The effect of environmental and management associated factors on prevalence and diversity of *Streptococcus suis* in clinically healthy pig herds of China and UK

Geng Zou[1], Jianwei Zhou[1], Ran Xiao[1], Liangsheng Zhang[1], Yuting Cheng[1], Hui Jin[1,2], Lu Li[1], Lijun Zhang[1], Bin Wu[1], Ping Qian[1], Shaowen Li[1], Lixin Ren[5], Jinhong Wang[2], Olusegun Oshota[2], Juan Hernandez-Garcia[2], Thomas M. Wileman[2], Stephen Bentley[6], Lucy Weinert[2], Duncan J. Maskell[2], A.W. (Dan) Tucker[2, *, $], Rui Zhou[1,3,4, *, $]

[1] State Key Laboratory of Agricultural Microbiology and Key Laboratory of Veterinary Diagnosis (Ministry of Agriculture), College of Veterinary Medicine, Huazhong Agricultural University, Wuhan 430070, China

[2] Department of Veterinary Medicine, University of Cambridge, Madingley Road, Cambridge, CB3 0ES, UK

[3] International Research Center for Animal Diseases (MOST), Wuhan 430070, China

[4] Cooperative Innovation Center of Sustainable Pig Production, Wuhan 430070, China

[5]Animal Health Supervision Institute in Jiangxia District of Wuhan City, Wuhan 430200, China

[6] The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

* Corresponding authors:

Prof. Rui Zhou (Phone: +86 27 87281878; e-mail: rzhou@mail.hzau.edu.cn); Dr. A.W. (Dan) Tucker (Phone: +44 1223 330855; e-mail: awt1000@cam.ac.uk).

$ Authors made equal contributions.

Abstract

Streptococcus suis (*S. suis*), a global zoonosis of pigs, shows regional differences in prevalence of human associated disease for Asian and non-Asian countries. The isolation rate and diversity of *S. suis* on tonsils of healthy slaughter pigs in China and the UK were studied for effects of geography, temperature, pig age and farm type. Isolates underwent analysis of molecular serotype, multilocus sequence type, and virulence-associated genotyping. Although we found no significant difference in positive isolation rates between Chinese and UK farms, the prevalences of serotypes previously associated with human disease were significantly greater in the Chinese collection ($p = 0.003$). A significant effect of temperature was found on the positive isolation rate of the Chinese samples and prevalence of human disease associated serotypes in the UK *S. suis* population (China, $p = 0.004$; UK, $p = 0.024$), and on the prevalence of isolates carrying key virulence genes in China ($p = 0.044$). Finally, we found marked diversity among *S. suis* isolates with statistically significant temperature effects on detection of multiple strain types within individual pigs. This study highlighted the high carriage prevalence and diversity of *S. suis* among clinically healthy pig herds of China and the UK. The significant effect of temperature on prevalence of isolation, human disease associated serotypes and diversity carried by individual pigs may shed new light on geographic variations in human *S. suis* associated disease.

Importance

Streptococcus suis is a global zoonotic pathogen and also a normal colonizer mainly carried on the tonsil of pigs. Thus, it is important to study the effect of environmental and management associated factors on the *S. suis* populations in clinically healthy

pigs. In this research, we investigated the similarities and differences between the *S. suis* populations obtained from different pig ages, seasons and farm management systems and discovered the relationship between high climatic temperature and the prevalence of *S. suis*.

Introduction

Streptococcus suis, a globally important pig pathogen that infects pigs and humans (1), is a normal colonizer of the upper respiratory tract (URT) of healthy pigs (2) and is a major worldwide driver of antibiotic administration to pigs. Colonization of pigs by *S. suis* is widespread globally but disease-associated strains are found only rarely in the upper respiratory tract (3). The high level of genomic diversity between isolates of *S. suis*, and evidence for high levels of recombination, indicates the potential for emergence of additional disease associated strains (3) and the importance of gaining a deeper understanding of *S. suis* population diversity in the URT of healthy pigs.

Classical typing methods identified 35 serotypes of *S. suis* with a large number of non-serotypable isolates, but serotypes 32 and 34 were excluded after analysis of the 16S rRNA gene sequence showed they should be reclassified as *S. orisratti* (4). It was recently proposed that serotypes 20, 22, 26 and 33 should also be reclassified as another species based on the phylogeny of one or two genes (5) but whole-genome analysis of a larger collection of divergent isolates of *S. suis* found them to be more closely associated with *S. suis* than any other species (6, 7). Of the 33 serotypes, several are over-represented among the isolates from clinical samples. Serotypes 1, 2, 7, 9 and 14 were the most prevalent disease-associated serotypes in Europe, Southeast Asia and South America while serotypes 3 and 8 were more frequently isolated from cases in North America (8). However, human cases of disease are dominated by

serotype 2, followed by serotype 14 with serotypes 4, 5, 16, 21, 24 and 31 occasionally reported (9-13). The introduction of molecular typing by PCR, based on polymorphisms in the *cps* loci (14) has assisted in assigning molecular serotype to most, but not all, previously non-serotypable isolates commonly found in healthy pigs. A new approach, known as NCLs (novel *cps* loci), was introduced to address the classification of remaining non-serotypable *S. suis* based on 8 novel *cps* loci (15).

Multi-locus sequence typing (MLST) (16), based on polymorphisms in 7 housekeeping gene fragments to identify discrete sequence types (ST), not only showed serotype to be a poor indicator of genetic relatedness but also identified significant geographical differences in ST distribution. For example, ST7 was only found in China where it was associated with a severe outbreak of pig and human disease in Sichuan Province in 2005, while ST1 was predominant in meningitis and septicaemia among humans and pigs in Asia, Europe and South America (8, 17). To date 956 STs have been described (PubMLST: https://pubmlst.org/ssuis/), the majority of these STs belonging to disease associated strains. However, the MLST method is based on a small fraction of the genome and, although useful as epidemiological tool, it has been proved ineffective in discrimination of disease and non-disease associated isolates.

Reports of the first cases of human disease dated back to the 1950s in Europe, where disease was historically linked to occupational exposure, but recent surveillance shows that most human cases currently occur in Asia (18). In China, *S. suis* has been implicated in summer-time outbreaks of toxic shock-like syndrome (19, 20) and was identified as the third most common cause of adult bacterial meningitis in Hong Kong (21). Meningitis and septicaemia are widely linked to *S. suis* in other Asian countries including Thailand and especially Vietnam where it was the most frequent reported

cause of adult bacterial meningitis (22). The widespread occurrence in Asia of zoonotic cases of *S. suis* might be explained by the high density of pigs in the region, traditional slaughtering practices without preventive measures, and the consumption of uncooked or lightly cooked pig products (23). While these factors have been shown to be relevant, to date no large-scale genomic comparative studies have been undertaken on the diversity and prevalence of carriage of *S. suis* in pigs entering the pig meat supply chain in Asian versus European countries.

This study set out to apply emergent molecular epidemiological methods to gain insights into the factors affecting *S. suis* carriage in pigs entering the human food chain. A longitudinal study was performed to investigate the genomic characteristics of the *S. suis* populations in clinically healthy pig populations in pig meat supply chains in China and the UK. We set out to describe relationships between strain prevalence and diversity with pig age, climatic temperature and farm management system.

Results

Factors associated with *S. suis* positive isolation rate at farm level

Overall and farm level isolation rates. The Chinese collection included 223 isolates of *S. suis,* confirmed by whole genome sequence (WGS), obtained from 137/500 (27.4%) tonsil swabs of clinical healthy pigs. The positive isolation rate of each farm at the 5 weeks of age sample point ranged from 8.00% to 62.96%, and at 20 weeks the range was from 4.00% to 52.00%. For the UK collection, 127 isolates, confirmed by WGS, were isolated from 89/250 (35.6%) tonsil swabs. The positive isolation rate ranged from 4.00% to 92.00% at the 5 weeks old sampling occasion and from 4.00% to 40.00% at 20 weeks old (**Table 1**). There was no significant difference between the

positive isolation rate of the Chinese samples and the UK samples (two tailed *p*-value = 0.174, t-test).

The effect of temperature, age and farm type on positive isolation rate at farm level.

The 20 Chinese sampling occasions and the 10 UK sampling occasions were respectively divided into two groups across the mean values of the corresponding set of daily mean air temperature values of the sampling times.

For the Chinese sampling occasions, a three-way ANOVA was used to evaluate the impact of temperature, farm type and age, which showed that temperature was the only significant factor affecting positive isolation rate at farm level. The positive isolation rates for pig farms sampled on high temperature occasions were significantly greater than for those sampled on low temperature occasions (temperature, $p = 0.004$; age, $p = 0.203$; farm type, $p = 0.064$) (**Table 3**). For the UK sampling occasions, since the equal variance test for two-way ANOVA was not passed, t-tests and paired t-tests were used to evaluate the effect of temperature and age separately. Although there was no detectable significant effect of temperature or age (temperature, two tailed *p*-value = 0.111, t-test; age, two tailed *p*-value = 0.285, paired t-test), the distribution of the UK samples showed a similar pattern to that of the Chinese samples, where the frequency of positive samples was higher for sampling occasions of 5 weeks old pigs at high air temperatures (**Fig. 1**).

