on the presence and absence of each gene from the pangenome. The candidates that were significant in the models (P value < 0.1) were taken forward to the second round. As combinations of genes could contribute to virulence, situations in which up to five genes at a time were interacting were investigated in the series of models. For example, the addition of genes A to E with interactions to the model provided an estimation of the importance of each gene individually, the importance of all combinations of two genes, three genes, and four genes, and finally, the importance of the combination of all five genes together and whether this improved the model. Finally, these two lists were combined and the model was assessed using stepwise selection again to remove spurious interactions.

The output of the model was a fitted value between 0 and 1. We assessed the performance of the model using the ROCR R package (42) to predict virulence by varying the cutoff on the scale of the fitted values; below the cutoff was considered "carriage" and above was considered "virulent." The accuracy, sensitivity, and specificity at each threshold were then calculated by comparing the predictions to the known clinical category for the original isolate collection. Accuracy was defined as the sum of true positives (TP) and true negatives (TN) divided by all isolates tested. Sensitivity was defined as the TP rate: TP divided by TP and false negatives (FN). Specificity was defined as the TN rate: TN divided by TN and false positives (FP). An ROC curve was built using the ROCR R library (42), which is a plot of the sensitivity and specificity for each cutoff, and was used to identify a cutoff for the model that achieved greater than 80% sensitivity and 80% specificity. We also proposed that a buffer zone of ± 0.1 be added to the model cutoff, which we considered to be a category for "potentially virulent" isolates for the interpretation of the user.

(iii) Primer design. Primer3 was used to design primers between 21 and 30 bases in length with 40 to 60% G+C content based on the New England BioLabs (NEB) master mix guidelines for mPCR design. The primers for each gene were compared to the *H. parasuis* genomes using BLASTn (word size 7) to identify those that matched the presence/absence pattern of the genes in the original isolate collection. Alignments of genes and primers were performed using MEGA (43) and were visually inspected to ensure the best choice of primers given the desired product sizes. Errors of up to 10% in the matches of the

Downloaded from http://jcm.asm.org/ on December 14, 2017 by LONDON SCHOOL OF HYGIENE & TROPICAL MED

Downloaded from http://jcm.asm.org/ on December 14, 2017 by LONDON SCHOOL OF HYGIENE & TROPICAL MED

TABLE 3 Pathotyping genes and multiplex primer information[a]

Gene no.	Existing reference	Predicted function	Prevalence (%)		*P* value	Marker	Primer sequence (5′–3′)		Primer Ratio	Product Size (bp)
			DA	C			Forward	Reverse		
HPS_21058	WP_021114360.1	Inosine-5′-monophosphate dehydrogenase	41	16	0.07	Virulent	CCGAAAGCATAGATCCAAATGC	CCACCTTGTTTACTTGCTTCTGC	1	590
HPS_21059	K756_02725 (AGO15787.1)	Unknown function	45	11	0.06	Virulent	CGTAGCATACGCACACCTAAAG	GAAAGGGCAATAGATACATTTCGG	2	720
HPS_21068	gb|EQA03873.1|	Aspartate kinase monofunctional class	17	53	0.006	Carriage	TGATAATGCACAGATAGTGGGTAGCTCT	TATGACTACTCGCAATGAATTGCTCTG	2	520
HPS_22970	HPS9_02935 (gb|KDB47008.1|)	Unknown function	67	37	0.79	Virulent	CAAGGAAGTGTTTATTTGGGGAAGAAGG	GCTCGATCCAACCCTGAATTTATC	1.5	240
HPS_23060	gb|EQA03144.1|	Helix-turn-helix family protein	43	81	0.0009	Carriage	CCTGTGATTGAATGGGTTCCTCG	GTTGTTTAGATCATCATCCATTAC	2	300
HPS_23300	HS327_01128 (KEZ22556.1)	Unknown function	26	60	0.06	Carriage	GGATATACACTCAGTCTTAGCCCTA	GCAGGTTCTCTTGATTTAGCTTTTC	2	120
HPS_23505	HS327_02008 (KEZ17271.1)	Unknown function	30	76	0.004	Carriage	GTAACAGCTAACCGAGGAAATTGT	TAAGAGTATATGTCCAATCTCCCCG	1.5	170
HPS_23879		Unknown function	13	9	0.04	BAPS 4	GGATAACCCTGATGAACTTGATGAAGAGC	CGAATGTTCGGATTTGTTCAAGTTGG	1.5	210
HPS_23887	HPSSW114_1459 (EQA00635.1)	Putative glycosyltransferase	29	48	0.006	Carriage	GATTTAGCTGAAGCTCAAGACA	GCAAAGCGACATAACCTGCTGATCTTCTC	2	470
HPS_22976	SVR5_00410 (CDH98961.1)	Unknown function	55	41	0.34	BAPS 4, BAPS 5	ATTGAGCAGATTGAGCAACGGAAAA	TTACCAAAGCCCTAATGTTGGGTTT	1	375
HPS_sp-sp		Unknown function	100	100	NA	Species marker	ACAACCTGCAAGTACTTATCGGGAT	TAGCCTCCTGTCTGATATTCCCACG	1	275

