Linking phenotype, genotype and ecology: antimicrobial resistance in the zoonotic pathogen *Streptococcus suis*

Nazreen F. Hadjirin, Eric L. Miller, Gemma G. R. Murray, Phung L. K. Yen, Ho D. Phuc, Thomas M. Wileman, Juan Hernandez-Garcia, Susanna M. Williamson, Julian Parkhill, Duncan J. Maskell, Rui Zhou, Nahuel Fittipaldi, Marcelo Gottschalk, A.W. (Dan) Tucker, Ngo Thi Hoa, John J. Welch, Lucy A. Weinert

Abstract

Antimicrobial resistance (AMR) is a major problem for human and animal health, and can co-vary with features of the bacterial genome, its ecological context, and with other traits. Here, we present an integrated study of AMR in the porcine zoonotic pathogen, *Streptococcus suis*. Collecting a diverse panel of 678 isolates, we obtained replicated measures of Minimum Inhibitory Concentration (MIC) for 16 widely used antimicrobials, from eight distinct classes. Then using whole genome sequences for each isolate, we identified 42 candidate AMR alleles, of which 21 were novel for *S. suis*. The presence of these alleles explains most of the observed variation in MIC but there was also unexplained variation, partly due to alleles of small effect, and epistatic interactions. We find that MIC varies with *S. suis* genetic background and country, but less so with clinical phenotype, serotype or over time. These ecological patterns are largely explained by allele prevalence. MIC values for different antibiotic classes positively correlate with one another, but there is limited evidence of direct pleiotropy, where single alleles confer resistance to multiple drug classes. Instead, this covariation is due to alleles for separate classes appearing together, but not necessarily in the same genomic region. The result is a significant excess of multi-drug resistant isolates. Indeed, we found more isolates that carried resistance alleles to five classes of antibiotics than isolates than carried no resistance alleles. We discuss the implications of our results for combatting AMR in *S. suis*, and for understanding multi-drug resistance in bacteria.

Importance

Streptococcus suis is a bacterium that causes serious disease with significant welfare and economic impacts for the global pig industry, and meningitis in humans. Antibiotics are the principal treatment for these bacteria, and so the evolution of antibiotic resistance presents a major problem for global food security and human health. We tested susceptibility to 16 different antibiotics in a global sample of 678 *S. suis* isolates. Using comparative analysis of whole genome sequences, we identified existing and novel genetic variants that affect resistance. Results show that, in most cases, the presence/absence of these alleles is a good predictor of resistance levels, which could help facilitate resistance monitoring through genome sequencing. We also observe a tendency for AMR variants to co-occur together in some *S. suis* isolates. These multi-drug resistant isolates should be monitored as a priority.

Introduction

The ability of bacterial pathogens to evolve resistance to antimicrobials is one of the gravest threats to human health and food security worldwide. Antimicrobial resistance (AMR) to a given drug can be quantified via the minimum inhibitory concentration (MIC), i.e., the concentration of the drug that is sufficient to inhibit growth of a bacterial culture. While, for practical use, bacterial isolates are often categorised as either susceptible or resistant, MIC is, in principle, a continuously varying trait. Like any other such trait, it may co-vary with aspects of the population ecology, with features of their genomes, and with other phenotypic traits. AMR cannot be fully understood without considering all of these contexts. In particular, the prevalence of multi-drug resistance, implies that resistance cannot be understood on a drug-by-drug basis. In this study, we present a comprehensive portrait of AMR in the bacterium *Streptococcus suis*.

S. suis primarily exists as a commensal in pigs, colonising the nasopharynx, gut and vagina, but it also causes systemic and respiratory disease, particularly in young pigs (1). *S. suis* is also a serious zoonotic disease, being the leading cause of adult bacterial meningitis in Vietnam (2). While some autogenous vaccines are used in pig production, they are serotype-specific and give inconsistent cross protection against heterogenous *S. suis* (3). Antimicrobials therefore remain the standard treatment for *S. suis*, and as such, *S. suis* is a leading driver of antimicrobial usage in pig farms (4).

As well as being a serious problem in itself, *S. suis* also has unique benefits as a model for studying AMR. By weight, more pork is consumed globally than any other meat (5, 6), and *S. suis* is found in most, if not all pigs (7). Furthermore, antibiotic consumption is higher in pigs (172 mg per population corrected unit) than any other livestock (e.g. cattle (45 mg) and chicken (148 mg)) (8). As a result, *S. suis* is extremely diverse, and yet most lineages experience antibiotics. These include not only antibiotics administered directly against *S. suis*, whether as a therapeutic, prophylaxis or metaphylaxis, but also, and perhaps more commonly, in response to many other bacterial infections.

The strong selection pressure caused by widespread use of antimicrobials in pig farming is expected to give rise to AMR in *S. suis*. Consistent with this, several phenotypic studies show high MICs for each of the major classes of antibiotics in one or more *S. suis* collections (9-13). In addition, there have been demonstrations of individual resistance alleles affecting MIC in *S. suis* (14-16), and also some mining of *S. suis* genome collections for known resistance alleles (17-19). However, to our knowledge, no study has combined complete genomic and phenotypic information in large numbers of isolates, collected from a diverse range of populations.

To this end, we obtained replicated measures of MIC for 16 antibiotics, each widely used in pigs, for six different *S. suis* collections, comprising 678 isolates. These collections were chosen to include the three main pig-producing regions of the globe - namely the Americas, South East Asia and

Europe - and targeted both human and pig hosts, as well as a range of years, serotypes and clinical phenotypes (20-22). We obtained whole genome sequences for each of our isolates, allowing us to compare phenotype and genotype on an unprecedented scale.

Results

Measurement of MIC for 16 antibiotics in *S. suis*

We scored 678 isolates of *Streptococcus suis* for replicated measures of Minimum Inhibitory Concentration (MIC) for 16 antibiotics (Table S1). Four of these are beta-lactams (amoxicillin, cefquinome, ceftiofur and penicillin), which are typically used to treat *S. suis* infection in pigs. The other antibiotics are all widely used in the pig industry, but found in different drug classes. They comprise macrolide-lincosamide-streptogramin B (MLS_B: erythromycin, lincomycin, tilmicosin and tylosin), tetracyclines (doxycycline and tetracycline), and fluoroquinolones (enrofloxacin and marbofloxacin), plus one each of an aminoglycoside (spectinomycin), a pleuromutilin (tiamulin), trimethoprim (TMP), and a phenicol (florfenicol). Results, shown in the left hand plot of Figure 1, reveal wide variation among our isolates in MIC for most of the antibiotics. For a few antibiotics, especially the MLS_B and tetracyclines, the MIC values are clearly bimodal, suggesting a meaningful division between "resistant" and "wild-type" isolates. However, for most antibiotics, the distributions are either roughly lognormal (fluoroquinolones and phenicols) or positively skewed (spectinomycin, pleuromutilin, TMP and beta-lactams). It is striking that the distribution for beta-lactams, the most common treatment against *S. suis*, are not clearly outlying. Given these distributions, and the general lack of defined clinical breakpoints for these antibiotics in *S. suis* (23), we treated MIC as a quantitative trait in our subsequent analyses.

Ecological and genomic predictors of MIC

Our *S. suis* isolates were collected at different times, and from different countries, hosts and body sites, including sites associated with respiratory and systemic disease. Isolates were also genetically heterogeneous, and differed in their serotype; including serotypes with a known disease association (see methods and Table S1). We first used linear models, to ask whether MIC levels varied systematically with these ecological and genomic factors.

S. suis is highly recombining, such that no single genealogy characterises its diversity. As such, to model overall genomic diversity in our sample, we identified 30 clusters of genetically-similar strains (see methods). The resulting clusters were a highly significant predictor of MIC ($p<10^{-15}$; Table 1a) showing that genetically similar isolates tend to have similar MIC values. To account for this genetic structure, "cluster" was included as a random effect in subsequent analyses.

A model including all potential predictors showed that country of isolation was also a highly significant predictor of MIC. In addition, there were weaker effects of year of collection (with a slight trend for increasing MIC over time); for serotype (with non-disease-associated serotypes tending to have higher MIC); and for host disease status (with non-clinical isolates having higher MIC) (Table 1b).

Separate analyses for each antibiotic, showed that the effect of country was driven by consistently higher MIC in the samples from Canadian pigs. This

applied to 15/16 antibiotics, and to various models and subsets of the data (Figure S1a; Table 1c). By contrast, the other predictors (year, serotype and clinical status) had effects that were less consistent or robust (Table 1d-f). For example, the trend for higher MIC in non-clinical isolates (Table 1d), applied consistently only in tetracyclines and TMS (Figure S2a).

Identification of candidate AMR alleles within *S. suis* genomes

Using a candidate locus approach, we next identified 42 known and putative AMR alleles that were present in our dataset: nine novel alleles of known resistance genes, 12 known resistance alleles detected in *S. suis* for the first time, and 21 previously reported in *S. suis* (Table 2 and Table S1). Our candidate alleles come from three sources: the antimicrobial resistance database CARD (24), previously published *S. suis* variants not present in the CARD database (25-27) and novel alleles detected using a subset of our data (see methods).

Of the nine novel alleles, five are variants in genes previously associated with resistance in other bacteria (including a promoter variant). In particular, we discovered novel haplotypes within the loop region of the central transpeptidase domains of the *pbp2B* and *pbp2X* genes and within the signal peptide region of the *mraY* gene (encoding the enzyme, acetylmuramoyl-pentapeptide-transferase, essential to cell wall biosynthesis) that were associated with variation in beta-lactam MIC (Table 2). In addition, a variant of the chromosomal dihydrofolate reductase gene *dhfr* (allele frequency of 0.34)

and its promoter (allele frequency of 0.13) were associated with reduced susceptibility against trimethoprim (Table 2). Furthermore, these alleles were independently associated with changes in MIC in at least seven different genetic clusters (Table S2) implying that the association between these haplotypes and MIC is either causal or directly compensatory of the causal mechanism (28).

The remaining four novel alleles, were genes with homologies to known resistance genes, three located on mobile genetic elements (MGEs) and one chromosomal (Table 2). Three novel aminoglycoside genes, carried by isolates with high spectinomycin MIC, were identified based on aminoglycoside resistance determinant protein homologies. Although previously undescribed, we found homologues in other bacteria using a blastn search of the non-redundant database in Genbank (Table 2). We also characterised a chromosomally-encoded *vgaC* homologue (37% protein homology to *vgaC*), encoding an ABC-F ATP-binding cassette ribosomal protection protein, that we designate *vgaF* (allele frequency of 0.25) associated with reduced susceptibility to tiamulin.