When the Chinese and the UK sampling occasions were combined and two-way ANOVA was used to evaluate the effect of temperature and age, there is no significant interaction between temperature and age, and the temperature showed the significant effect (*p*-value = 0.001) while the age also had noteworthy effect that 5 weeks old pigs may carry more *S. suis* (*p*-value = 0.051).

Typing characteristics of the *S. suis* populations

Molecular serotyping. In total, 71.75% (160/223) of the Chinese *S. suis* isolates could be allocated a molecular serotype across 18 serotypes which included all 8 serotypes that had human infection cases reported (serotype 2, 4, 5, 14, 16, 21, 24 and 31). A total of 68.50% (87/127) of the UK isolates were serotypeable covering 15 serotypes and 5 human disease associated serotypes (serotype 4, 5, 16, 21 and 31). Serotype 16 was the most dominant isolates in both collections (China, n = 55; UK, n = 20), followed by serotypes 15 (n = 17) and 31 (n = 16) in the Chinese collection and serotypes 9 (n = 13) and 15 (n = 11) in UK collection. Isolates that could not be allocated a molecular serotype by this method were analysed according to the eight novel categories identified in the NCL scheme. Both collections showed seven NCL types, while NCL type 6 and NCL type 1 could not be found in the Chinese and the UK collections respectively (**Fig. 2**).

Multilocus sequence typing. For the Chinese collection, a total of 207 of the original 223 *S. suis* isolates could be classified into 93 different kinds of STs. The STs with 6/7 identical alleles were grouped as a Clonal Complex while the STs belonged to no Clonal Complex were named as Singletons. The 93 STs of Chinese isolates were finally grouped as 15 Clonal Complexes and 60 Singletons. Out of the 93 STs only 11 STs were reported previously. The 2 serotype 1&14 isolates typed as ST11, and for the 4 serotype 2&1/2 isolates, 1 was typed as ST1 and the other 3 were ST7.

Of the original 127 UK *S. suis* isolates, 118 were classified into 70 different kinds of STs, clustered as 53 Singletons and 7 Clonal Complexes, only one of which had been previously reported (ST29). There were 30 isolates that could not be typed by the MLST system due to absence of one or more alleles. There was no same ST found in both the Chinese and the UK collection (**Fig. 3**) (**Table S1**).

Genotypes of mrp, epf and sly. Genotype was allocated to 223 Chinese isolates and, in total, 8 VGTs were found: *epf mrp*$^{-}$*sly*$^{-}$ (n = 174); *epf mrp*^{NA1}*sly*$^{-}$ (n = 10); *epf mrp*^{NA1}*sly*$^{+}$ (n = 12); *epf mrp*^{NA2}*sly*$^{+}$ (n = 12); *epf mrp*^{EU}*sly*$^{+}$ (n = 4); *epf*$^{+}$*mrp*^{NA1}*sly*$^{+}$ (n = 1); *epf*$^{+}$*mrp*^{NA2}*sly*$^{+}$ (n = 5); *epf*$^{+}$*mrp*^{EU}*sly*$^{+}$ (n = 4). All the 49 PVGT isolates (representing 21.9% of the collection) possessed the *mrp* gene. Furthermore, 10/223 (4.38%) isolates in the Chinese collection were both PVGT and belonged to serotypes previously reported in human disease - 1&14 (2/2 in collection were PVGT), 2&1/2 (4/4), 4 (3/5), 21 (1/2). One non-typeable isolate possessed all 3 studied virulence genes. There were 6 of the total 20 sampling occasions (5 weeks old = 1, 20 weeks old = 5) that had no PVGT isolate. The genotypes *epf mrp*^{NA2}*sly*$^{+}$ and *epf mrp*^{NA1}*sly*$^{+}$ were the most prevalent PVGTs that could be detected in all subgroups.

The UK collection (n=127) had possessed 5 different VGTs: *epf mrp*$^{-}$*sly*$^{-}$ (n = 109); *epf mrp*^{NA1}*sly*$^{-}$ (n = 1); *epf mrp*^{NA1}*sly*$^{+}$ (n = 7); *epf mrp*^{NA2}*sly*$^{+}$ (n = 9); *epf*$^{+}$*mrp*^{NA1}*sly*$^{+}$ (n = 1). The only *mrp*, *epf* and *sly* positive isolate was non-serotypeable. Same as with the Chinese collection, in total there were 18 PVGT isolates (representing 14.1% of the collection) and, the *mrp* gene was existed detected in all the virulence genes harboring PVGT isolates and the genotypes *epf mrp*^{NA2}*sly*$^{+}$ and *epf mrp*^{NA1}*sly*$^{+}$ were the most prevalent PVGTs that could be detected in all subgroups. (**Table 2**). No isolate in the UK collection was both PVGT and HDAS. There were 3 of the total 10 sampling occasions (5 weeks old = 1, 20 weeks old = 2) that had no PVGT isolate.

Factors associated with pig-level diversity of *S. suis*.

It was not rare for more than one *S. suis* isolate to be obtained from the same tonsil scrape, each with different genomic characteristics. In summary, 28.47% (39/137) carried more than one STs *S. suis* isolates, 27.01% (37/137) of positive Chinese

samples carried more than one serotype of *S. suis*, and 13.87% (19/137) carried isolates with multiple virulence gene profiles. For UK samples, 25.84% (23/89) carried more than one STs *S. suis* isolates, 22.47% (20/89) of positive tonsil scrapes carried more than one serotype of *S. suis* and 10.11% (9/89) carried isolates with multiple virulence gene profiles.

Individual isolates were also identified based on their ST. The proportion of pigs carrying *S. suis* isolates with multiple STs at each sampling occasion was used to evaluate the influence of temperature, farm type and age on the diversity of *S. suis* within sampled pigs. For the Chinese samples, three-way ANOVA results showed that there was a significant interaction between temperature, age and farm type (p = 0.043). Since the further three-way interaction term analysis found no significant interaction between temperature and age at both intensive and small traditional type farm (intensive level, p = 0.062; small level, p = 0.198), simple main effect tests were done at intensive level and small level farm types respectively. The results showed that, at both levels of farm type, temperature was the significant factor while the influence of age was not significant (temperature: intensive level, p = 0.01, small level, p = 0.02; age: intensive level, p = 0.945, small level, p = 0.454). Since the UK dataset failed to pass the equal variance test for two-way ANOVA, t-test and paired t-test were used instead to evaluate temperature and age factors respectively. The UK samples from high temperature sampling occasions had significantly higher pig-level diversity (more STs per tonsil scrape) than those from low temperature sampling occasions (two tailed p-value = 0.0296, t-test), while no significant difference was found between 5 weeks old and 20 weeks old samples (two tailed p-value = 0.401, paired t-test). To compare the pig-level diversity of *S. suis* between the China and UK *S. suis* positive pigs, a filtering process was undertaken to maximize the direct

226 comparability of the Chinese and the UK collections, reducing the former collection
227 from 223 to 209 and the latter from 127 to 125. The result showed that the pig-level
228 diversity of *S. suis* isolates was not significantly different for the Chinese and UK
229 collections (p = 1.000; Rank Sum Test) (**Table 3**).
230
231 **Factors associated with human disease associated serotype prevalence, positive**
232 **virulence genotype prevalence and diversity of *S. suis* at population level.**
233 ***The high temperature subgroup versus the low temperature subgroup.*** In summary,
234 higher temperatures were associated with an increased proportion of human disease
235 associated serotypes (HDAS) which reached statistical significance in the UK
236 subgroups, but not in the Chinese subgroups. HDAS represented 43.45% (73/168) of
237 the Chinese high temperature subgroup (with all 8 HDAS represented) while 32.73%
238 (18/55) of the Chinese low temperature subgroup isolates were HDAS (serotypes 2, 5,
239 16, 21 and 31) (p = 0.213, Chi-square). The UK high temperature subgroup had
240 28.86% (28/97) HDAS isolates (serotype 4, 5, 16, 21 and 31) while the low
241 temperature subgroup had 6.67% (2/30) HDAS isolates (serotype 4 and 16) (p =
242 0.024, Chi-square).
243 Higher temperatures were also associated with increased prevalence of PVGT (i.e.
244 isolates carrying at least 1 virulence associated gene), reaching statistical significance
245 for the Chinese high temperature subgroup, at 25.00% (42/168) (PVGT), compared to
246 that of the Chinese low temperature subgroup (10.91% (6/55) p = 0.044, Chi-square).
247 For the UK collection, the high and the low temperature subgroups included 14/97
248 isolates and 4/30 PVGT isolates respectively, with no significant difference. (p =
249 0.882, *Chi*-square) (**Table 3**).
250 The diversity indexes for high and low temperature subgroups based on MLST

sequence types (STs) were not significantly different, neither in China nor the UK (**Table 2**).