[a]The genes are numbered according to the pangenome reference (gene no.) and their respective prevalences among disease-associated (DA) and carriage (C) isolates from the original collection and category of the marker are also shown. NA, not available.

primers to the *H. parasuis* genomes were allowed from the BLASTn results, allowing for the draft nature of the genome sequences. The primers were then compared (using BLASTn with a word size of 7) to the NCBI nonredundant (nr) nucleotide database and the closely related *Pasteurellaceae* bacterial genomes to check for nonspecific primer matches. In addition to the primers for the virulence-associated genes, a previously published species-specific primer pair was added to the mPCR (25). Primers were ordered in dehydrated desalted form from Sigma-Aldrich (Haverhill, Cambridge).

One-step mPCR. The PCR master mix was prepared using One*Taq* Quick-load 2× master mix with standard buffer (New England BioLabs) in accordance with product specifications and protocols. After optimization, the final protocol for amplification of the targets was initiation at 94°C for 30 s, followed by 30 cycles of denaturation at 94°C for 30 s, annealing at 54°C for 30 s, and extension at 68°C for 60 s, followed by a final extension at 68°C for 5 min. Each PCR mixture contained 12.5 μl One*Taq* Quick-load 2× master mix, 0.75 μl of dimethyl sulfoxide ([DMSO] Sigma-Aldrich), 2.2 μl of the primer mix, 2 μl of gDNA for each isolate (at >10 ng/μl),and 7.55 μl UltraPure H_2O (Life Technologies) to a final volume of 25 μl. Ratios of the primers in the primer mix can be found in Table 3. Positive-control isolates were chosen based on the presence of the target gene in their genome sequence. Negative-control isolates were chosen based on the absence of the target gene in their genome sequence. A negative control of UltraPure H_2O was also used for each PCR performed in this study. Gel electrophoresis was performed using 2.0% agarose gel with 5% SybrSafe dye (Invitrogen) and run at 110 V for 90 min using the Quick-load 100 bp DNA ladder (New England BioLabs) for all mPCRs. All mPCR results were visualized using the GelDoc imager (Bio-Rad). The accuracy of the results for these individual PCRs was compared with the expected results from the genome sequences and BLAST matches for the primers.

Validation of the pathotyping mPCR. The mPCR was evaluated using gDNA of 143 isolates from the original collection and gDNA of 84 isolates of the additional collection (gDNA concentration > 10 ng/μl). The expected amplicon pattern for each isolate in the original collection was based on the presence/absence of the 10 genes from their genome sequence (23). After the mPCR was performed, the amplicon pattern was entered into the final model and the fitted model value was calculated. The accuracy, sensitivity, and specificity of the mPCR results were calculated using the cutoff of 0.72 and the known clinical category. For the additional isolate collection, the mPCR results were entered into the model, and the output was compared with clinical metadata. All mPCRs were repeated on three occasions, each time using a separate master mix, to demonstrate the repeatability and accuracy of the mPCR.