The 12 AMR alleles detected in *S. suis* for the first time are mobile genetic element (MGE)-linked genes that confer resistance to aminoglycosides (*n*=2), MLS_B (*n*=2), tetracyclines (*n*=5), TMS (*n*=2) and phenicol (*n*=1*)*. Based on the CARD and NCBI databases, these genes are associated with other gram-positive and gram-negative bacteria (Table 2).

Overall, most of our candidate alleles (28/42) were found in fewer than 5% of the isolates, but *ermB* (which confers resistance to the MLS_B class) and *tetO* (which confers resistance to tetracyclines) were present in a majority of the isolates (Table 2; Figure S3). Only about a quarter of isolates carry a candidate allele for the beta-lactams.

Candidate AMR alleles explain most of the variation in MIC

We next investigated how well our candidate AMR alleles explain the variation we observe in MIC. The right-hand panel of Figure 1 compares the distribution of MIC values for isolates carrying one or more of the candidate alleles for that antibiotic class (denoted A for 'allele'), to the remaining isolates, which carry no such allele (N for 'no allele').

The proportion of the variation in MIC explained by our alleles varied between antibiotics. Explanatory power was strongest (r^2>0.75) for antibiotics where the distributions of MIC were most clearly bimodial (MLS_B and tetracyclines). By contrast, we found no association between our alleles and MIC for the fluoroquinolones and phenicol, where the distribution of MIC values was roughly lognormal, consistent with our isolates representing a "wild-type" population. For the remaining antibiotics, our alleles had intermediate explanatory powers (Figure 1), but there were very few isolates with high MIC that did not carry a candidate allele. This suggests that we have detected most of the causal alleles of large effect in our dataset, removing the need to perform additional genome-wide associations. Where our alleles show low

explanatory power (r^2<0.5), this appears to be caused mainly by isolates that carry an allele but whose MIC is low (Figure 1).

We conjectured that the variable presence of our candidate alleles in different ecological settings would explain the significant ecological predictors of MIC in our linear models (Table 1).

Consistent with this hypothesis, we found that the fitted MIC in genetic clusters was highly correlated with the candidate allele frequency in that cluster (Figure S4), and that candidate allele frequencies are consistently higher in Canada (Figure S3). Accordingly, when we include candidate allele presence in our models, the effect size for country greatly decreases (Figure S1b). Nevertheless, this cannot be the whole picture, because genetic cluster and country remain significant predictors for all 16 antibiotics, even when allele presence is included (Table S3). Including allele presence does however remove the effects of the other weaker ecological predictors. For example, disease status is no longer a significant predictor for 15/16 antibiotics (enrofloxacin remains significant but, as mentioned above, the candidate alleles explain very little variation in this case) (Figure S2b; Table S3).

A possible explanation for the unexplained variation in Figure 1, and the continuing predictive power of genetic cluster and country, is the way in which we have included our candidate alleles as predictor variables (i.e. an isolate is scored as positive if it contains any one allele for that class).

In some cases, simple amendments to the model increase explanatory power. Spectinomycin MIC, for example, is unchanged by some AMR genes from the CARD database (e.g. *ant(6')-Ib* and *aph(3')-IIIa* in Figure S5). Including each of the candidate alleles as a separate predictor (i.e., a two-level factor: present/absent) removes the influence of these non-functional alleles, and improves the r^2 value from 0.37 (Figure 1) to 0.69 (Figure S6a).

A similar improvement is also seen for the beta-lactams (Figure S7), and in this case, it appears that resistance has a polygenic architecture with alleles acting additively on log MIC (Figure S7). This implies that isolates will have high MIC only if they carry multiple alleles. Accordingly, if we predict MIC from the total number of candidate alleles that are carried (treating "number of candidate alleles carried" as a continuous predictor variable), the beta-lactam r^2 values improve from 0.16-0.53 (Figure 1) to 0.35-0.63 (Figure S6b).

These simple explanations (ineffective or small-effect alleles) apply only to some of the antibiotics (Figure S6). In other cases, our data suggest other complexities (Figure S8-S12). Sometimes, as is well known, alleles act against only some of the antibiotics in a class. For example, while modification of the ribosomal target confers cross resistance to macrolides and lincosamides, alleles encoding mechanisms such as efflux and enzymatic inactivation affect only some of the antibiotics within the class (29). This is illustrated in Figure 2, where *mel* acts against erythromycin, but not against lincomycin, while *linB* inactivates lincomycin, but not erythromycin. We also

see that some alleles generally cause high MIC, but are non-functional in some isolates. For example, for isolates carrying *ermB*, the distributions of MIC values are clearly bimodal (Figure 2). Some of this variation appears to be allelic. Examining the 363 isolates with a complete *ermB* sequence in our assemblies, we found that 68/363 were associated with low level resistance (MICs ≤32mg/ml). Of the 363, 304 isolates carried one or more rare amino acid variants, and these isolates had a lower mean MIC for erythromycin (Welch's t test t = 4.1231, df = 148.35, $p<10^{-4}$). Nevertheless, some isolates carrying the wild-type *ermB* also had low MIC. This is likely an epistatic effect between *ermB* and its genetic background. One possible interaction of this kind involves another candidate allele: *lnuB*, which alone has no effect on erythromycin, but decreases MIC in the presence of *ermB* (Figure 2). The non-linear effects of genotype, evident in Figure 2, are the most probable explanation of the remaining unexplained variation in MIC in our data.

Covariation in MIC between antibiotic classes

Results above consider each antibiotic independently, but our data also show strong positive covariation in MIC between different drug classes (Figure 3). Of 120 pairwise correlations, 112 are positive, and all 8 exceptions involve enrofloxacin, where there is little variation in MIC (Figure 1a; Figure 3 inset). Covariances are particularly strong for beta-lactams/pleuromutilin/TMP/aminoglycoside, along with MLS_B/tetracyclines and MLS_B/phenicol (Figure 3).

The simplest hypothesis to explain this pattern would be that some alleles act against more than one antibiotic class (i.e. direct pleiotropy). To test for pleiotropy in our data, we asked whether isolates with and without candidate alleles for a different antibiotic class, tended to differ in their MICs (Figure S13). The strongest support for the hypothesis of direct pleiotropy was found in the tetracyclines. Here, isolates containing candidate alleles for non-tetracycline drugs showed higher MIC for tetracycline, whether or not they also carried tetracycline-resistance alleles (Figure S13). The same result was observed in doxycycline, but only for isolates carrying tetracycline-resistance alleles. Overall, however, these were isolated results, with modest effect sizes. No other class showed consistent evidence of pleiotropy (Figure S13), and there was little evidence of pleiotropic effects for the most common single alleles (*ermB*, *tetO*, and *dhfr102*; see Figure S14).

If direct pleiotropy cannot explain the observed covariation in MIC, an alternative explanation is that alleles of different classes tend to occur together in isolates. Consistent with this, we found that the co-occurrence between candidate alleles from different antibiotic classes strongly predicts the correlation in their MIC values (upper right graph in Figure 3, Mantel test $p<10^{-4}$).

An obvious explanation for the co-occurrence of our alleles is their physical linkage, and consistent with this, many of our candidate alleles occur together on mobile genetic elements. One example is the integrative conjugative element (ICE) ICESsu584 (GenBank: KT716758), which carries two

aminoglycoside resistance genes *ant6-1a* and *ant9-1a*, a lincosamide resistance gene *lnuB* and a tiamulin resistance gene *lsaE*. This may explain the observed covariation in MIC between spectinomycin, tiamulin and lincosamide (“spc”, "tia" and "lcm" in Figure 3). In addition, the two most common alleles, *ermB* and *tetO* often co-occur on ICEs (19) and this may explain the co-variation between tetracycline and MLS_B.

We also see physical linkage of alleles within the chromosome. The *vgaF* homologue is often located upstream of the *pbp2B* gene (Figure S15), which may explain the covariation between beta-lactams and pleuromutilin (Figure 3).

Furthermore, there is a general and significant trend in the data for co-occurring alleles to be closer together in the genome (see Figure S15; Mantel test $p<10^{-4}$). Nevertheless, the strength of the correlation is weak, and linkage clearly cannot explain all of the co-occurrence. Some highly correlated alleles are very distant in our genomes. For example, the *vgaF* homologue co-occurs with the *dhfr102* variant (ρ=0.51), yet when they assemble on the same contig, they are located a median of 367,404bp apart (Figure S15).

AMR allele clustering

A corollary of allelic co-occurrence is a high level of clustering in the number of candidate AMR alleles carried. Our data contained a class of “globally susceptible” isolates, which carry no candidate allele (60/678), and out of the many isolates (525/678) that carry multiple candidate alleles (Figure 4a),

almost all of these (508/525) were "multi-drug resistant" (MDR), carrying candidate alleles for two or more antibiotic classes. Indeed, more isolates carry alleles for five distinct classes of antibiotic than carry no candidate allele at all (Figure 4b).

Without a well-defined null model, it is not clear whether this co-occurrence calls for any special explanation, (e.g. this level of MDR may simply reflect the high number of individual AMR alleles in our *S. suis* samples). To generate null expectations, we used three randomization procedures. First, we randomly permuted allelic states across the isolates. Results (Table 3; "over isolates") show that, compared to the real data, the randomized data sets had increased levels of resistance overall, with significantly fewer susceptible isolates, but significantly more MDR isolates (an illustrative randomisation of this kind is shown as the narrow red bars in Figure 4).

Randomisation across isolates implicitly assumes that all alleles evolve independently, so we next made the opposite assumption, assuming clonal evolution, and placing each of our resistance alleles, at random, on the consensus phylogeny of our isolates. As shown in Table 3 ("across phylogeny") this type of randomisation had the opposite effect to randomisation across isolates, leading to significantly more susceptible isolates, but significantly less MDR (an illustrative randomisation is shown as the narrow yellow bars in Figure 4).

Because these procedures make extreme assumptions of different kinds, we next combined them, applying each to a subset of our candidate alleles. For the 31/42 alleles that were found on MGEs (see Table 2), we randomised across isolates, while the remaining 11 alleles were placed at random on the phylogeny. These randomised data sets (Table 3; "both") were much closer to the true data. However, they were much less clustered, with significantly fewer susceptible isolates, and significant fewer MDR isolates than the true data. Together, then, we can conclude that the clustering of AMR in our data is greater than would be expected by chance.