The 5 weeks subgroup versus the 20 weeks subgroup. The prevalence of HDAS was not significantly different for 5 weeks versus 20 weeks subgroups in China or in the UK. Briefly, there was a 42.76% (62/145) prevalence of HDAS in the Chinese 5 weeks old subgroup and 41.02% (32/78) of the 20 weeks old subgroup. The proportions of HDAS in the UK 5 weeks old and 20 weeks old sub group were 28.57% (26/91) and 11.11% (4/36) respectively (China: $p = 0.914$, UK: $p = 0.063$, *Chi*-square).

Similarly, there were no significant differences in prevalence of PVGT isolates between age subgroups in either country. For the Chinese collection, the 5 weeks old subgroup had 24.83% (36/145) isolates carrying virulence gene(s), and that of the 20 weeks old subgroup was 15.38% (12/78). The UK 5 weeks old subgroup had 15.38% (14/91) virulence gene positive isolates, while the 20 weeks old subgroup had 11.11% (4/36). The differences between the 5 weeks old subgroups and the 20 weeks subgroups were not significantly (China: $p = 0.143$, UK: $p = 0.743$, *Chi*-square) (**Table 3**).

The diversity indexes for the 5 and 20 weeks subgroups, based on STs, were not significantly different in China nor the UK (**Table 2**).

The intensive farm subgroup versus the small pig farm subgroup. There were no significant differences in prevalence of HDAS between the Chinese intensive pig farm subgroup 39.86% (55/138) and the small pig farm subgroup 42.35% (36/85) ($p = 0.819$, *Chi*-square). Neither were there any significant differences between these two subgroups in terms of prevalence of PVGT isolates. In the intensive pig farm subgroup, 23.19% (32/138) of isolates carried virulence gene(s), and in the small pig

farm subgroup the prevalence was 18.82% (16/85). The differences between the subgroups were not significantly (p = 0.547, *Chi*-square) (**Table 3**). Finally, significant differences were not found in diversity index between these two subgroups, based on STs (**Table 2**).

The Chinese subset versus the UK subset. Human disease associated serotypes (HDAS) were found at a significantly higher prevalence (39.71%, 83/209) of the Chinese filtered subset compared to the UK filtered subset (23.20%, 29/125) (p = 0.003, *Chi*-square). However, there was no significant difference in prevalence of PVGT isolates, nor in the diversity index based on STs, between the two filtered subsets (**Tables 2 and 3**).

Discussion

Increased awareness of *S. suis* as a zoonotic pathogen has directed attention to the role of pigs as healthy carriers of *S. suis*, enabling the transmission to humans by exposure to unprocessed pig meat or pork products and also through close contact with pigs (18). The strikingly high incidence of human cases of *S. suis* in some Asian countries such as Vietnam, versus non-Asian countries, was linked to culturally-based practices including consumption of lightly cooked pig meat products (24) but, so far, there has been much less focus on other potential predisposing differences in prevalence and or diversity of carriage of *S. suis* among healthy slaughter pigs. China and UK each has a large pig industry but with different farming systems and differing climates. Unlike the UK, China has reported occasional large-scale outbreaks of human disease (20). In this study, we provided the comparative characterization of the prevalence, population structure and diversity of *S. suis* among clinically healthy pig herds entering the pig meat supply chain in Asian and non-Asian countries. Direct comparison of the

301 Chinese and the UK collections was hampered by subtle but potentially important
302 differences in the methods used for isolate collection – notably up to six colonies
303 were selected per swab in China, but only three in the UK. This was addressed by
304 applying filters to the two country level collections to result in a reduced but more
305 comparable subset of isolates.
306 This one-year longitudinal study found that *S. suis* is widely carried in clinically
307 healthy pig herds of China and UK, with no significant difference in positive isolation
308 rate for *S. suis* among the 10 Chinese pig farms and the 5 UK pig farms. It should be
309 noted that the use of API Strep 20 kit as a selection step for suspected *S. suis* isolates
310 prior to their confirmation by whole genome sequencing data might have resulted in
311 the exclusion of true *S. suis* isolates from the study (25), with a systematic
312 underestimation of true prevalence across the collections. The Chinese collection had
313 18 different serotypes based on molecular typing, 9 of which carried one or more of
314 the virulence genes *epf, mrp* and *sly.* The UK collection contained 15 different
315 serotypes, only 3 of which contained *epf, mrp* and (or) *sly* genes (**Table 4**). Serotype
316 16 was the most prevalent serotype in both countries, followed by serotypes 15 and 31
317 in China and serotypes 9 and 15 in the UK respectively. Since, serotypes 16 and 31
318 have been reported in human infections (9, 10) we used a filtered subset to directly
319 compare the prevalence of human disease associated serotypes (HDAS) in the
320 Chinese and UK collections and found that the Chinese collection had more HDAS
321 serotypes represented, and also a significantly higher proportion of HDAS over all,
322 compared to the UK (China: 83/209 (39.71%) isolates including serotypes 1&14,
323 2&1/2, 4, 5, 16, 24, 21 and 31; UK: 29/125 (23.20%) isolates including serotypes 4, 5,
324 16, 21 and 31, p = 0.003, *Chi*-square). These findings indicate that healthy pig
325 populations of UK and China are important reservoirs of zoonotic *S. suis* and also that

geographic differences in the prevalence of HDAS isolates of *S. suis*, and therefore the risk for zoonotic transmission, may exist.

A key finding of this study was the significant association between air temperature not only with positive isolation rate at farm level, but also with increased prevalence of HDAS isolates and of isolates carrying virulence associated genes – a finding that sheds light on previous observations of increased levels of *S. suis* on pig meat in hot weather (26) and potential associations with increased incidence of human disease in summer time (21). Observations on the prevalence of isolates bearing the virulence associated genes *sly*, *epf* and *mrp* should be tempered by the caveat that associations between the presence of these genes and virulence is has strictly only be studied for serotype 2 isolates (27). Sampling occasions on relatively hot days had significantly higher positive isolation rate than on cold days, and furthermore the proportion of isolates carrying virulence genes was also significantly higher in the high temperature subgroup in China. Meanwhile, increased temperature was associated with increased prevalence of HDAS isolates in the UK collection. These results align with observations that *S. suis* infection in humans occurs more frequently during the warmer months of the year but also emphasize the potential importance of enhanced carriage and shedding of zoonotic *S. suis* from healthy pigs, and potential contamination of pig meat, during warmer months. The biological processes underlying this observed effect of temperature on prevalence of disease associated *S. suis* are worthy of deeper investigation. The relationship is likely to be complex since an opposite relationship between temperature and transmission of the human pathogen *Streptococcus pneumonia* was recently described, with transmission and prevalence of carriage being enhanced during cooler and drier months across studied populations (28). That study tracked regional monthly minimum temperatures rather

than local mean temperatures on the specific day of sampling, as in the current study, and did not include an analysis of diversity of carriage at individual host or population level. Recent comparative genomic studies of disease associated and non-disease associated *S. suis* found a significantly reduced genome size among disease associated isolates (3). It might be hypothesized that disease associated isolates have comparatively reduced competitive fitness in the upper respiratory tract and or environment, with survival and successful transmission of this globally prevalent bacterium being supported by higher ambient temperatures. Nevertheless, although a statistical association cannot be interpreted as causality these findings indicate the potential for further exploration of the relationship between ambient temperature and infection dynamics of *S. suis*.

The effects of the age of pigs and the farm type were also evaluated as factors affecting positive isolation rate of *S. suis*. No significant difference was found between intensive and smaller pig farms in China but both types were operated in continuous flow systems. A larger and more detailed risk factor analysis might reveal underlying managemental effects. Although not statistically significant, markedly more *S. suis* isolates were obtained from 5 weeks old pigs than from 20 weeks old pigs in both countries (**Table 2**). Aligning this finding with the reported observation that clinical *S. suis* disease of pigs is most prevalent in the post-weaning period (29) – when pigs are commonly mixed and maternal derived passive antibody titres are declining, whereas *S. suis* associated disease is less prevalent in older pigs approaching slaughter age, suggests that future and larger studies focused on carriage of disease associated strains might reveal subtle but significant age associations with positive isolation rate.