To determine whether the mPCR could detect *H. parasuis* directly from a single colony, thus not requiring prior gDNA extraction, colony PCR was performed on a subset of the additional isolate collection ($n = 20$). For the colony PCR, a loopful of bacteria was resuspended in 50 μl of UltraPure H_2O and heated at 100°C for 30 min, and then centrifuged at 4,000 × *g* for 1 min before the supernatant was used in the mPCR. The supernatant (2 μl) was used in the mPCR reaction.

Limit of detection of the mPCR. The concentration of gDNA was measured using a Qubit fluorometer (Life Technologies) with broad-range standards for five isolates (strain name-serovar: HS145-S1, SW140-S2, Nagasaki-S5, C5-S8, D74-Aus–S9, and IA84/17975-S13). The gDNA was diluted six times in a serial dilution using UltraPure H_2O; this was then used as the template in the mPCR to estimate its limit of detection. The number of copies of the genome per microliter was calculated using an average genome size of 2.26 Mb.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/JCM.02464-16.

SUPPLEMENTAL FILE 1, XLSX file, 0.1 MB.

ACKNOWLEDGMENTS

We thank several people for their highly valued help with this study: Patrick Blackall of the University of Queensland, Australia, for the provision of copies of the Australian reference strain set; Feifei Shen for assistance in purification and genomic DNA preparation of isolates used for validation of the mPCR assay; and Brian Hunt, Jon Rogers, and Sarah Howie of the Animal Health and Veterinary Laboratories Agency (now APHA) for collection and biochemical profiling of United Kingdom isolates.

This work was supported by a BPEX Ph.D. studentship and a Longer and Larger (LoLa) grant from the Biotechnology and Biological Sciences Research Council (grants BB/G020744/1, BB/G019177/1, BB/G019274/1, and BB/G003203/1), the United Kingdom Department for Environment, Food and Rural Affairs, and Zoetis, awarded to the Bacterial Respiratory Diseases of Pigs-1 Technology (BRaDP1T) Consortium. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. L.A.W. is supported by both a Dorothy Hodgkin Fellowship funded by the Royal Society (grant DH140195) and a Sir Henry Dale Fellowship funded by the Wellcome Trust (grant 109385/Z/15/Z).

The list of BRaDP1T Consortium members is as follows: from the University of

Downloaded from http://jcm.asm.org/ on December 14, 2017 by LONDON SCHOOL OF HYGIENE & TROPICAL MED

Cambridge, Duncan J. Maskell, Alexander W. (Dan) Tucker, Sarah E. Peters, Lucy A. Weinert, Jinhong (Tracy) Wang, Shi-Lu Luan, and Roy R. Chaudhuri (now at University of Sheffield, Sheffield, United Kingdom); from the Royal Veterinary College, Andrew N. Rycroft, Gareth A. Maglennon, and Dominic Matthews; from the Imperial College London, Paul R. Langford, Janine T. Bossé, and Yanwen Li; and from the London School of Hygiene and Tropical Medicine, Brendan W. Wren, Jon Cuccui, and Vanessa S. Terra.