We also examined our data for a "nested" pattern of resistance in which lower frequency AMR alleles tend to appear in genetic backgrounds that already contain an AMR allele (see (30) for details). In total 63% of our isolates were nested in this way (Figure S16), but this value was not clearly higher than expectations under the null model of evolution over the genealogy (Table 3, bottom row).

Discussion

We have examined antimicrobial susceptibility to 16 commonly-used antibiotics in *Streptococcus suis*, using replicated measures of MIC. Data suggest that, even when some isolates have very high MIC, there is often no natural separation into "resistant" and "susceptible". This highlights the need to define clinical breakpoints for these antibiotics in *S. suis* to understand the relevance of these MIC measurements in treatment.

We have described 42 candidate resistance alleles, and showed that they explain the majority of the observed variation in MIC for 11/16 antibiotics, and a substantial fraction of the variation for another 2/16 (Figure 1; Figure S6).

While many of these alleles were previously described, and most were found at low frequency (Table 2; Figure S3), we detected three novel candidates that appear in many isolates with clinically-relevant changes in MIC: *vgaF*, *dhfr102* and the *dhfr* promoter. If these resistance mechanisms are functionally verified, they should be included in routine AMR gene testing for *S. suis*.

Of particular interest are the novel haplotypes that are linked to changes in beta-lactam MIC, because penicillin is the first line treatment for *S. suis* infection. Functional studies will be needed to show whether the *S. suis* PBP alleles uncovered here confer a similar mode of resistance to beta-lactams as is seen in other streptococci (31-35). Studies have shown that cefotaxime, a third-generation cephalosporin like ceftiofur, selectively inactivates PBP2X but

not PBP2B (36) and that mutations within the catalytic domain in PBP2X give rise to cefotaxime, but not to penicillin, resistance (32). Our results echo this pattern because our PBP2B haplotype has a smaller change on mean MIC of ceftiofur than penicillin (Figure S7), and the explanatory effects of our PBP haplotypes are different for the different beta-lactams (Figure 1). Previous variants in PBP2X thought to confer cefuroxime resistance (25), whilst validated by our data, were found at very low frequency in our collection (Table 2). We show that our PBP haplotypes have additive effects on mean log MIC such that some of our strains reach clinically relevant changes in beta-lactam MIC (e.g. MIC > 1 mg/L for Penicillin) but that no one allele is associated with this change. Additive effects have also previously been shown for *tet* alleles in *S. suis* (16). Our work illustrates a need to address how these additive effects can be incorporated in to predictive genome models of MIC.

Unlike the PBPs, the contribution of mutations in *mraY* to changes in beta-lactam MIC might not be directly causal. It has been suggested that altered PBPs, while conferring resistance, might be less active transpeptidases than their wild type counterparts, and that organisms are likely to evolve compensatory mechanisms to circumvent this problem (31). For example, the amidotransferase enzyme encoded by *murT* enables cross-linking of cell wall peptidoglycans by preparing lipid II used by pneumococcal PBPs (37, 38). Mutations in *mraY* are therefore suggestive of a possible compensatory mechanism for reduced enzymatic activity of altered PBPs in *S. suis.* In agreement with this, Figure S7 shows that the *mraY* haplotypes are only found in concert with an altered PBP. Polymorphisms in *mraY* were also

identified in a genome-wide association study of beta-lactam resistance in *S. pneumoniae* (39) and could make a promising drug target for beta-lactam resistant streptococci.

While our candidate alleles were explanatory, we also found several complexities in the genotype-phenotype map. These include epistatic interactions, where the effects of some candidate genes vary with their *S. suis* genetic background. Phenotypic reversion (i.e. mutations in transcriptional regulators or elsewhere in the genome, counteracting a resistance allele) can be common if resistance has a fitness cost (40), and this could explain why we see many isolates carrying candidate alleles but with low MIC (Figure 1). These complexities have consequences for diagnostic investigation of AMR of *S. suis* using whole genome sequences. While the predictive power is quite high for many antibiotics, we should expect many Type I errors. These results also caution against using AMR allele presence as a measure of phenotypic resistance more generally.

We have also shown suggestive links between antibiotic exposure and non-susceptibility. In particular, antibiotic usage is much higher in Canada than the UK (41, 42) and this is reflected in consistently higher MIC, and consistently higher frequency of the same candidate AMR alleles (Figures S1, S3). Results for Vietnam are more difficult to interpret, partly due to a lack of official records of antibiotic usage (although see (43)), and partly because our Vietnamese sample is relatively small and genomically homogeneous (see methods).

When considering results for all of the 16 antibiotics together, we have shown that our data contain a significant excess of MDR. This is evident both in highly correlated MIC values and, indirectly, via the co-occurrence of alleles. Notably, we have found that isolates with candidate resistance alleles to five different drug classes are more common than isolates with no resistance alleles at all (Figure 4b).

While linkage between resistance alleles plays some role in explaining this pattern (Figure S15), we argue that it cannot do so alone. One possible factor is resistance to one antibiotic somehow facilitates the evolution of resistance to a second. We found limited evidence of direct pleiotropy (where AMR alleles to a previously administered drug provide some action against a novel drug; Figure S13-S14), and so this might plausibly involve indirect pleiotropy, (e.g., mitigation of a cost to antibiotic resistance that applies to multiple drug classes).

However, a simpler explanation for the observed pattern in MDR is correlated selection pressures, where isolates are exposed to many antibiotics, or to none (30). Some forms of correlated selection lead to a characteristic pattern of nested MDR, where rarer alleles tend to be found on backgrounds containing more common alleles (30). But while we found some nestedness in our data, it was not clearly above chance levels (Figure S16; Table 3). Correlated selection pressures might also result if isolates have very different durations of carriage (44). It is notable, in this respect, that we have found

higher levels of non-susceptibility in isolates from the tonsils of pigs without clinical disease (Table 1; Figure S2).

Two final aspects of our results highlight the fact that AMR is not a problem that can be tackled one disease at a time. First, we have found high levels of resistance in *S. suis* even to antibiotics other than beta-lactams that are typically used to treat this infection. In fact, resistance alleles to four of the seven other classes were more common (Figure S16). This must reflect a "bystander selection effect", common in opportunistic pathogens that are part of the healthy microbiota and frequently exposed to antibiotics used as growth promoters or therapeutics. Second, many of the rarer AMR alleles in our sample were not previously known in *S. suis*, but are common in other bacteria, both gram negative and gram positive. This raises the possibility of *S. suis* acting as a reservoir for AMR alleles in other bacteria. Indeed *S. suis* shares conserved chromosomal insertion sites with many human pathogens, such as *S. pyogenes, S. pneumoniae*, and *S. agalactiae* (17). Given that *S. suis* causes human clinical disease, a direct exchange of AMR alleles between *S. suis* and other human pathogens is plausible.

Material and Methods

Streptococcus suis isolates

We assembled a collection of 678 *S. suis* isolates, including isolates from the UK (*n*=423), Canada (*n*=205) and Vietnam (*n*=50) (Table S1).

In the UK, isolates came from three different collections. The first collection in 2009-2011 sampled non-clinical and clinical isolates from pigs across England and Wales (described in Weinert *et al.* 2015 (20)). The second collection in 2013-2014 sampled non-clinical isolates from five farms (described in Zou *et al.* 2018 (22)). The third collection during 2013–2014 targeted clinical isolates from pigs across England and Wales (described in Wileman *et al.* 2019 (21)). In pigs that showed clinical symptoms consistent with *S. suis* infections (e.g. meningitis, septicaemia and arthritis), the site of isolation was classified as 'systemic' if recovered from systemic sites. The site of recovery was classified as 'respiratory' if derived from lungs with gross lesions of pneumonia. *S. suis* isolates from the tonsils or tracheo-bronchus of healthy pigs or dead pigs without any typical signs of *S. suis* infections were defied as 'non-clinical'. Isolates that could not confidently be assigned to these categories (e.g. a tonsil isolate from a pig with systemic signs) were classified as unknown. Altogether, the UK isolates were classified by clinical status as 'systemic' (*n*=94), 'respiratory' (*n*=50), 'non-clinical' (*n*=197) or 'unknown' (*n*=82), respectively.

The Canadian pig *S. suis* isolates from 1983-2016 were collected to target similar numbers of clinical and non-clinical isolates and were also classified by clinical status as 'systemic' (*n*=81), 'respiratory' (*n*=30), 'non-clinical' (*n*=55) or 'unknown' (*n*=39).

The Vietnamese isolates were collected to sample related populations from human and pig (described in Weinert *et al.* 2015 (20)). These comprised 'systemic' isolates (*n*=28) from human clinical cases of meningitis from provinces in southern and central Vietnam, and 'systemic' (*n*=4) or 'non-clinical' isolates (*n*=18) from pigs, collected between 2000 and 2010. These isolates were exclusively serotype 2 or serotype 14 and belong to one genetic population (Table S1).

Antimicrobial susceptibility testing

The minimum inhibitory concentrations (MIC) for a range of antibiotics were determined by the micro-broth dilution method, which was performed and the results interpreted in accordance with CLSI Approved Standards, M100-S25 (2015), Vet01S 3rd Edition (2015) and de Jong *et al.* (2014) (9, 45). MIC measurements for some of our isolates were previously published (9). For the remaining MIC measurements, the MICs were determined for sixteen different antimicrobial compounds, representing nine antimicrobial classes at LGC, Fordham, UK (formerly Quotient Bioresearch, Fordham, UK) for the UK and the Canadian isolates. MIC testing of antibiotics for the Vietnamese isolates was performed at the Oxford University Clinical Research Unit, Ho Chi Minh

City, Vietnam, in collaboration with the Department of Veterinary Medicine, University of Cambridge, UK.

Whole genome sequencing, assembly and inference of population structure

Genome DNA extractions and whole genome sequencing of newly sequenced *S suis* in this study were as previously described by Weinert *et al.* (2015) (20). Briefly, single colonies of strains were grown up in broth culture, DNA was extracted using DNeasy kits (Qiagen), Illumina library preparations were performed as described by Quail *et al.* 2008 (46) and the whole genomes sequenced on the HiSeq2000 according to the manufacturer's instructions (Illumina, San Diego, CA, USA) at the Welcome Trust Sanger Institute, Cambridge, UK. Sequencing generated 125bp paired end reads, which were assessed for quality using Sickle (47) after the removal of adapter sequences.