The extraordinary diversity of *S. suis* has been described elsewhere (3). Our findings

376 confirmed that it was not rare that an individual pig could carry multiple *S. suis*
377 isolates (strains) as determined by different MLSTs, serotypes or different virulence
378 genotypes. The higher temperature sampling occasions were associated with a
379 statistically increased number of isolates recovered per pig, based on MLSTs,
380 highlighting the likelihood that any temperature effect on diversity of *S. suis*
381 populations is active at the individual pig level, and not only at pig population level.
382 The isolates of *S. suis* from clinical healthy pigs in this study also showed a high
383 diversity at pig population level, again based on MLST typing (**Fig. 3**). In total, 163
384 STs were identified in total from the China and UK collections in this study, only 12
385 of which were previously reported. In our study, the sequence types found among
386 serotypes with strong potential of virulence (serotypes 2&1/2 and 1&14 isolates)
387 found in the Chinese collection were ST1, ST7 and ST11, similar to other published
388 reports (16, 30, 8). It suggests that *S. suis* from the healthy pigs are more diverse than
389 those from diseased pigs. Similar finding has been mentioned in a previous study in
390 which 115 STs were identified among 179 *S. suis* isolates from throat swabs of
391 healthy pigs (31).
392 The low overlap of serotypes and STs from pigs sampled on the same farm at 5 and
393 20 weeks (**Fig. 3**) was unexpected and might reflect high rates of exchange and
394 infection dynamics between sampled and non-sampled pigs or the environment; or it
395 might be a consequence of the low sensitivity of the sampling approach taken – it was
396 practical to pick and characterize only a small number of isolates from each tonsil
397 swab culture. Future studies, exploiting direct metagenomic sequencing methods
398 might enable greater depth of characterization of the highly diverse populations of *S.*
399 *suis* in healthy and diseased pig populations. Nevertheless, further analysis of the
400 genome sequences and background data of the isolates from these new collections

promises to shed further light on the genomic diversity and evolutionary characteristics of this important pathogen.

Taken together, this study highlighted the high diversity and carriage prevalence of *S. suis* among clinically healthy pig herds of China and UK. The significant effect of temperature on prevalence of isolation, prevalence of human disease associated serotypes and virulence genotypes, and diversity at individual pig level may shed new light on geographic variations in human *S. suis* associated disease.

Materials and Methods

Sample collection and *S. suis* isolation.

Sampling and S. suis isolation in China. Five intensive pig farms (500-3000 sows, Herd 1 to Herd 5) and 5 traditional small pig farms (70-300 sows, Herd 6 to Herd 10) were sampled to study the prevalence of *S. suis* in different farm types. The intensive pig farms (> 500 sows) followed standardized management with modern feeding systems while the traditional small pig farms (< 300 sows) used variable management and feeding methods. All 10 farms held slaughter pigs from farrowing until slaughter at around 5-6 months of age on a continuous flow basis. To study any effect of pig age on *S. suis* carriage or diversity tonsil scrapes were collected twice from each farm - at 5 weeks of age and again at 20 weeks, sampling from between 23 and 27 individuals randomly selected from the same group within the herd. Finally, this total of 20 sampling visits were scheduled across different seasons from March 2013 to March 2014, enabling each sample set to be classified into different groups according to ambient air temperature, farm type and pig age. Tonsil scrape material was transferred to swabs and then cultured on Columbia Sheep Blood Agar at 37℃ for 24 h aerobically. Six colonies with alpha-haemolytic activity and compatible colony

morphology were then sub-cultured. Isolates were biochemically profiled using the API 20 Strep kit (bioMérieux) and those with positive results were stored as *S. suis*.

Sampling and S. suis isolation in the UK. Five intensive farms were sampled, each holding pigs from weaning at 4 weeks of age until slaughter at 20 - 25 weeks of age on an all-in all-out basis. In accordance with the common sampling protocol applied in China, swabs from tonsil scrapes were collected at 5 and 20 weeks of age from 25 randomly selected pigs from the same group within the herd. Swabs were cultured on Columbia Sheep Blood Agar in aerobic conditions at 37℃ for 24 h. A maximum of three alpha-haemolytic colonies per plate, with compatible *S. suis* morphology, were sub-cultured and profiled with a biochemical kit (API 20 Strep kit, bioMérieux) and those with positive results stored as *S. suis*.

Subset for direct comparison of China and UK isolates. In order to directly compare prevalence and diversity of *S. suis* between the UK and China, a subset of each collection was generated by filtering using the following rules: (i) Since 6 Chinese colonies were picked per swab for identification compared to 3 in the UK, if more than 3 *S. suis* isolates were obtained from a single swab, only the first 3 isolates were kept to limit the *S. suis* isolates obtained from a swab from either country to not more than 3. (ii) For this maximum of 3 isolates selected per pig, where any of these isolates shared the same API profile then only one representative of a given API profile per pig was kept. The filtered Chinese subset included 209 isolates while the UK subset included 125 isolates.

Data grouping.

The sampling points under different conditions were grouped as 3 pairs of 'subgroups' (age, farm type and temperature subgroups for UK and for China). Based

on the age of sampled pigs, the 20 sampling points of China and the 10 sampling points of UK were divided into 2 pairs of 'age subgroups'; '5 weeks old subgroup' and '20 weeks old subgroup', respectively. The 'farm type subgroups' were relevant only to the Chinese collection, and included 10 sampling points of the 5 intensive pig farms ('intensive pig farm subgroup') and 10 sampling points of small pig farms ('small pig farm subgroup'). As a stable and unambiguous attribution, in this study, air temperature was used to represent the seasonal variation of the *S. suis* positive isolation rate or diversity. The average value of the daily maximum and the minimum temperature was used to represent the air temperature of the sampling day. For the three-way ANOVA, the different temperature values of the Chinese and the UK sampling occasions were divided into binary groups by the mean values respectively (China = 20.8℃, UK = 11.9℃). The 'high temperature subgroups' included sampling points where the temperature was above the mean value, and 'low temperature subgroups' represented sampling points that were equal to or below the mean value. The binary temperature grouping strategy was also used when evaluating the diversity difference.

Sequencing and assembly.

For Chinese isolates, genomic DNA was prepared from isolates grown overnight at 37℃ in Tryptone Soy Broth (TSB, BD Biosciences) plus 10% bovine serum, using Bacterial DNA kits (OMEGA bio-tek). Genomic DNA was prepared from UK isolates grown overnight at 37℃ in Todd-Hewitt broth plus 0.2% yeast (Oxoid Ltd.) using a MasterPure Gram Positive DNA isolation kit (Epicentre). All genomic DNA samples were qualified with an OD260/280 ratio between 1.8 and 2 (Nanodrop, Thermo Scientific). Genomic DNA (typically 500ng) was used to prepare multiplexed libraries

476 (32, 33) for sequencing on Illumina HiSeq 2000 instruments operated according to the
477 manufacturer's instructions with 100 cycle paired end runs.
478 For the *S. suis* genomes of both countries a Perl program, Fastq_screen
479 (http://www.bioinformatics.babraham.ac.uk/projects/fastq_screen), was used to map
480 the raw reads to published genomes of *S. suis*. Samples with high levels of unmapped
481 data (> 50%) were confirmed by blasting against the NCBI genome database and
482 those found to be contaminated with genomes from other species were excluded from
483 further analysis. Genome assemblies of China and UK isolates were generated using
484 SPAdes 3.5.0 and Velvet 1.2.08 (with Velvet Optimiser 2.2.5), respectively.
485 Assemblies with an n50 < 15000 were discarded.
486
487 **Molecular serotyping.**
488 A BLAST database was built, based on a published multiple PCR method (14),
489 including all unique *cps* genes for *S. suis* serotype identification. Cutoff values were
490 used to determine gene presence or absence (sequence homology of > 95% with an
491 alignment length > 95% of the target gene). The *cps* genes of serotype 2 and 1/2,
492 serotype 1 and 14 are too similar to distinguish at the draft genomic level (14) thus
493 serotypes 2 and 1/2 were grouped together, as were serotypes 1 and 14. The non-
494 serotypeable isolates were further classified into 8 different groups according to 8
495 novel *cps* loci (NCL) (15). The isolates that could not be typed by both methods were
496 recorded as “non-typeable”.
497
498 **Multilocus sequence typing.**
499 A database including all published alleles of 7 housekeeping gene fragments was built
500 to run the blast process (16). Blast results with 100% homology were treated as the

same alleles, while results with sequence homology > 80% and alignment length > 80% were considered as new alleles. Isolates yielding results with lower similarity were not assigned a sequence type (ST) and were treated as MLST non-typeable isolates, as the corresponding housekeeping genes were considered to be absent.

Genotyping of *mrp, sly* and *epf* gene.

The full length of *mrp*, *sly* and *epf* genes were extracted from the draft genome sequences. Based on the three classic virulence associated genes, the genotype of the three virulence genes, namely the 'virulence genotype' (VGT), of each isolate was detected. Sequences with homology of > 80% and alignment length of > 80% of the corresponding reference gene were considered as alleles. Isolates in which none of the three virulence genes were detected were termed 'negative VGT' (NVGT), otherwise isolates were termed 'positive VGT' (PVGT).

Statistical analysis.

Statistical analyses were performed with SigmaPlot software (version 12.0). A value of $p \leq 0.05(5)$ was considered to be significant in this study. ANOVA, t-test or paired t-test were used to evaluate the influence of different factors on the distribution of *S. suis* populations, while *Chi*-square test was used to evaluate the differences between the composition of different *S. suis* populations. For the data that failed to pass the normality test, Mann-Whitney rank sum test, Wilcoxon signed rank test and the ANOVA on Ranks were used instead of t-test, paired t-test and ANOVA respectively.

Analysis of diversity.