REFERENCES

1. Aragon V, Cerdà-Cuéllar M, Fraile L, Mombarg M, Nofrarías M, Olvera A, Sibila M, Solanes D, Segalés J. 2010. Correlation between clinicopathological outcome and typing of *Haemophilus parasuis* field strains. Vet Microbiol 142:387–393. https://doi.org/10.1016/j.vetmic.2009.10.025.
2. Rapp-Gabrielson VJ, Oliveira SR, Pijoan C. 2006. *Haemophilus parasuis*, p 681–690. *In* Straw BE, Zimmerman JJ, D'Allaire S, Taylor DJ (ed), Diseases of Swine, 9th ed. Blackwell Publishing, Oxford, UK.
3. Smart NLN, Miniats OP, Rosendal S, Friendship RRM. 1989. Glasser's disease and prevalence of subclinical infection with *Haemophilus parasuis* in swine in southern Ontario. Can Vet J 30:339–343.
4. Oliveira S, Pijoan C. 2004. *Haemophilus parasuis*: new trends on diagnosis, epidemiology and control. Vet Microbiol 99:1–12. https://doi.org/10.1016/j.vetmic.2003.12.001.
5. White M. 2010. NADIS pig health—March 2010 Glässers Dis BPEX Knowledge Transfer, p 1–3.
6. Animal and Plant Health Agency. 2015. Pig diseases quarterly report. GB emerging threats quarterly report. 19:Q4. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/507457/pub-survrep-p0415.pdf.
7. Holtkamp D, Rotto H, Garcia R. 2007. Economic cost of major health challenges in large US swine production systems—part 1. Swine News 30:85–89. http://www.thepigsite.com/articles/1935/economic-cost-of-major-health-challenges-in-large-us-swine-production-systemspart-1/.
8. Brockmeier S, Halbur P, Thacker E. 2002. Porcine respiratory disease complex, p 231–258. *In* Brogden K, Guthmiller J (ed), Polymicrobial diseases. ASM Press, Washington, DC.
9. Kielstein P, Rapp-Gabrielson VJ. 1992. Designation of 15 serovars of *Haemophilus parasuis* on the basis of immunodiffusion using heat-stable antigen extracts. J Clin Microbiol 30:862–865.
10. Turni C, Blackall PJ. 2005. Comparison of the indirect haemagglutination and gel diffusion test for serotyping *Haemophilus parasuis*. Vet Microbiol 106:145–151. https://doi.org/10.1016/j.vetmic.2004.12.019.
11. Del Río ML, Gutiérrez CB, Rodríguez Ferri EF. 2003. Value of indirect hemagglutination and coagglutination tests for serotyping *Haemophilus parasuis*. J Clin Microbiol 41:880–882. https://doi.org/10.1128/JCM.41.2.880-882.2003.
12. Sack M, Baltes N. 2009. Identification of novel potential virulence-associated factors in *Haemophilus parasuis*. Vet Microbiol 136:382–386. https://doi.org/10.1016/j.vetmic.2008.11.008.
13. Brockmeier SL, Register KB, Kuehn JS, Nicholson TL, Loving CL, Bayles DO, Shore SM, Phillips GJ. 2014. Virulence and draft genome sequence overview of multiple strains of the swine pathogen *Haemophilus parasuis*. PLoS One 9:e103787. https://doi.org/10.1371/journal.pone.0103787.
14. Costa-Hurtado M, Aragon V. 2013. Advances in the quest for virulence factors of *Haemophilus parasuis*. Vet J 198:571–576. https://doi.org/10.1016/j.tvjl.2013.08.027.
15. Pina S, Olvera A, Barceló A, Bensaid A. 2009. Trimeric autotransporters of *Haemophilus parasuis*: generation of an extensive passenger domain repertoire specific for pathogenic strains. J Bacteriol 191:576–587. https://doi.org/10.1128/JB.00703-08.
16. Olvera A, Pina S, Macedo N, Oliveira S, Aragon V, Bensaid A. 2012. Identification of potentially virulent strains of *Haemophilus parasuis* using a multiplex PCR for virulence-associated autotransporters (*vtaA*). Vet J 191:213–218. https://doi.org/10.1016/j.tvjl.2010.12.014.
17. Galofré-Milà N, Correa-Fiz F, Lacouture S, Gottschalk M, Strutzberg-Minder K, Bensaid A, Pina-Pedrero S, Aragon V. 2017. A robust PCR for the differentiation of potential virulent strains of *Haemophilus parasuis*. BMC Vet Res 13:124. https://doi.org/10.1186/s12917-017-1041-4.