Sequencing reads that passed the quality threshold were put forward for *de-novo* assembly generation using Spades v.3.10.1 (48) utilising a variety of parameters conditions. Our 678 isolates were combined with 401 additional genomes given in Table S4 to increase robustness of *S. suis* population structure estimation, although these isolates were unavailable for MIC testing. Draft genomes were annotated using Prokka v1.12-beta (v2.8.2). (49) and bacterial species assignment was performed by a combination of MLST assignment and FastQ Screen v.0.11.1 using a custom database (50). We mapped the Illumina reads back to the *de-novo* assembly to investigate polymorphic reads in the samples (indicative of mixed cultures) using BWA

v.0.7.16a (51). Genomes that exhibited poor sequencing quality (i.e poor assembly as indicated by a large number of contigs, low N50 values or a high number of polymorphic reads) or that which were inconsistent with an *S. suis* species assignment were excluded from the analysis.

In order to group our *S. suis* isolates in to different genetic clusters, we used the program hierBAPs in R (52). First, we inferred core genes from our isolates using Roary (v2.8.2) (53), aligned them using MACSE (54), and stripped regions that could not be aligned unambiguously due to high divergence, indels or missing data. This conserved region of the core genome was used as input in hierBAPS. We also constructed a neighbour joining tree of the core genome by importing the fasta alignment in to R v.3.4.1 (55). We computed a distance matrix using the function *dist.dna* with Kimura's 2-parameter distance model of sequence evolution ("K80") (56) and created a neighbour-joining tree using the function *nj*, both from the R package *ape* (57, 58). We midpoint rooted the tree using the *midpoint* function in the R package *Phangorn* (59).

Known AMR allele detection

ARIBA (60) identifies AMR determinants (or any sequence of interest) directly from paired sequencing reads using a public or custom reference database and relies on mapping the reads to reference sequence clusters followed by the formation of local assemblies of the mapped reads. The tool is further able to confirm the intactness of resistance genes as well to identify known or novel SNPs within a gene of interest.

The AMR alleles in *S. suis* were identified using ARIBA v2.10.0 based on the public AMR database, CARD (24), which was further supplemented with a custom database (given in the Supplement). The custom database contained gene sequences not present within the CARD database at the time of testing. These were previously published *S. suis* AMR genes, single nucleotide polymorphisms (SNPs) in genes known to confer antibiotic resistance and AMR genes found in other bacteria. In addition, we investigated whether there were additional novel resistance alleles in known resistance genes (described below). The sequence identity threshold against a reference was set at 90%. Only paired-end sequence reads that passed the quality control thresholds were used as input for ARIBA. To identify genes falling under the 90% identity threshold, we also performed blast searches of the draft assemblies against the non-redundant NCBI protein or nucleotide databases to identify variants or chimeric alleles of known resistance determinants that might not be present in the CARD database. Examples of allelic variants identified this way include *tet*, *aad*, *ant* and *cat* (Table 2).

Novel AMR candidate allele detection

We identified novel resistance alleles by a range of methods. Candidate alleles that might encode novel resistance mechanisms were identified by scanning literature describing experimental studies and others found by Genome Wide Association Studies (GWAS) in related bacterial species, for example, alleles of *pbp*s, *mraY*, *dhfr* and *folP* (33, 61-63) along with their promoter regions. To avoid over-fitting our generalised linear models, gene or

promoters of interest in a “training” subset of the collection (*n*=205) were then extracted, aligned and ranked from the highest to lowest MIC values, using MUSCLE (64) in SEAVIEW (65). Manual sequence analysis was then performed to identify either amino acid or nucleotide variations that associated with high MIC.

Kinetic and structural studies have previously established that beta-lactam resistance is conferred by substitutions within PBPs in *Streptococcus pneumoniae* (33, 66) and other streptococci (34, 35, 67). While mutations are present throughout the entire PBPs, we noted statistically significant mutations (Table S2) in altered PBP2B and PBP2X expressing alleles in strains exhibiting high penicillin (≥ 1mg/L) and ceftiofur MICs (≥ 2mg/L), respectively. In PBP2B, the residue variations were present in loop regions, K479T/A, D512E/Q/K/A, K513E/D and T515S, within the *S. suis* transpeptidase domain (residues 351-681). In PBP2X, the mutations were located at positions M437L, S445T, T467S and Y525F, also in loop regions within its catalytic domain (residues 265–619). As multiple amino acid variants were identified within the *pbp2X* and *pbp2B*, many of which were in the same genomic region (Table 2), we grouped the variants in to haplotypes. We did this by scoring haplotype presence if the isolate had all of the amino acid variants associated with high MIC in a given gene.

In addition to allelic differences in PBP2B and PBP2X, we also detected mutations in *mraY*. Found immediately upstream of *pbp2X*, *mraY* encodes an acetylmuramoyl-pentapeptide-transferase enzyme which is essential to lipid

cycle reactions in the peptidoglycan cell wall biosynthesis pathway. Residue substitutions were present at either A4S/T together with M6I/L or M6I/L together with G8S within the signal peptide regions of the protein and were associated with variation in both penicillin and ceftiofur MIC. These were scored as two separate haplotypes in further analyses. The distribution of these mutations was statistically significant at ≥ 1mg/L for both ceftiofur and penicillin (Table S2).

Not all isolates with high tiamulin MICs possessed a *lsaE* gene, suggesting an additional mechanism in *S. suis* conferring resistance to pleuromutilins. Using the CARD RGI (resistance gene identifier) protein homology models, a *vgaC* homologue was identified as a candidate gene associated with variation in tiamulin MIC. The gene, designated as *vgaF* in this work, encodes an ABC-F ATP-binding cassette ribosomal protection protein and shares 37.14% homology to the reference. Located within the chromosome, *vgaF* is found intact (~1386bp) in statistically significant numbers in isolates with MIC ≥ 8mg/L (Table S2) but is truncated (<750bp) in others. Plasmid borne *vga* homologues including *vgaC* are frequently detected in staphylococci and have been shown to confer cross resistance to pleuromutilins, lincosamides and streptogramin A antibiotics. A chromosome based *vga*(A) gene variant, encoding an ATP-binding cassette protein conferring resistance to streptogramin A and related antibiotics in *S. aureus* has also been described (68).

An amino acid substitution (I102-L) was detected in the dihydrofolate reductase (DHFR) gene, *dhfr*, in the majority of isolates with trimethoprim MICs of 0.12mg/L or greater. Similar substitutions of isoleucine to leucine at position 100 in *S. pneumoniae* (62, 69) and *Streptococcus pyogenes* (70) is known to cause resistance to TMP. Additionally, we also identified polymorphisms within the promoter region (0-30bp upstream) of the *dhfr* gene; an A5G substitution and insertions, in isolates exhibiting MICs ≥ 1mg/L.

All of these alleles might be causal, compensatory or linked to high MIC because of population structure. We therefore examined our alleles and showed that susceptible and resistance alleles were found repeatedly in different BAPs clusters, which were significant using binomial sign tests when defining a cutoff (Table S2).

The *pbp*, *mraY* and *vgaF* genes were unable to be included into the custom database in ARIBA because of the sequence divergence. Therefore, we manually aligned the genes in all 678 isolates using MUSCLE (64) in SEAVIEW (65).

Similarly, analysis for known variants in *gyrA* and *parC* were also performed by manually aligning the alleles. *tet(O/W/32/O)* was scored by using blastn as ARIBA was unable to differentiate mosaic sequence patterns

Serotype inference from whole genomes

Serotypes were determined *in silico* using the Athey *et al.* (2016) (71) serotype database, which we implemented in ARIBA. Failed ARIBA runs and sequence non-matches were designated as not available (NA).

Statistical analyses

All general linear models were fit in R v. 3.3 (55) using the built-in function *lm* for models with solely fixed effects, or via Reduced Maximum Likelihood, using the function *lme* from the package nlme v. 3.1-141 (72), when BAPs cluster was included as a random effect. The response MIC values were log transformed. Our data for host and country is confounded because all of our human samples came from Vietnam (human *S. suis* has a higher prevalence in South East Asia (73)). Therefore, we coded “Country” as separate populations – Canada, UK, Vietnam-pig and Vietnam-human. We classified “serotypes” into either disease-associated or non-disease associated according to Wileman *et al.* (2019) (21).

For the Mantel tests using the data plotted in Figure 3 and Figure S15, the test statistic was the Pearson correlation coefficients of the off-diagonal matrix elements (excluding the diagonals for correlation matrices, and, with missing data, including all observations that were pairwise complete), and generated a null distribution from 10,000 random permutations of the row/column ordering for one of the matrices. Given the hypotheses, these tests were one-tailed.

Inferring physical linkage of candidate alleles

The coordinates of candidate alleles in *S. suis* draft genomes were obtained using a custom python script (given in the Supplement). This script identifies the allele locations by performing blastn of the alleles against local databases of each draft genome. For each allele pairing, the number of genomes that contained both alleles were computed and whether these alleles were found on the same contig. For the isolates where the alleles were found on the same contig, a distribution of base pair distances was computed. The median base pairwise distance between two alleles was used to generate the heatmap in Figure S15.

Description of randomisation tests

For the permutation tests shown in Table 3, and illustrated in Figure 4, we used two approaches. For permutation "over isolates", we randomly permuted the present/absent records for each allele, independently, over the strains. This preserves the overall frequency of each allele, and the mean number of alleles per strain, but effectively assumes free recombination among all sites. We then assigned each of the candidate alleles, at random, to one of the 679 branches of our core genome phylogeny (described above), choosing each branch with a probability proportional to its molecular branch length (thereby accounting for the fact that the substitutions are more likely to occur along lineages that have undergone more molecular evolution). The resistance allele was then treated as present in all strains that descend from the randomly chose branch. For the combined approach (labelled "both"), permutation over isolates was applied to candidate alleles found on Mobile Genetic Elements (Table 2), and permutation over phylogeny applied to the

remaining alleles. To assign *p*-values, we report the proportion of the randomly permuted data sets whose value of the test statistic was more extreme (i.e., closer to the tails of the distribution) than the value for the true data, doubled for a two-tailed test.

Data availability

Whole genome sequence assemblies are available from NCBI under Project ID PRJNA628943. The gene *cat* from isolate FX419 has accession MT367165. The gene *aadE1* from isolate CF2D3-1A has accession MT383663. The gene *aadE2* from isolate BH3D7-4E has accession MT383664. The gene *ant1* from isolate CF2D3-1C has accession MT383665. The gene *tetM1* from isolate 1129711 has accession MT383666. The gene *tetM2* from isolate LSS85 has accession MT383667. The gene *tetM3* from isolate SS981 has accession MT383668. The gene *tetO1* from isolate D16-010262 has accession MT383669. All R and Python scripts for conducting statistical tests and figures are given in the supplement. The ARIBA custom database is given in the supplement.