Diversity was analysed at multiple levels; namely individual pig level, group level

and country level. At pig level diversity was assessed for the China and UK collections separately by comparing the number of different STs obtained from each tonsil scrape. Simpson's diversity indexes (Ds), again based on STs, were used to evaluate the diversity of the *S. suis* populations at subgroup and country level. The STs grouped into a Clonal Complex were calculated as the same type. The MLST non-typeable isolates with different ST genes patterns were also identified as different types. Due to differences in the isolation protocols in China and the UK, the diversity between the subgroups within countries was calculated first, and then diversity at country level was compared directly using the filtered subsets produced as described above. The Simpson's diversity indexes of different *S. suis* subgroups were calculated using the equation defined by Hunter and Gaston (34):

$$\boldsymbol{D_s = 1 - [\frac{1}{N(N-1)}] \sum_{j=1}^{s} n_j(n_j - 1)}$$

D_s represents the probability of two independent isolates being placed into different types. N is the total number of the population. S is the total number of the types while n_j is number of the individuals belonging to the j^{th} type. The variance for Ds (σ^2) was estimated with the equation:

$$\boldsymbol{\sigma^2 = \frac{4}{N}\left(\sum_{j=1}^{s} \left(\frac{n_j}{N}\right)^3 - \left[\sum_{j=1}^{s} \left(\frac{n_j}{N}\right)^2\right]^2 \right)}$$

The Confidence Interval (CI) of the Ds was used to evaluated the statistical significance when comparing the diversity of different populations. The difference between two Ds without overlapped CI was treated as significant. According to the Chebyshev's Theorem, for normal distributions, about 95% of the results will fall between +2 and -2 standard deviations from the mean. Thus, the proposed equation of the CI is:

$$\mathbf{CI} = \left(\boldsymbol{D_s} - \mathbf{2}\sqrt{\boldsymbol{\sigma^2}}\,,\ \boldsymbol{D_s} + \mathbf{2}\sqrt{\boldsymbol{\sigma^2}} \right)$$

547

548 **Accession numbers**

549 The GenBank accession numbers of the genes used in this study are: *epf*
550 (AY341262.1, https://www.ncbi.nlm.nih.gov/nuccore/AY341262.1), *mrp*^{EU}
551 (X64450.1, https://www.ncbi.nlm.nih.gov/nuccore/X64450.1), *mrp*^{NA2} (FJ685609.1,
552 https://www.ncbi.nlm.nih.gov/nuccore/FJ685609.1), *sly* (Z36907.1,
553 https://www.ncbi.nlm.nih.gov/nuccore/Z36907.1) and the *mrp*^{NA1} reference gene was
554 extracted from the Canadian isolate 89/1591 (GenBank accession: AAFA00000000,
555 https://www.ncbi.nlm.nih.gov/genome/?term=AAFA00000000) (35). The GenBank
556 accession numbers of the *S. suis* genome sequences of 223 Chinese isolates and the
557 Short Read Archive (SRA) accession numbers of the *S. suis* genome sequences of 127
558 UK isolates are in the supplemental material **Table S1**.

559

560 ## Acknowledgments

561 This study was supported by the International S & T Cooperation Program of China
562 (ISTCP grant No. 2013DFG32360 for RZ), the Biotechnology and Biological
563 Sciences Research Council (BBSRC grant No. BB/L003902/1 for AT), National Key
564 R&D Program of China (2017YFD0500201 for RZ) and the Hubei Province Natural
565 Science Foundation for Innovative Research Groups (2016CFA015 for RZ).

566

References

1. **Feng YJ, Zhang HM, Wu ZW, Wang SH, Cao M, Hu D, Wang CJ.** 2014. *Streptococcus suis* Infection. Virulence **5**(4):477-497.
2. **Gottschalk M, Xu JG, Calzas C, Segura M.** 2010. *Streptococcus suis*: a new emerging or an old neglected zoonotic pathogen? Future Microbiol. **5**(3):371-391.
3. **Weinert LA, Chaudhuri RR, Wang JH, Peters SE, Corander J, Jombart T, Baig A, Howell KJ, Vehkala M, Valimaki N, Harris D, Chieu TTB, Chau NVV, Campbell J, Schultsz C, Parkhill J, Bentley SD, Langford PR, Rycroft AN, Wren BW, Farrar J, Baker S, Hoa NT, Holden MTG, Tucker AW, Maskell DJ.** 2015. Genomic signatures of human and animal disease in the zoonotic pathogen *Streptococcus suis*. Nat. Commun. **6**:6740. doi:10.1038/ncomms7740.
4. **Hill JE, Gottshalk M, Brousseau R, Harel J, Hemmingsen SM, Goh SH.**2005. Biochemical analysis, cpn60 and 16S rDNA sequence data indicate that *Streptococcus suis* serotypes 32 and 34, isolated from pigs are *Streptococcus orisratti*. Vet. Microbiol. **107**:63–69.
5. **Ishida S, Tien LHT, Osawa R, Tohya M, Nomoto R, Kawamura Y, Takahashi T, Kikuchi N, Kikuchi K, Sekizaki T.** 2014. Development of an appropriate PCR system for the reclassification of *Streptococcus suis*. J. Microbiol. Meth. **107**:66-70.
6. **Baig A, Weinert LA, Peters SE, Howell KJ, Chaudhuri RR, Wang J, Holden MT, Parkhill J, Langford PR, Rycroft AN, Wren BW, Tucker AW, Maskell DJ.** 2015. Whole genome investigation of a divergent clade of the pathogen *Streptococcus suis*. Front. Microbiol. **6**:1191. doi: 10.3389/fmicb.2015.01191.

7. **Okura M, Osaki M, Nomoto R, Arai S, Osawa R, Sekizaki T, Takamatsu D.** 2016. Current Taxonomical Situation of *Streptococcus suis*. Pathogens. 24;5(3).

8. **Huang JH, Shang KX, Kashif J, Wang LP.** 2015. Genetic diversity of *Streptococcus suis* isolated from three pig farms of China obtained by acquiring antibiotic resistance genes. J. Sci. Food Agr. **95**:1454-1460.

9. **Huong VTL, Ha N, Huy NT, Horby P, Nghia HDT, Thiem VD, Zhu XT, Hoa NT, Hien TT, Zamora J, Schultsz C, Wertheim HFL, Hirayama K.** 2014. Epidemiology, clinical manifestations, and outcomes of *Streptococcus suis* infection in humans. Emerg. Infect. Dis. **20**(7):1105-1114.

10. **Callejo R, Prieto M, Salamone F, Auger JP, Goyette-Desjardins G, Gottschalk M.** 2014. Atypical Streptococcus suis in man, Argentina, 2013.Emerg Infect Dis. 20(3):500-2.

11. **Callejo R, Zheng H, Du P, Prieto M, Xu J, Zielinski G, Auger JP, Gottschalk M.** 2016. Streptococcus suis serotype 2 strains isolated in Argentina (South America) are different from those recovered in North America and present a higher risk for humans. JMM Case Rep. 2016. 3(5): e005066.

12. **Goyette-Desjardins G, Auger JP, Xu J, Segura M, Gottschalk M.** 2014. Streptococcus suis, an important pig pathogen and emerging zoonotic agent-an update on the worldwide distribution based on serotyping and sequence typing. Emerg Microbes Infect. 3(6):e45.

13. **Hatrongjit R, Kerdsin A, Gottshalk M, Takeuchi D, Hamada S, Oishi K, Akeda Y.**2015. First human case report of sepsis due to infection with *Streptococcus suis* serotype 31 in Thailand. BMC Infect. Dis. **15**:392.

14. **Okura M, Lachance C, Osaki M, Sekizaki T, Maruyama F, Nozawa T, Nakagawa I, Hamada S, Rossignol C, Gottschalk M, Takamatsu D.** 2014.

Development of a two-step multiplex PCR assay for typing of capsular polysaccharide synthesis gene clusters of *Streptococcus suis*. J. Clin. Microbiol. **52**(5):1714-9.

15. **Zheng H, Ji SB, Liu ZJ, Lan RT, Huang Y, Bai XM, Gottschalk M, Xu JG.** 2015. Eight novel capsular polysaccharide synthesis gene loci identified in nontypeable *Streptococcus suis* isolates. Appl. Environ. Microb. **81**(12):4111-4119.

16. **King SJ, Leigh JA, Heath PJ, Luque I, Tarradas C, Dowson CG, Whatmore AM.** 2002. Development of a multilocus sequence typing scheme for the pig pathogen *Streptococcus suis*: identification of virulent clones and potential capsular serotype exchange. J. Clin. Microbiol. **40**(10):3671-3680.

17. **Segura M, Fittipaldi N, Calzas C, Gottschalk M.** 2017. Critical Streptococcus suis Virulence Factors: Are They All Really Critical? Trends Microbiol. 2017 Jul;**25**(7):585-599.

18. **Goyette-Desjardins G, Auger JP, Xu JG, Segura M, Gottschalk M.** 2014. *Streptococcus suis*, an important pig pathogen and emerging zoonotic agent-an update on the worldwide distribution based on serotyping and sequence typing. Emerg. Microbes. Infect. **3**(6):e45. doi:10.1038/emi.2014.45.