18. Wilkinson JM, Sargent CA, Galina-Pantoja L, Tucker AW. 2010. Gene expression profiling in the lungs of pigs with different susceptibilities to Glässer's disease. BMC Genomics 11:455. https://doi.org/10.1186/1471-2164-11-455.
19. Vahle JL, Haynes JS, Andrews JJ. 1995. Experimental reproduction of *Haemophilus parasuis* infection in swine: clinical, bacteriologic, and morphologic findings. J Vet Diagn Invest 7:476–480. https://doi.org/10.1177/104063879500700409.
20. Bak H, Riising HJ. 2002. Protection of vaccinated pigs against experimental infections with homologous and heterologous *Haemophilus parasuis*. Vet Rec 151:502–505. https://doi.org/10.1136/vr.151.17.502.
21. Bello-Orti B, Costa-Hurtado M, Martinez-Moliner V, Segalés J, Aragon V. 2014. Time course *Haemophilus parasuis* infection reveals pathological differences between virulent and non-virulent strains in the respiratory tract. Vet Microbiol 170:430–437. https://doi.org/10.1016/j.vetmic.2014.01.011.
22. Brockmeier SL, Loving CL, Mullins MA, Register KB, Nicholson TL, Wiseman BS, Baker RB, Kehrli ME. 2013. Virulence, transmission, and heterologous protection of four isolates of *Haemophilus parasuis*. Clin Vaccine Immunol 20:1466–1472. https://doi.org/10.1128/CVI.00168-13.
23. Howell KJ, Weinert LA, Chaudhuri RR, Luan S, Peters SE, Corander J, Harris D, Angen Ø Aragon V, Bensaid A, Williamson SM, Parkhill J, Langford PR, Rycroft AN, Wren BW, Holden MTG, Tucker AW, Maskell DJ. 2014. The use of genome wide association methods to investigate pathogenicity, population structure and serovar in *Haemophilus parasuis*. BMC Genomics 15:1179. https://doi.org/10.1186/1471-2164-15-1179.
24. Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11:94. https://doi.org/10.1186/1471-2156-11-94.
25. Howell KJ, Hernandez-Garcia J, Weinert LA, Luan S-L, Wang J, Peters SE, Angen Ø Aragon V, Williamson SM, Parkhill J, Langford PR, Rycroft AN, Wren BW, Maskell DJ, Tucker AW. 2015. A molecular serotyping multiplex PCR for *Haemophilus parasuis*. J Clin Microbiol 53:3812–3821. https://doi.org/10.1128/JCM.01991-15.
26. Zhou J, Zhang X, He K, Wang W, Ni Y, Zhu H, Yu Z, Mao A, Lv L. 2014. Characterization and proteome analysis of inosine 5-monophosphate dehydrogenase in epidemic *Streptococcus suis* serotype 2. Curr Microbiol 68:663–669. https://doi.org/10.1007/s00284-014-0527-6.
27. Boerlin P, Poljak Z, Gallant J, Chalmers G, Nicholson V, Soltes GA, Macinnes JI. 2013. Genetic diversity of *Haemophilus parasuis* from sick and healthy pigs. Vet Microbiol 167:459–467. https://doi.org/10.1016/j.vetmic.2013.07.028.
28. Cai X, Chen H, Blackall PJ, Yin Z, Wang L, Liu Z, Jin M. 2005. Serological characterization of *Haemophilus parasuis* isolates from China. Vet Microbiol 111:231–236. https://doi.org/10.1016/j.vetmic.2005.07.007.
29. Olvera A, Cerdà-Cuéllar M, Nofrarías M, Revilla E, Segalés J, Aragon V. 2007. Dynamics of *Haemophilus parasuis* genotypes in a farm recovered from an outbreak of Glässer's disease. Vet Microbiol 123:230–237. https://doi.org/10.1016/j.vetmic.2007.03.004.
30. da Costa MM, Klein CS, Balestrin R, Schrank A, Piffer IA, da Silva SC, Schrank IS. 2004. Evaluation of PCR based on gene *apxIVA* associated with 16S rDNA sequencing for the identification of *Actinobacillus pleuropneumoniae* and related species. Curr Microbiol 48:189–195. https://doi.org/10.1007/s00284-003-4162-x.
31. Giménez-Lirola LG, Jiang YH, Sun D, Hoang H, Yoon KJ, Halbur PG, Opriessnig T. 2014. Simultaneous detection of antibodies against Apx toxins ApxI, ApxII, ApxIII, and ApxIV in pigs with known and unknown *Actinobacillus pleuropneumoniae* exposure using a multiplexing liquid array platform. Clin Vaccine Immunol 21:85–95. https://doi.org/10.1128/CVI.00451-13.
32. Turni C, Blackall PJ. 2010. Serovar profiling of *Haemophilus parasuis* on