Acknowledgements

This work was primarily supported by a Challenge grant from the Royal Society (reference: CH160114) and an Isaac Newton Trust Research Grant (reference: 17.24(u)). In addition, LAW and GGRM are supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (109385/Z/15/Z). We also acknowledge a BBSRC grant to AWT (BBSRC grant BB/L003902/1), a LoLa grant to DJM and AWT (BB/G019274/1) and a China MoST grant to RZ (2013DFG32360) for sample collections. In addition, we wish to thank Abiyad Baig, Andrew Balmer and Olivier Restif for helpful discussion on antimicrobial resistance in *S. suis*.

Contributions

LAW and NTH conceived the study. AWT, SMW, RZ, MG, NF, TMW, JP, DJM contributed isolates. NFH, ELM, PLKY, HDP, NTH collected new data. NFH, ELM, GGRM, JJW and LAW analysed data. JJW produced all figures. LAW, NFH, ELM and JJW wrote the manuscript.

Table 1. Ecological and genomic predictors of MIC

Analysis	#strains	Fixed effects	*Df*	*F*	*p*
(a)	678	Antibiotic	15	1404.125	$<10^{-15}$
		genetic cluster	29	40.644	$<10^{-15}$
(b)	450	Antibiotic	15	972.3107	$<10^{-4}$
		Year	1	6.6078	0.0102
		Serotype	1	6.6261	0.0101
		Disease Status	2	2.9760	0.0511
		Country	3	92.5155	$<10^{-4}$
(c)	678	Antibiotic	15	1457.4043	$<10^{-4}$
		Country	3	139.5902	$<10^{-4}$
(d)	557	Antibiotic	15	1144.3846	$<10^{-4}$
		Disease Status	2	9.9619	$<10^{-4}$
(e)	542	Antibiotic	15	1131.6347	$<10^{-4}$
		Serotype	1	8.7216	0.0032
(f)	652	Antibiotic	15	1350.8781	$<10^{-4}$
		Year	1	0.1591	0.69

Table 2. AMR determinants of *Streptococcus suis*

AMR gene	Name in figures	Resistance conferred (Antibiotic Class)	No. of isolates	Reported in *S. suis*	Reported in other species (CARD and NCBI)	Reported genetic location in *S. suis* or other species
ant(4')-Ib	ant4Ib	Aminoglycoside	1	This report	*Staphylococcus aureus, Staphylococcus epidermidis, Enterococcus spp.*	Plasmid
ant(6')-Ia/aadE	ant6Ia	Aminoglycoside	56	(19)	*Staphylococcus epidermidis, Streptococcus spp., Enterococcus fecalis*	Plasmid, ICE
ant(6')-Ib	ant6Ib	Aminoglycoside	12	This report	*Campylobacter spp., Bacillus subtilis, Clostridium spp.*	Pathogenicity Island
ant(9')-Ia/spw/aad9	ant9Ia	Aminoglycoside	45	(19)	*Staphylococcus aureus, Enterocoocus spp., Staphylococcus spp.*	Plasmid, ICE
aph(3')-IIIa	aph3IIIa	Aminoglycoside	46	(74)	*Campylobacter spp., Staphylococcus spp., Streptococcus spp., Enterococcus spp.*	Plasmid, transposon
*ant1**	ant1	Aminoglycoside	12	This report	*Campylobacter spp., Clostridium difficile, Lactobacillus amylovorus*	Plasmid, ICE
*aadE1**	aadE1	Aminoglycoside	2	This report	*Enterococcus faecium, Staphylococcus aureus*	Plasmid
*aadE2**	aadE2	Aminoglycoside	5	This report	*Clostridium spp., Campylobacter fetus*	ICE
ermB	ermB	MLS_B (Macrolide)	370	(75)	*Campylobacter spp., Clostridium spp., Enterococcus spp., Escherichia coli, Klebsiella pneumoniae, Proteus mirabilis, Shigella spp., Staphylococcus spp., Streptococcus spp., Vibrio vulnificus.*	Plasmid, ICE
ermG	ermG	MLS_B (Macrolide)	1	This report	*Clostridium difficile, Enterococcus faecium*	Conjugative transposon
ermT	ermT	MLS_B (Macrolide)	3	This report	*Enterococcus faecium, Staphylococcus aureus, Streptococcus spp.*	Plasmid
mefA	mefA	MLS_B (Macrolide)	11	(75)	*Streptococcus spp., Enterococcus faecium, Haemophilus parainfluenzae, Proteus mirabilis*	Transposon, ICE
mel/msrD	mel	MLS_B (Macrolide)	9	(75)	*Streptococcus spp., Enterococcus faecium, Haemophilus parainfluenzae, Proteus mirabilis.*	Transposon, ICE
lnuB	lnuB	MLS_B (Lincosamide)	3	(17)	*Enterococcus spp., Staphylococcus spp., Streptococcus agalactiae*	Plasmid
linB	linB	MLS_B (Lincosamide)	42	(76)	*Streptococcus suis*	
lnuC	lnuC	MLS_B (Lincosamide)	9	(77)	*Campylobacter spp., Clostridium difficile, Streptococcus agalactiae.*	Plasmid
lsaE	lsaE	Pleuromutilin	43	(17)	*Enterococcus spp., Staphylococcus spp., Streptococcus agalactiae.*	Plasmid, ICE
tet44	tet44	Tetracycline	4	This report	*Clostridium spp.*	ICE
tetM	tetM	Tetracycline	81	(19)	*Streptococcus spp., Escherichia coli, Staphylococcus spp., Enterococcus spp., Klebsiella pneumonia, Neisseria meningitides, Clostridium difficile, Salmonella enterica*	Plasmid, ICE
*tetM1**	tetM1	Tetracycline	34	This report	*Staphylococcus spp., Streptococcus spp., Enterococcus spp., Bacillus spp., Clostridium spp.*	Plasmid, transposon
*tetM2**	tetM2	Tetracycline	43	This report	*Enterococcus spp., Clostridium difficile.*	
*tetM3**	tetM3	Tetracycline	13	This report	*Streptococcus pneumoniae, Streptococcus pyogenes, Enterococcus spp., Clostridium spp.*	Transposon, ICE
tetO	tetO	Tetracycline	465	(19)	*Streptococcus suis, Campylobacter spp.*	ICE, plasmid
*tetO1**	tetO1	Tetracycline	2	This report	*Dorea spp.*, uncultured bacterium	Conjugative transposon
tetW	tetW	Tetracycline	8	(16)	*Streptococcus agalactiae, Clostridium difficile, Chlamydia trachomatis.*	Plasmid

tet40	tet40	Tetracycline	7	(17, 26)	*Clostridium difficile*	Composite mobile genetic element, ICE
tetL	tetL	Tetracycline	12	(19)	*Streptococcus spp., Escherichia coli, Staphylococcus spp., Enterococcus spp., Klebsiella pneumoniae.*	Plasmid
tet(O/W/32/O) †	tetO:W:32:O	Tetracycline	20	(73)	*Streptococcus suis*	ICE
dfrF	dfrF	TMP	11	This report	*Clostridium difficile, Enterococcus spp., Staphylococcus aureus, Streptococcus agalactiae*	plasmid
dfrK	dfrK	TMP	7	This report	*Staphylococcus aureus, Staphylococcus epidermidis.*	plasmid
*vgaF**	vgaF	Pleuromutilin	173	This report	Not known	Chromosome. Found as an intact gene (~1386bp) or truncated (<750bp)
*cat**	cat	Florfenicol	3	This report	*Enterococcus faecalis, Streptococcus iniae, Listeria monocytogenes, Staphylococcus aureus*	Plasmid, Transposon
*dhfr**	DHFR102	TMS	235	This report	Mutations in I100L in *Streptococcus spp.* and *Staphylococcus aureus*	*dhfr* : I102L
*dhfr promoter**	DHFRPromoter	TMS	86	This report	None reported	*dhfr* : A5G upstream/indels 0-30bp upstream
*pbp2B/penA**	PBP2B:hap	Penicillin	181	This report	Mutations in other locations in *Streptococcus pneumonia* and Group B *Streptococcus spp*	*pbp2B* : K479T/A, D512E/Q/K/A, K513E/D,T515S
*pbp2X**	PBP2X:hap	Ceftiofur	93	This report	None reported	*pbp2X* : M437L, S445T, T467S Y525F
pbp2X	PBP2X594:596	Penicillin	15	(25)	*Streptococcus suis.* Mutations in other locations in *Streptococcus pneumonia,* Group B *Streptococcus sp.* and *Streptococcus uberis*	*pbp2X* : L594Y/F, V596G
pbp2X	PBP2X569	Penicillin	10	(25)	*Streptococcus suis*	*pbp2X* : N569Q
*mraY**	MraY:hap	Penicillin	62	This report	None reported	*mraY* : A4S/T, M6I/L G8S
parC†	ParC79	Fluoroquinolone	4	(27)	*Escherichia coli, Pseudomonas aeruginosa, Streptococcus spp., Staphylococcus aureus*	*parC* : S79
gyrA†	GyrA81	Fluoroquinolone	13	(27)	*Escherichia coli, Klebsiella spp., Streptococcus spp., Staphylococcus aureus*	*gyrA* : S81
gyrA†	GyrA85	Fluoroquinolone	1	(27)	*Escherichia coli, Klebsiella spp., Streptococcus spp., Staphylococcus aureus*	*gyrA* : E85

*novel determinants reported in this study
† Not documented in the CARD database
MLS_B Macrolide Linocsomide Streptogramin B
TMP Trimethoprim

		over isolates		over genealogy		both	
statistic	value	null: mean (range)	*p*	null: mean (range)	*p*	null: mean (range)	*p*
proportion susceptible	0.088	0.014 (0.001-0.032)	$<10^{-4}$	0.621 (0.001-0.855)	0.033	0.054 (0.001-0.086)	$<10^{-4}$
proportion MDR	0.748	0.881 (0.842-0.928)	$<10^{-4}$	0.109 (0.006-0.900)	8×10^{-4}	0.667 (0.609-0.928)	0.020
proportion nested	0.631	0.272 (0.226-0.326)	$<10^{-4}$	0.753 (0.487-0.923)	0.065	0.708 (0.655-0.860)	$<10^{-4}$

Table 3: Evidence of clustering of AMR in *S. suis*

Figure legends

Figure 1. Candidate AMR alleles explain most of the variation in MIC. a) Histograms of log transformed MIC measures for each of our 16 different antibiotics, across our panel of 678 *S. suis* isolates. Antibiotics are coloured by their class (Beta: beta-lactams; MLS_B: macrolide-lincosamide-streptogramin B; Tetra: tetracyclines; Fluoro: fluoroquinolones; Amino: aminoglycoside; Pleuro: pleuromutilin; TMS: trimethoprim/sulfamethoxazole and Phen: phenicol). **b)** Histograms showing the predictive power of our candidate AMR alleles. The left-hand histograms (labelled N) show the MIC values for isolates that carry no allele for that antibiotic class, while the right-hand histograms (labelled A) show the MIC values for isolates carrying one or more such allele. If all resistance alleles perfectly explain MIC, then we expect to see histogram distributions on the bottom left (low MIC, no allele) and the top right (high MIC, presence of candidate allele(s)). For the first antimicrobial in each class, we show the number of candidate AMR alleles in square brackets, along with the number of isolates where candidate alleles are absent or present. r^2 values show the proportion of the variance explained by the binary presence/absence of these alleles.