19. **Tang J, Wang C, Feng Y, Yang W, Song H, Chen Z, Yu H, Pan X, Zhou X, Wang H, Wu B, Wang H, Zhao H, Lin Y, Yue J, Wu Z, He X, Gao F, Khan AH, Wang J, Zhao GP, Wang Y, Wang X, Chen Z, Gao GF.** 2006. Streptococcal toxic shock syndrome caused by *Streptococcus suis* serotype 2.PLoS Med. **3**(5):e151.

20. **Yu H, Jing H, Chen Z, Zheng H, Zhu X, Wang H, Wang S, Liu L, Zu R, Luo L, Xiang N, Liu H, Liu X, Shu Y, Lee SS, Chuang SK, Wang Y, Xu J, Yang**

642 **W.** 2006. Human *Streptococcus suis* outbreak, Sichuan, China. Emerg. Infect.
643 Dis. **12**(6):914-20.
644 21. **Ma E, Chung PH, So T, Wong L, Choi KM, Cheung DT, Kam KM, Chuang**
645 **SK, Tsang T.** 2008. *Streptococcus suis* infection in Hong Kong: an emerging
646 infectious disease? Epidemiol. Infect. **136**(12):1691-7.
647 22. **Wertheim HF, Nguyen HN, Taylor W, Lien TT, Ngo HT, Nguyen TQ, Nguyen**
648 **BN, Nguyen HH, Nguyen HM, Nguyen CT, Dao TT, Nguyen TV, Fox A,**
649 **Farrar J, Schultsz C, Nguyen HD, Nguyen KV, Horby P.** 2009. *Streptococcus*
650 *suis*, an important cause of adult bacterial meningitis in northern Vietnam. PLoS
651 One. **4**(6):e5973. doi:10.1371/journal.pone.0005973.
652 23. **Wertheim HF, Nghia HD, Taylor W, Schultsz C.** 2009. *Streptococcus suis*: an
653 emerging human pathogen. Clin. Infect. Dis. **48**(5):617-25.
654 24. **Nghia HD, Tu le TP, Wolbers M, Thai CQ, Hoang NV, Nga TV, Thao le TP,**
655 **Phu NH, Chau TT, Sinh DX, Diep TS, Hang HT, Truong H, Campbell J,**
656 **Chau NV, Chinh NT, Dung NV, Hoa NT, Spratt BG, Hien TT, Farrar J,**
657 **Schultsz C.** 2011. Risk factors of *Streptococcus suis* infection in Vietnam. A
658 case-control study. PLoS One. **6**(3):e17604. doi: 10.1371/journal.pone.0017604.
659 25. **Gottschalk M, Higgins R, Jacques M, Beaudoin M, Henrichsen J.** 1991.
660 Isolation and characterization of Streptococcus suis capsular types 9-22. J Vet
661 Diagn Invest. 3(1):60-5.
662 26. **Cheung PY, Lo KL, Cheung TT, Yeung WH, Leung PH, Kam KM.** 2008.
663 *Streptococcus suis* in retail markets: How prevalent is it in raw pork? Int. J. Food
664 Microbiol. **127**:316-320.
665 27. **Staats JJ, Plattner BL, Stewart GC, Changappa MM.** 1999. Presence of the
666 Streptococcus suis suilysin gene and expression of MRP and EF correlates with

667 high virulence in Streptococcus suis type 2 isolates. Vet Microbiol. 70(3-4):201-
668 11.
669 28. **Numminen E, Chewapreecha C, Turner C, Goldblatt D, Nosten F, Bentley**
670 **SD, Turner P, Corander J**. 2015. Climate induces seasonality in pneumococcal
671 transmission. Sci. Rep. **5**:11344.
672 29. **Gottschalk M.** 2012. Streptococcosis, p. 841–853. In Zimmerman JJ, Karriker
673 LA, Ramirez A, Schwartz KJ and Stevenson GW (ed.), Diseases of swine, 10th
674 ed. John Wiley & Sons, Inc.
675 30. **Wang KC, Zhang W, Li XC, Lu CP, Chen JM, Fan WX, Huang BX.** 2013.
676 Characterization of *Streptococcus suis* isolates from slaughter swine. Curr.
677 Microbiol. **66**:344–349.
678 31. **Zheng H, Ji SB, Lan RT, Liu ZJ, Bai XM, Zhang W, Gottschalk M, Xu JG.**
679 **2014.** Population analysis of Streptococcus suis isolates from slaughtered swine
680 by use of minimum core genome sequence typing. J. Clin. Microbiol. 2014,
681 **52**(10): 3568-3672.
682 32. **Quail MA, Kozarewa I, Smith F, Scally A, Stephens PJ, Durbin R, Swerdlow**
683 **H, Turner DJ.** 2008. A large genome center's improvements to the Illumina
684 sequencing system. Nat. Methods **5**(12):1005-10.
685 33. **Quail MA, Otto TD, Gu Y, Harris SR, Skelly TF, McQuillan JA, Swerdlow**
686 **HP, Oyola SO.** 2011. Optimal enzymes for amplifying sequencing libraries. Nat.
687 Methods **9**(1):10-1.
688 34. **Hunter PR, Gaston MA.** 1988. Numerical index of the discriminatory ability of
689 typing systems: an application of Simpson's index of diversity. J. Clin. Microbiol.
690 **26**(11):2465-2466.
691 35. **Fittipaldi N, Fuller TE, Teel JF, Wilson TL, Wolfram TJ, Lowery DE,**

692 **Gottschalk M.** 2009. Serotype distribution and production of muramidase-
693 released protein, extracellular factor and suilysin by field strains of *streptococcus*
694 *suis* isolated in the united states. Vet. Microbiol. **139**:310–317.
695

Figure legends

Fig. 1 The distribution of *S. suis* positive and negative samples across the sampling occasions accounting for different pig ages and air temperature. Kernel density estimation was used to describe the bivariate distribution of 500 Chinese and 250 UK samples across the air temperatures of the sampling times and the ages of the sampled pig herds.

Fig.2 The molecular serotyping composition of different *S. suis* subgroups. The absence/presence and size of the bubbles represents the number of isolates of each serotype in each population. The red colour indicates serotypes that have previously been obtained from human cases of disease. The Chinese filtered subset and the Chinese high temperature subgroup had all 8 human disease associated serotypes. Serotype 2 & 1/2 isolates were identified in all the Chinese subgroups but not in the UK collation. Serotype 16 isolates could be identified in all subgroups.

Fig. 3 The organization of STs within the *S. suis* population. To show the relationship between the STs of the whole *S. suis* population in this study, the 93 STs of Chinese *S. suis* collection and 70 STs of UK *S. suis* collection were clustered with eBURST (http://eburst.mlst.net/). Each dot in the figure stand for a ST, the STs with 6/7 identical alleles were treated as clonal complexes (CC) and connected with lines. The size of dots represents the number of isolates. All previously reported 12 STs are labeled. There were 7 STs (all new STs) of UK and 6 STs of Chinese collection (4 reported STs and 2 new STs) identified in more than 1 pig farms. 4 STs (3 of Chinese collection and 1 of UK collection) could be identified from both 5 weeks old and the corresponding 20 weeks old sampling occasions.

Table 1. Distribution of *S. suis* positive isolation rates, mean temperatures and farm types in China and UK.

Country	Farm ID	Farm type	5 weeks old					20 weeks old				
			MT (°C)	Size	N (F)	PP	PIR%	MT (°C)	Size	N (F)	PP	PIR (%)
CHINA	C1	Intensive	29.5	23	10 (9)	6	26.09	29	25	26 (21)	11	44.00
	C2	Intensive	32	27	33 (30)	17	62.96	19.5	25	1 (1)	1	4.00
	C3	Intensive	29.5	25	28 (26)	15	60.00	17.5	25	2 (2)	2	8.00
	C4	Intensive	15.5	25	12 (12)	9	36.00	4.8	25	2 (2)	2	8.00
	C5	Intensive	15.5	25	12 (12)	10	40.00	4.8	25	12 (12)	8	32.00
	C6	Small	28.5	24	21 (19)	12	50.00	29	25	20 (19)	13	52.00
	C7	Small	29	24	11 (11)	8	33.33	28.5	25	6 (6)	4	16.00
	C8	Small	30	27	13 (13)	8	29.63	17.5	25	7 (7)	4	16.00
	C9	Small	14	25	2 (2)	2	8.00	14	25	1 (1)	1	4.00
	C10	Small	8.5	25	3 (3)	2	12.00	14	25	1 (1)	1	4.00
UK	U1	Intensive	14	25	17 (17)	11	44.00	15	25	14 (14)	10	40.00
	U2	Intensive	19	25	42 (40)	23	92.00	12.5	25	1 (1)	1	4.00
	U3	Intensive	16.5	25	23 (23)	17	68.00	4	25	9 (9)	8	32.00
	U4	Intensive	11.5	25	8 (8)	7	28.00	7.5	25	5 (5)	4	16.00
	U5	Intensive	7	25	1 (1)	1	4.00	7	25	7 (7)	7	28.00

Note: MT, mean temperature, the average of the highest and lowest air temperature of the sampling day; Size, sample size, the number of pigs sampled each time; N, the number of *S. suis* isolates; (F), the number of isolates in the filtered subset for direct comparison of China and UK data; PP, positive pigs, the number of *S. suis* positive pigs; PIR, positive isolation rate, the proportion of *S. suis* positive pigs.