Downloaded from http://jcm.asm.org/ on December 14, 2017 by LONDON SCHOOL OF HYGIENE & TROPICAL MED

Australian farms by sampling live pigs. Aust Vet J 88:255–259. https://doi.org/10.1111/j.1751-0813.2010.00592.x.
33. de la Puente Redondo VA, Navas Mendez J, Garcia del Blanco N, Ladron Boronat N, Gutierrez Martin CB, Rodriguez Ferri EF. 2003. Typing of *Haemophilus parasuis* strains by PCR-RFLP analysis of the tbpA gene. Vet Microbiol 92:253–262. https://doi.org/10.1016/S0378-1135(02)00362-0.
34. del Rio ML, Martin CB, Navas J, Gutierrez-Muniz B, Rodriguez-Barbosa JI, Rodriguez Ferri EF. 2006. *aroA* gene PCR-RFLP diversity patterns in *Haemophilus parasuis* and *Actinobacillus* species. Res Vet Sci 80:55–61. https://doi.org/10.1016/j.rvsc.2005.03.004.
35. Rafiee M, Bara M, Stephens CP, Blackall PJ. 2000. Application of ERIC PCR for the comparison of isolates of *Haemophilus parasuis*. Aust Vet J 78:846–849. https://doi.org/10.1111/j.1751-0813.2000.tb10507.x.
36. Castilla KS, de Gobbi DDS, Moreno LZ, Paixão R, Coutinho TA, dos Santos JL, Moreno AM. 2012. Characterization of *Haemophilus parasuis* isolated from Brazilian swine through serotyping, AFLP and PFGE. Res Vet Sci 92:366–371. https://doi.org/10.1016/j.rvsc.2011.04.006.
37. Opriessnig T, Giménez-Lirola LG, Halbur PG. 2011. Polymicrobial respiratory disease in pigs. Anim Heal Res Rev 12:133–148. https://doi.org/10.1017/S1466252311000120.
38. Howell KJ, Weinert LA, Luan S-L, Peters SE, Chaudhuri RR, Harris D, Angen O, Aragon V, Parkhill J, Langford PR, Rycroft AN, Wren BW, Tucker AW, Maskell DJ. 2013. Gene content and diversity of the loci encoding biosynthesis of capsular polysaccharides of the fifteen serovar reference strains of *Haemophilus parasuis*. J Bacteriol 195:4264–4273. https://doi.org/10.1128/JB.00471-13.
39. Angen O, Oliveira S, Ahrens P, Svensmark B, Leser TD. 2007. Development of an improved species specific PCR test for detection of *Haemophilus parasuis*. Vet Microbiol 119:266–276. https://doi.org/10.1016/j.vetmic.2006.10.008.
40. Marttinen P, Baldwin A, Hanage WP, Dowson C, Mahenthiralingam E, Corander J. 2008. Bayesian modeling of recombination events in bacterial populations. BMC Bioinformatics 9:421. https://doi.org/10.1186/1471-2105-9-421.
41. Venables WN, Ripley BD (ed). 2002. Modern applied statistics with S (statistics and computing), 4th ed. Springer, New York, NY.
42. Sing T, Sander O, Beerenwinkel N, Lengauer T. 2005. ROCR: visualizing classifier performance in R. Bioinformatics 21:3940–3941. https://doi.org/10.1093/bioinformatics/bti623.
43. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729. https://doi.org/10.1093/molbev/mst197.

Downloaded from http://jcm.asm.org/ on December 14, 2017 by LONDON SCHOOL OF HYGIENE & TROPICAL MED