Figure 2. Genotype-specific MIC values for lincomycin and erythromycin show epistatic interactions. The middle panels show the MIC values for lincomycin and erythromycin (both in the MLS_B class), for a range of genotypes with mean MIC values shown as black empty circles. Genotypes are classified in the lower panel, according to the subset of the 8 candidate

alleles for the MLS_B class that they contain, and the number of isolates with each genotype are shown in the upper panel. The plots illustrate several complexities of the genotype-phenotype map, including alleles that confer resistance to only one of the two drugs, and epistatic interactions, i.e., resistance effects that vary with the genetic background.

Figure 3. Covariation in MIC values for different antibiotic classes. The matrix represents correlation coefficients between log(MIC) values for pairs of antibiotics. The greyscale runs from white (-0.14) to black (1). Correlations are strongest for most antibiotics in the same class (as indicated by colour), but there are also strong correlations between classes. All values are positive, except for some values involving enrofloxacin (enr), which does not even correlate strongly with the MIC for the other Fluoroquinolone, marbofloxacin (mar). The inset figure plots the correlation coefficients on the *y*-axis (with values involving enrofloxacin shown in yellow). The *x*-axis shows the co-occurrence of candidate alleles conferring resistance to each pair of antibiotics. These were calculated as $\rho_{\text{allele pres/abs}} = (p_{ij} - p_i p_j)/\sqrt{p_i(1-p_i)p_j(1-p_j)}$, where p_i and p_j are the proportions of the isolates carrying candidate resistance alleles against each of the classes, and p_{ij} is the proportion of isolates carrying alleles against both classes. The trend suggests that much of the covariance in MIC values is due to co-occurring resistance alleles.

Figure 4 Clustering of AMR alleles in *S. suis.* Histograms of the number of candidate resistance alleles carried by each isolate (upper panel), or the

number of antibiotic classes to which these alleles confer resistance (lower panel). The true data (thicker blue bars), are compared to two comparison data sets, in which the presence/absence records of each allele was randomized over isolates (thinner red bars), or where the appearance of each allele was placed at random on a consensus genealogy of the isolates (thinner yellow bars). These randomization procedures suggest that the co-occurrence of candidate resistance alleles, and especially alleles conferring resistance to different drug classes, is more clustered than would be expected by chance.

1. de Moor CE. 1963. Septicaemic infections in pigs, caused by haemolytic streptococci of new Lancefield groups designated R, S and T. Antonie van Leeuwenhoek 29:272-280.
2. Wertheim HFL, Nguyen HN, Taylor W, Lien TTM, Ngo HT, Nguyen TQ, Nguyen BNT, Nguyen HH, Nguyen HM, Nguyen CT, Dao TT, Nguyen TV, Fox A, Farrar J, Schultsz C, Nguyen HD, Nguyen KV, Horby P. 2009. *Streptococcus suis*, an Important Cause of Adult Bacterial Meningitis in Northern Vietnam. PLOS ONE 4:e5973.
3. Baums CG, Kock C, Beineke A, Bennecke K, Goethe R, Schröder C, Waldmann K-H, Valentin-Weigand P. 2009. *Streptococcus suis* bacterin and subunit vaccine immunogenicities and protective efficacies against serotypes 2 and 9. Clinical and vaccine immunology 16:200-208.
4. van Rennings L, von Münchhausen C, Ottilie H, Hartmann M, Merle R, Honscha W, Käsbohrer A, Kreienbrock L. 2015. Cross-sectional study on antibiotic usage in pigs in Germany. PloS one 10:e0119114.
5. McGlone JJ. 2013. The Future of Pork Production in the World: Towards Sustainable, Welfare-Positive Systems. Animals 3:401-415.
6. Steinfeld H, Gerber PJ, Wassenaar T, Castel V, Rosales M, De haan C. 2006. Livestock's Long Shadow: Environmental Issues and Options, vol 24.
7. Goyette-Desjardins G, Auger J-P, 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. Emerging Microbes and Infections 3:1-20.
8. Van Boeckel TP, Brower C, Gilbert M, Grenfell BT, Levin SA, Robinson TP, Teillant A, Laxminarayan R. 2015. Global trends in antimicrobial use in food animals. Proceedings of the National Academy of Sciences 112:5649-5654.

939 9. Hernandez-Garcia J, Wang J, Restif O, Holmes MA, Mather AE, Weinert LA,
940 Wileman TM, Thomson JR, Langford PR, Wren BW, Rycroft A, Maskell DJ,
941 Tucker AW, Consortium BT. 2017. Patterns of antimicrobial resistance in
942 *Streptococcus suis* isolates from pigs with or without streptococcal disease
943 in England between 2009 and 2014. Veterinary microbiology 207:117-
944 124.
945 10. Yongkiettrakul S, Maneerat K, Arechanajan B, Malila Y, Srimanote P,
946 Gottschalk M, Visessanguan W. 2019. Antimicrobial susceptibility of
947 *Streptococcus suis* isolated from diseased pigs, asymptomatic pigs, and
948 human patients in Thailand. BMC Veterinary Research 15:5.
949 11. Marie J, Morvan H, Berthelot-Hérault F, Sanders P, Kempf I, Gautier-
950 Bouchardon AV, Jouy E, Kobisch M. 2002. Antimicrobial susceptibility of
951 *Streptococcus suis* isolated from swine in France and from humans in
952 different countries between 1996 and 2000. Journal of Antimicrobial
953 Chemotherapy 50:201-209.
954 12. Šeol B, Kelnerić Ž, Hajsig D, Madić J, Naglić T. 1996. Susceptibility to
955 Antimicrobial Agents of *Streptococcus suis* Capsular Type 2 Strains
956 Isolated from Pigs. Zentralblatt für Bakteriologie 283:328-331.
957 13. Turgeon PL, Higgins R, Gottschalk M, Beaudoin M. 1994. Antimicrobial
958 susceptibility of *Streptococcus suis* isolates. British Veterinary Journal
959 150:263-269.
960 14. Gurung M, Tamang MD, Moon DC, Kim S-R, Jeong J-H, Jang G-C, Jung S-C,
961 Park Y-H, Lim S-K. 2015. Molecular Basis of Resistance to Selected
962 Antimicrobial Agents in the Emerging Zoonotic Pathogen *Streptococcus*
963 *suis*. Journal of Clinical Microbiology 53:2332-2336.
964 15. Yao J, Shang K, Huang J, Ran W, Kashif J, Wang L. 2014. Overexpression of
965 an ABC transporter and mutations of GyrA, GyrB, and ParC in contributing
966 to high-level ciprofloxacin resistance in *Streptococcus suis* type 2.
967 BioScience Trends 8:84-92.
968 16. Hoa NT, Chieu TT, Nghia HD, Mai NT, Anh PH, Wolbers M, Baker S,
969 Campbell JI, Chau NV, Hien TT, Farrar J, Schultsz C. 2011. The
970 antimicrobial resistance patterns and associated determinants in
971 *Streptococcus suis* isolated from humans in southern Vietnam, 1997-2008.
972 BMC Infectious Diseases 11:6.
973 17. Huang J, Ma J, Shang K, Hu X, Liang Y, Li D, Wu Z, Dai L, Chen L, Wang L.
974 2016. Evolution and Diversity of the Antimicrobial Resistance Associated
975 Mobilome in *Streptococcus suis*: A Probable Mobile Genetic Elements
976 Reservoir for Other Streptococci. Frontiers in Cellular and Infection
977 Microbiology 6.
978 18. Hu P, Yang M, Zhang A, Wu J, Chen B, Hua Y, Yu J, Chen H, Xiao J, Jin M.
979 2011. Comparative genomics study of multi-drug-resistance mechanisms
980 in the antibiotic-resistant *Streptococcus suis* R61 strain. PloS one
981 6:e24988.
982 19. Holden MT, Hauser H, Sanders M, Ngo TH, Cherevach I, Cronin A,
983 Goodhead I, Mungall K, Quail MA, Price C, Rabbinowitsch E, Sharp S,
984 Croucher NJ, Chieu TB, Mai NT, Diep TS, Chinh NT, Kehoe M, Leigh JA,
985 Ward PN, Dowson CG, Whatmore AM, Chanter N, Iversen P, Gottschalk M,
986 Slater JD, Smith HE, Spratt BG, Xu J, Ye C, Bentley S, Barrell BG, Schultsz C,
987 Maskell DJ, Parkhill J. 2009. Rapid evolution of virulence and drug