Table 2. Genomic characteristics of the different *S. suis* populations.

Items	Number of isolates	Simpson's diversity index			The proportion of HDAS (%)	Virulence genotype pattern (%)(*epf* / *mrp* / *sly*)							
		D_s	CI-	CI+		- - -	- *mrp*^{NA2} *sly*	- *mrp*^{NA1} -	- *mrp*^{NA1} *sly*	- *mrp*^{EU} *sly*	*epf* *mrp*^{NA2} *sly*	*epf* *mrp*^{NA1} *sly*	*epf* *mrp*^{EU} *sly*
All S. suis isolates													
China_HTS	168	0.971	0.962	0.981	43.45	75.00	6.55	5.95	6.55	2.38	1.19	0.60	1.79
China_LTS	55	0.976	0.960	0.991	32.73	89.09	1.82	0	1.82	0	5.45	0	1.82
UK_HTS	97	0.965	0.947	0.982	28.86	85.57	8.25	0	5.15	0	0	1.03	0
UK_LTS	30	0.966	0.929	1.000	6.67	86.67	3.33	3.33	6.67	0	0	0	0
China_5 WOS	145	0.969	0.959	0.979	42.76	75.17	7.59	5.52	6.90	0	2.07	0.69	2.07
China_20 WOS	78	0.972	0.957	0.987	37.18	84.62	1.28	2.56	2.56	5.13	2.56	0	1.28
UK_5 WOS	91	0.960	0.940	0.980	28.57	84.62	8.79	0	5.49	0	0	1.098	0
UK_20 WOS	36	0.987	0.975	0.999	11.11	88.89	2.78	2.78	5.56	0	0	0	0
China_IPFS	138	0.972	0.964	0.981	39.86	76.81	7.25	3.62	7.25	0	1.45	0.72	2.90
China_SPFS	85	0.941	0.908	0.973	42.35	81.18	2.35	5.88	7.06	0	3.53	0	0
China_all	223	0.979	0.973	0.985	40.81	78.48	5.38	4.48	5.38	1.79	2.24	0.45	1.79
UK_all	127	0.972	0.959	0.984	23.62	85.83	7.09	0.79	5.51	0	0	0.79	0
Filtered S.suis isolates													
China_ filtered	209	0.980	0.974	0.986	39.71	78.47	5.26	4.78	5.47	1.44	2.39	0.48	1.44
UK_ filtered	125	0.972	0.960	0.985	23.20	85.60	7.20	0.80	5.60	0	0	0.80	0

Note: HDAS, human disease associated serotype; HTS, high temperature subgroup; LTS, low temperature subgroup; 5 WOS, 5 weeks old subgroup; 20 WOS, 20 weeks old subgroup; IPFS, intensive pig farm subgroup; SPFS, small pig farm subgroup; Ds, Simpson's diversity index; CI+, confidence interval upper limit, 95%; CI-, confidence interval lower limit, 95%.

Table 3. Summary of statistical analysis of the influence of environmental and management associated factors on the distribution and the composition of *S. suis* populations.

Analysis	Dataset	Temperature	Age	Farm type	Country
Positive isolation rate at farm level	China	*	–	–	/
	UK	–	–	/	/
	filtered subsets	/	/	/	–
Prevalence of pigs with multiple STs per swab	China	*	–	–	/
	UK	*	–	/	/
	filtered subsets	/	/	/	–
Prevalence of HDAS isolates	China	–	–	–	/
	UK	*	–	/	/
	filtered subsets	/	/	/	*
Prevalence of PVGT isolates	China	*	–	–	/
	UK	–	–	/	/
	filtered subsets	/	/	/	–

Note: HDAS, human disease associated serotype; PVGT, positive virulence genotype, the virulence genes including *epf*, *mrp* and or *sly*; ST, sequence type; *, statistically significant, $p < 0.05$; –, not significant; /, not involved in the analysis.

Table 4 Distribution of virulence factor genotype patterns described for each serotype.

Serotype/NCL type	Country	VG pattern (*epf* / *mrp* / *sly*)	Number of isolates
Serotype 1&14	China	*epf* mrp^{EU} *sly*	2
Serotype 2&1/2	China	*epf* mrp^{EU} *sly*	1
		epf mrp^{NA2} *sly*	3
Serotype3	China	- mrp^{NA1} -	5
Serotype4	China	- mrp^{NA1} *sly*	3
	UK	- - -	4
Serotype5	China	- - -	6
	UK	- - -	1
Serotype6	UK	- - -	9
Serotype7	UK	- mrp^{NA1} -	1
Serotype8	China	- mrp^{NA2} *sly*	8
	UK	- - -	2
Serotype9	China	- - -	2
	UK	- - -	13
Serotype10	UK	- - -	4
Serotype11	China	- mrp^{NA1} -	2
Serotype12	China	- mrp^{NA1} *sly*	1
		- - -	9
Serotype15	China	- - -	1
		- mrp^{NA1} -	3
		- mrp^{NA1} *sly*	7
		- mrp^{NA2} *sly*	4
		epf mrp^{NA1} *sly*	1
		epf mrp^{NA2} *sly*	1
	UK	- mrp^{NA1} *sly*	7
		- mrp^{NA2} *sly*	4
Serotype16	China	- - -	55
	UK	- - -	20
Serotype21	China	- - -	1
		epf mrp^{NA2} *sly*	1
	UK	- - -	2
Serotype24	China	- - -	1
Serotype25	China	- - -	1
	UK	- - -	8
Serotype28	China	- - -	7
	UK	- mrp^{NA2} *sly*	4
Serotype29	China	- - -	12
	UK	- - -	1
Serotype30	China	- - -	5
	UK	- - -	4
Serotype31	China	- - -	16
	UK	- - -	5
NCL1	China	- - -	2

NCL2	China	- - -	2
	UK	- - -	2
NCL3	China	- - -	5
	UK	- - -	1
NCL4	China	- - -	6
	UK	- - -	6
NCL5	China	- - -	2
	UK	- - -	2
NCL6	UK	- - -	1
NCL7	China	- - -	6
	UK	- - -	3
NCL8	China	- - -	14
	UK	- - -	2

Note: NCL, novel capsule locus (NCL) type; VG, virulence gene.

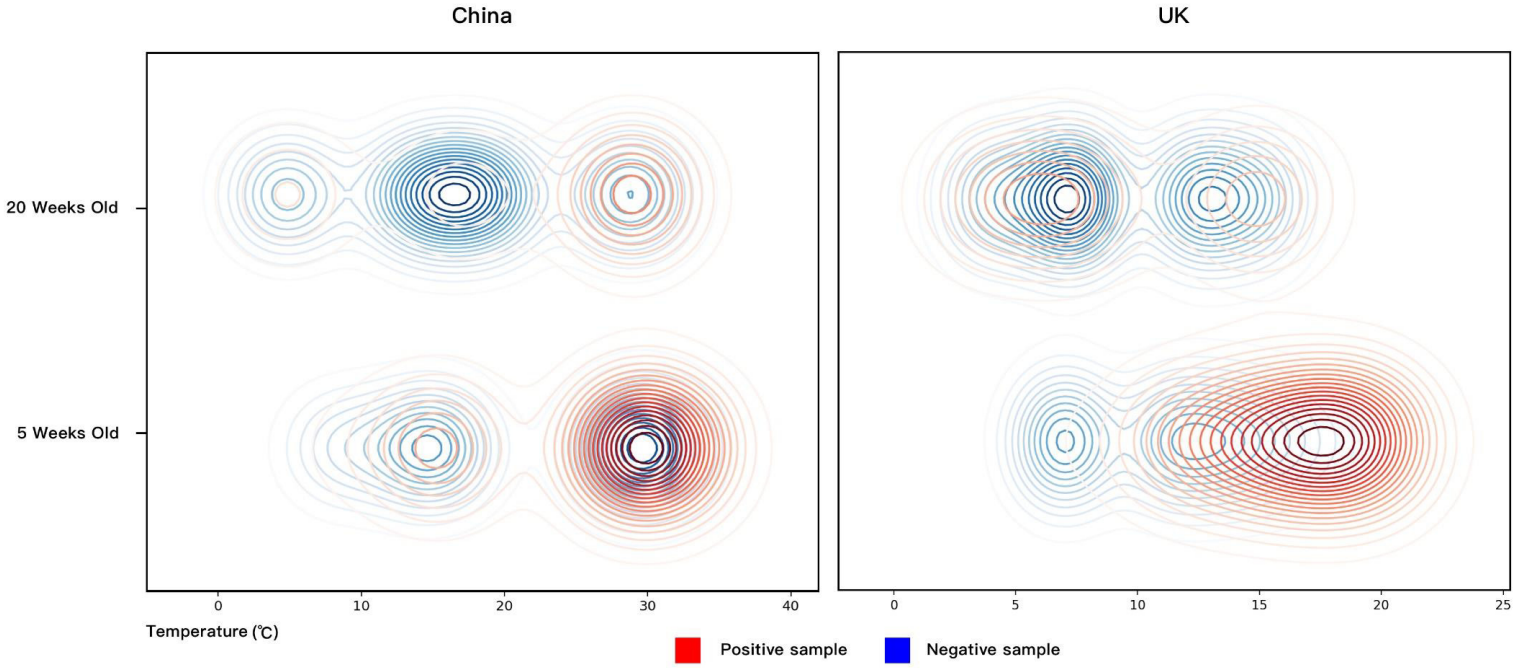

Fig. 1 The distribution of *S. suis* positive and negative samples across the sampling occasions accounting for different pig ages and air temperature. Kernel density estimation was used to describe the bivariate distribution of 500 Chinese and 250 UK samples across the air temperatures of the sampling times and the ages of the sampled pig herds.

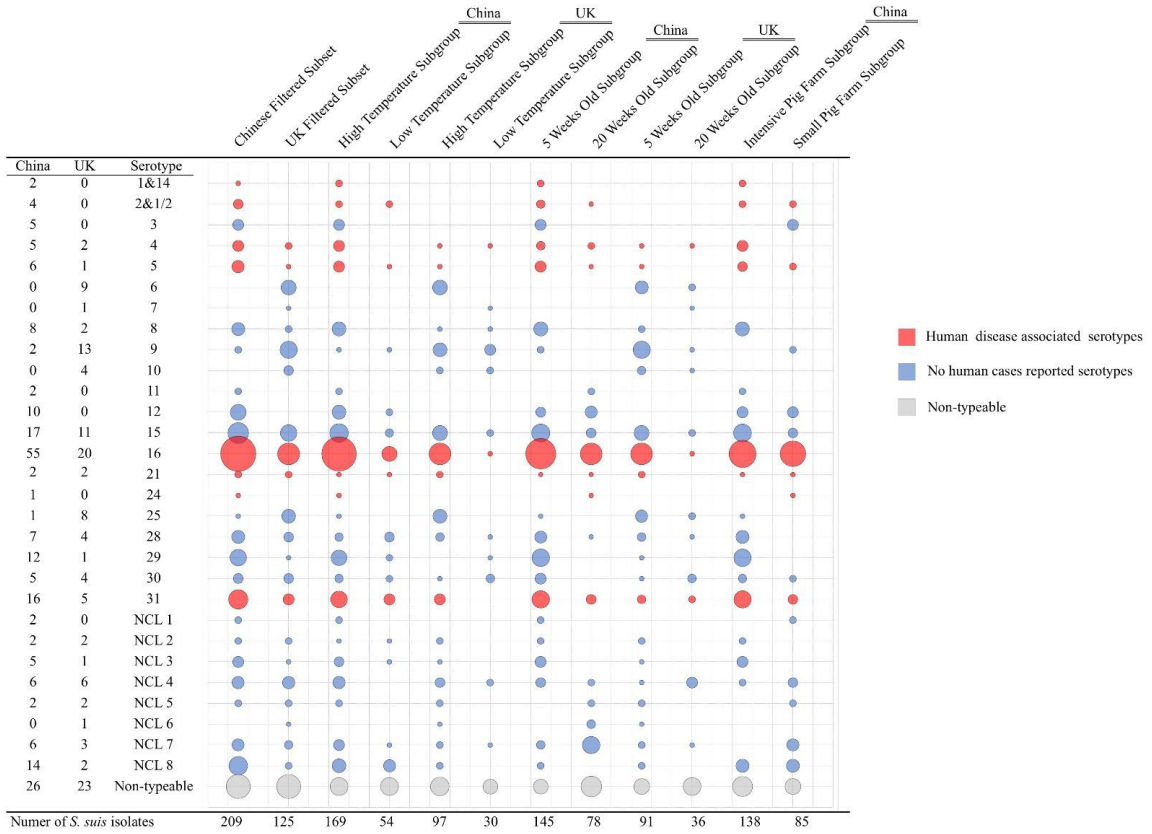

Fig.2 The molecular serotyping composition of different *S. suis* subgroups. The absence/presence and size of the bubbles represents the number of isolates of each serotype in each population. The red colour indicates serotypes that have previously been obtained from human cases of disease. The Chinese filtered subset and the Chinese high temperature subgroup had all 8 human disease associated serotypes. Serotype 2 & 1/2 isolates were identified in all the Chinese subgroups but not in the UK collation. Serotype 16 isolates could be identified in all subgroups.

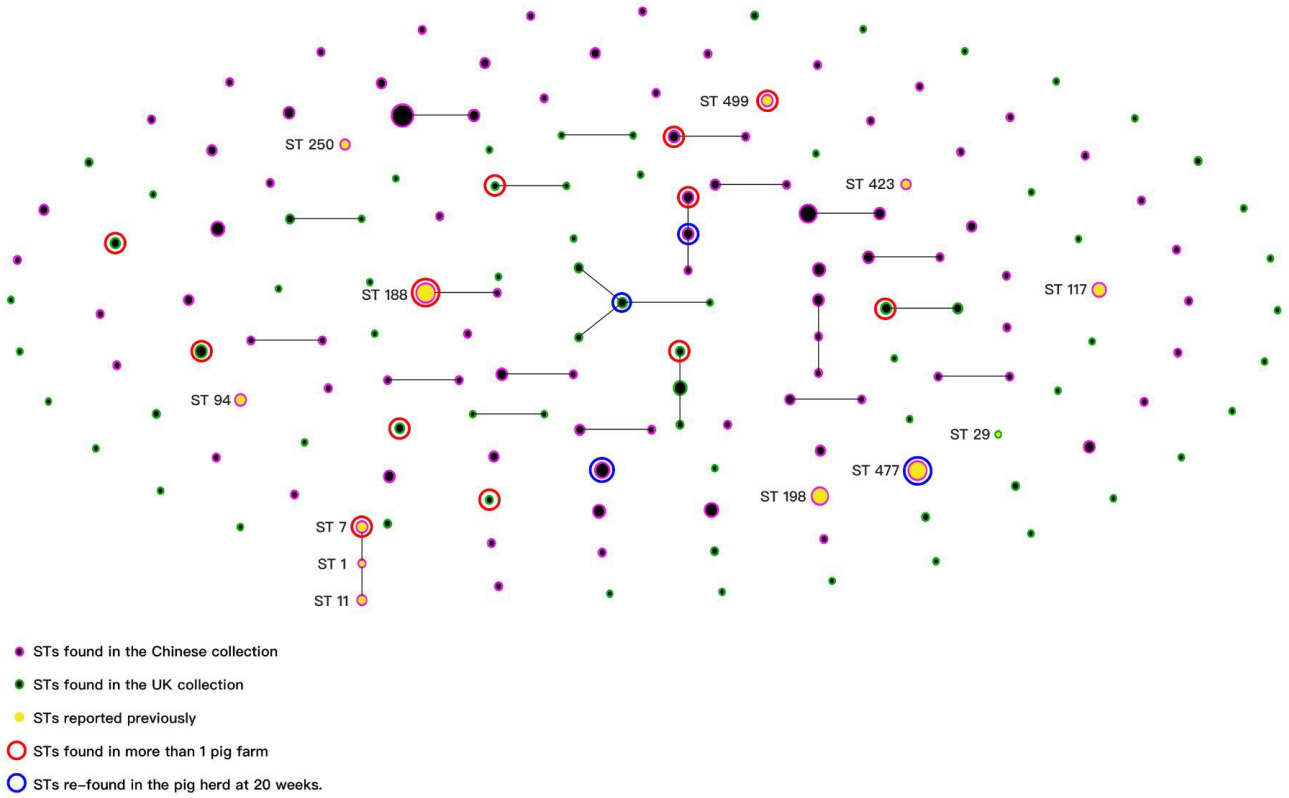

Fig. 3 The organization of STs within the *S. suis* population. To show the relationship between the STs of the whole *S. suis* population in this study, the 93 STs of Chinese *S. suis* collection and 70 STs of UK *S. suis* collection were clustered with eBURST (http://eburst.mlst.net/). Each dot in the figure stand for a ST, the STs with 6/7 identical alleles were treated as clonal complexes (CC) and connected with lines. The size of dots represents the number of isolates. All previously reported 12 STs are labeled. There were 7 STs (all new STs) of UK and 6 STs of Chinese collection (4 reported STs and 2 new STs) identified in more than 1 pig farms. 4 STs (3 of Chinese collection and 1 of UK collection) could be identified from both 5 weeks old and the corresponding 20 weeks old sampling occasions.