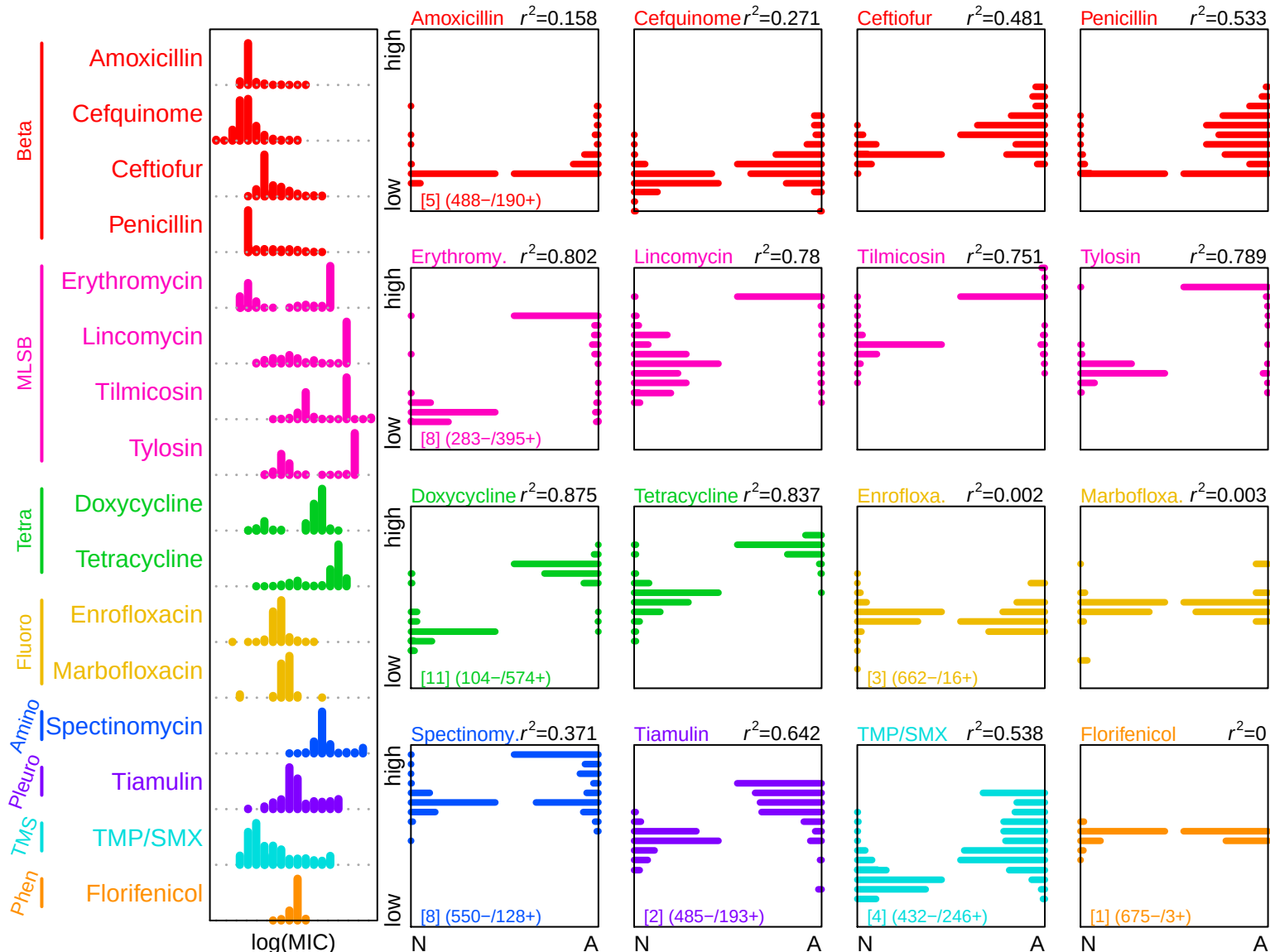
988 resistance in the emerging zoonotic pathogen *Streptococcus suis*. PLoS
989 One 4:e6072.
990 20. Weinert LA, Chaudhuri RR, Wang J, Peters SE, Corander J, Jombart T, Baig
991 A, Howell KJ, Vehkala M, Välimäki N, Harris D, Chieu TTB, Van Vinh Chau
992 N, Campbell J, Schultsz C, Parkhill J, Bentley SD, Langford PR, Rycroft AN,
993 Wren BW, Farrar J, Baker S, Hoa NT, Holden MTG, Tucker AW, Maskell DJ,
994 Bossé JT, Li Y, Maglennon GA, Matthews D, Cuccui J, Terra V, Consortium
995 BRT. 2015. Genomic signatures of human and animal disease in the
996 zoonotic pathogen *Streptococcus suis*. Nature Communications 6:6740.
997 21. Wileman TM, Weinert LA, Howell KJ, Wang J, Peters SE, Williamson SM,
998 Wells JM, Langford PR, Rycroft AN, Wren BW, Maskell DJ, Tucker AW.
999 2019. Pathotyping the Zoonotic Pathogen *Streptococcus suis*: Novel
1000 Genetic Markers To Differentiate Invasive Disease-Associated Isolates
1001 from Non-Disease-Associated Isolates from England and Wales. Journal of
1002 Clinical Microbiology 57:e01712-18.
1003 22. Zou G, Zhou J, Xiao R, Zhang L, Cheng Y, Jin H, Li L, Zhang L, Wu B, Qian P,
1004 Li S, Ren L, Wang J, Oshota O, Hernandez-Garcia J, Wileman TM, Bentley S,
1005 Weinert L, Maskell DJ, Tucker AW, Zhou R. 2018. Effects of Environmental
1006 and Management-Associated Factors on Prevalence and Diversity of
1007 *Streptococcus suis* in Clinically Healthy Pig Herds in China and the United
1008 Kingdom. Applied and Environmental Microbiology 84:e02590-17.
1009 23. van Hout J, Heuvelink A, Gonggrijp M. 2016. Monitoring of antimicrobial
1010 susceptibility of *Streptococcus suis* in the Netherlands, 2013–2015.
1011 Veterinary Microbiology 194:5-10.
1012 24. Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, Lago BA,
1013 Dave BM, Pereira S, Sharma AN, Doshi S, Courtot M, Lo R, Williams LE,
1014 Frye JG, Elsayegh T, Sardar D, Westman EL, Pawlowski AC, Johnson TA,
1015 Brinkman FSL, Wright GD, McArthur AG. 2017. CARD 2017: expansion
1016 and model-centric curation of the comprehensive antibiotic resistance
1017 database. Nucleic acids research 45:D566-D573.
1018 25. Ge Y, Wu J, Xia Y, Yang M, Xiao J, Yu J. 2012. Molecular dynamics
1019 simulation of the complex PBP-2x with drug cefuroxime to explore the
1020 drug resistance mechanism of *Streptococcus suis* R61. PLoS One 7:e35941.
1021 26. Palmieri C, Magi G, Mingoia M, Bagnarelli P, Ripa S, Varaldo PE, Facinelli B.
1022 2012. Characterization of a *Streptococcus suis* tet(O/W/32/O)-carrying
1023 element transferable to major streptococcal pathogens. Antimicrobial
1024 Agents and Chemotherapy 56:4697-702.
1025 27. Escudero JA, San Millan A, Catalan A, de la Campa AG, Rivero E, Lopez G,
1026 Dominguez L, Moreno MA, Gonzalez-Zorn B. 2007. First characterization
1027 of fluoroquinolone resistance in *Streptococcus suis*. Antimicrobial Agents
1028 and Chemotherapy 51:777-82.
1029 28. Albarracin Orio AG, Pinas GE, Cortes PR, Cian MB, Echenique J. 2011.
1030 Compensatory evolution of pbp mutations restores the fitness cost
1031 imposed by beta-lactam resistance in *Streptococcus pneumoniae*. PLoS
1032 Pathogens 7:e1002000.
1033 29. Leclercq R. 2002. Mechanisms of Resistance to Macrolides and
1034 Lincosamides: Nature of the Resistance Elements and Their Clinical
1035 Implications. Clinical Infectious Diseases 34:482-492.

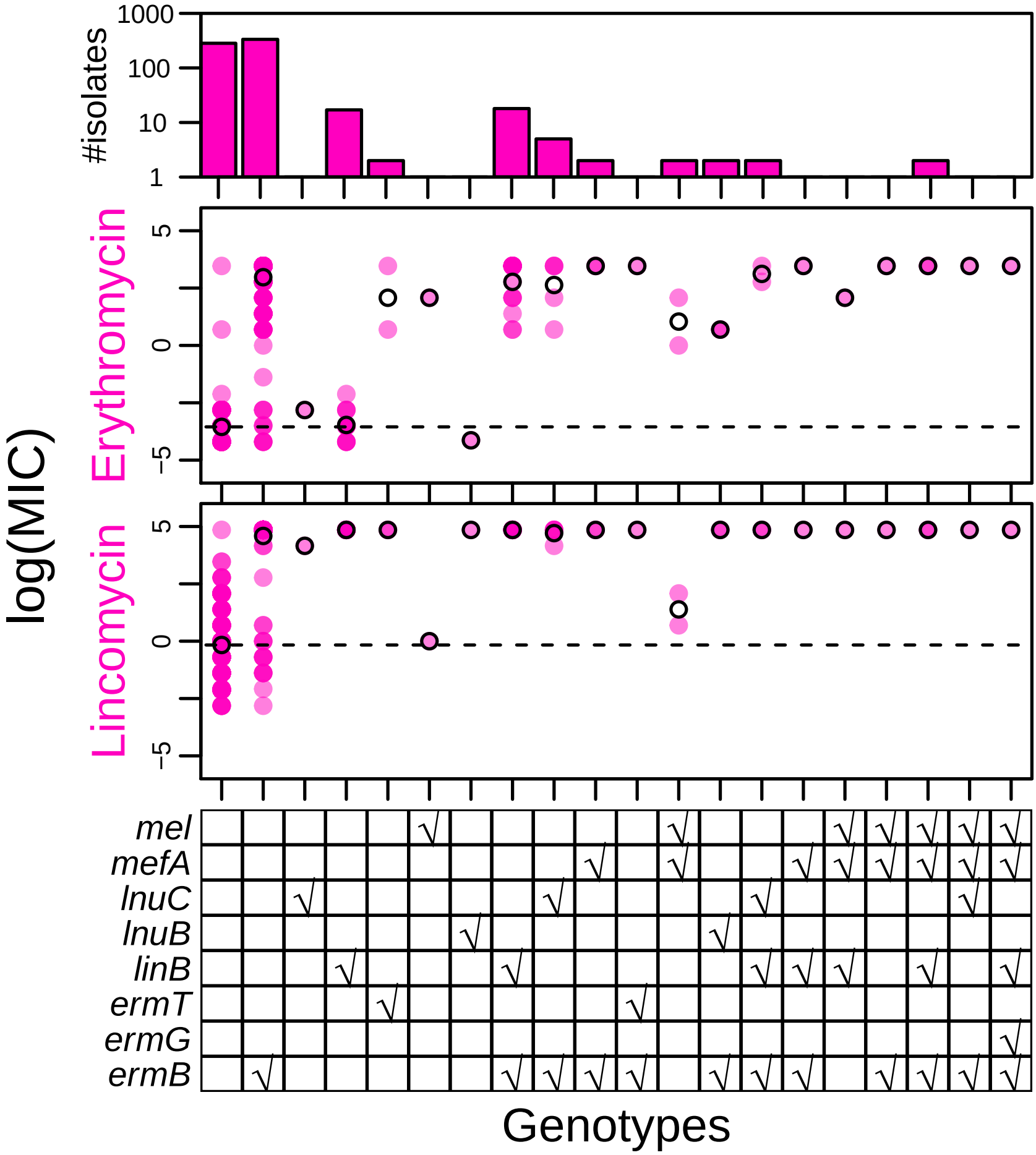
1036 30. Lehtinen S, Blanquart F, Lipsitch M, Fraser C, with the Maela
1037 Pneumococcal C. 2019. On the evolutionary ecology of multidrug
1038 resistance in bacteria. PLoS Pathogens 15:e1007763.
1039 31. Calvez P, Breukink E, Roper DI, Dib M, Contreras-Martel C, Zapun A. 2017.
1040 Substitutions in PBP2b from beta-Lactam-resistant *Streptococcus*
1041 *pneumoniae* Have Different Effects on Enzymatic Activity and Drug
1042 Reactivity. Journal of Biological Chemistry 292:2854-2865.
1043 32. Coffey TJ, Daniels M, McDougal LK, Dowson CG, Tenover FC, Spratt BG.
1044 1995. Genetic analysis of clinical isolates of *Streptococcus pneumoniae*
1045 with high-level resistance to expanded-spectrum cephalosporins.
1046 Antimicrobial Agents and Chemotherapy 39:1306-13.
1047 33. Grebe T, Hakenbeck R. 1996. Penicillin-binding proteins 2b and 2x of
1048 *Streptococcus pneumoniae* are primary resistance determinants for
1049 different classes of beta-lactam antibiotics. Antimicrobial Agents and
1050 Chemotherapy 40:829-34.
1051 34. Haenni M, Galofaro L, Ythier M, Giddey M, Majcherczyk P, Moreillon P,
1052 Madec JY. 2010. Penicillin-binding protein gene alterations in
1053 *Streptococcus uberis* isolates presenting decreased susceptibility to
1054 penicillin. Antimicrobial Agents and Chemotherapy 54:1140-5.
1055 35. Piccinelli G, Carlentini G, Gargiulo F, Caruso A, De Francesco MA. 2017.
1056 Analysis of Point Mutations in the pbp2x, pbp2b, and pbp1a Genes of
1057 *Streptococcus agalactiae* and Their Relation with a Reduced Susceptibility
1058 to Cephalosporins. Microbial Drug Resistance 23:1019-1024.
1059 36. Kocaoglu O, Tsui HC, Winkler ME, Carlson EE. 2015. Profiling of beta-
1060 lactam selectivity for penicillin-binding proteins in *Streptococcus*
1061 *pneumoniae* D39. Antimicrobial Agents and Chemotherapy 59:3548-55.
1062 37. Figueiredo TA, Sobral RG, Ludovice AM, Almeida JM, Bui NK, Vollmer W,
1063 de Lencastre H, Tomasz A. 2012. Identification of genetic determinants
1064 and enzymes involved with the amidation of glutamic acid residues in the
1065 peptidoglycan of *Staphylococcus aureus*. PLoS Pathogens 8:e1002508.
1066 38. Zapun A, Philippe J, Abrahams KA, Signor L, Roper DI, Breukink E, Vernet
1067 T. 2013. In vitro reconstitution of peptidoglycan assembly from the Gram-
1068 positive pathogen *Streptococcus pneumoniae*. ACS Chemical Biology
1069 8:2688-96.
1070 39. Chewapreecha C, Marttinen P, Croucher NJ, Salter SJ, Harris SR, Mather
1071 AE, Hanage WP, Goldblatt D, Nosten FH, Turner C, Turner P, Bentley SD,
1072 Parkhill J. 2014. Comprehensive identification of single nucleotide
1073 polymorphisms associated with beta-lactam resistance within
1074 pneumococcal mosaic genes. PLoS Genetics 10:e1004547.
1075 40. Dunai A, Spohn R, Farkas Z, Lázár V, Györkei Á, Apjok G, Boross G,
1076 Szappanos B, Grézal G, Faragó A, Bodai L, Papp B, Pál C. 2019. Rapid
1077 decline of bacterial drug-resistance in an antibiotic-free environment
1078 through phenotypic reversion. eLife 8:e47088.
1079 41. Directorate VM. 2018. UK veterinary antibiotic resistance and sales
1080 surveillance report.
1081 42. Dunlop RH, McEwen SA, Meek AH, Black WD, Clarke RC, Friendship RM.
1082 1998. Individual and group antimicrobial usage rates on 34 farrow-to-
1083 finish swine farms in Ontario, Canada. Preventive Veterinary Medicine
1084 34:247-264.

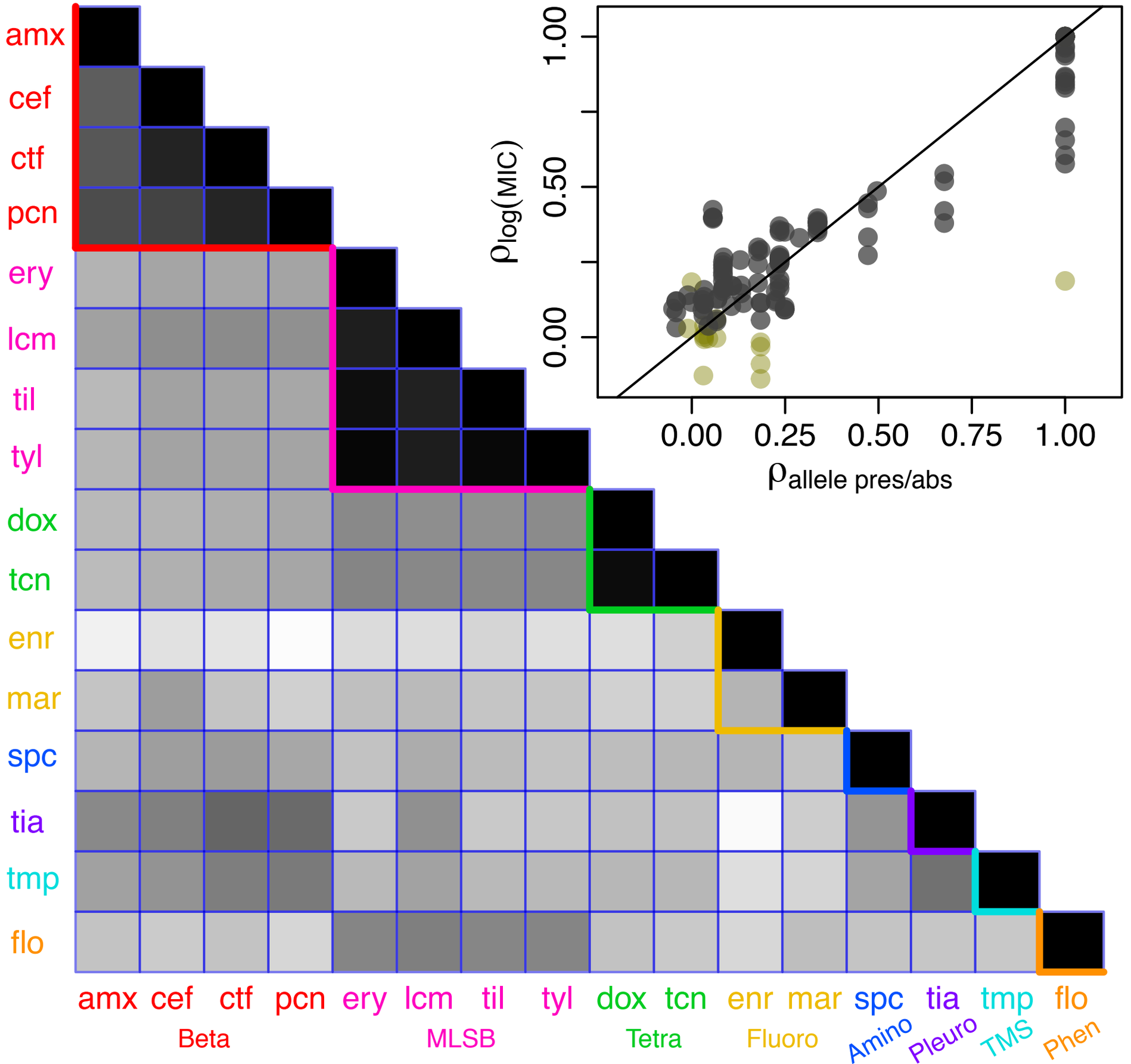
1085 43. Carrique-Mas JJ, Choisy M, Van Cuong N, Thwaites G, Baker S. 2020. An
1086 estimation of total antimicrobial usage in humans and animals in Vietnam.
1087 Antimicrobial Resistance and Infection Control 9:16.
1088 44. Lehtinen S, Blanquart F, Croucher NJ, Turner P, Lipsitch M, Fraser C. 2017.
1089 Evolution of antibiotic resistance is linked to any genetic mechanism
1090 affecting bacterial duration of carriage. Proceedings of the National
1091 Academy of Sciences 114:1075-1080.
1092 45. de Jong A, Thomas V, Simjee S, Moyaert H, El Garch F, Maher K, Morrissey
1093 I, Butty P, Klein U, Marion H, Rigaut D, Vallé M. 2014. Antimicrobial
1094 susceptibility monitoring of respiratory tract pathogens isolated from
1095 diseased cattle and pigs across Europe: The VetPath study. Veterinary
1096 Microbiology 172:202-215.
1097 46. Quail MA, Kozarewa I, Smith F, Scally A, Stephens PJ, Durbin R, Swerdlow
1098 H, Turner DJ. 2008. A large genome center's improvements to the Illumina
1099 sequencing system. Nature Methods 5:1005-1010.
1100 47. Joshi N, Fass J. 2011. Sickle: A sliding-window, adaptive, quality-based
1101 trimming tool for FastQ files.
1102 48. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS,
1103 Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV,
1104 Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new
1105 genome assembly algorithm and its applications to single-cell sequencing.
1106 Journal of Computational Biology 19:455-477.
1107 49. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation.
1108 Bioinformatics 30:2068-2069.
1109 50. Wingett SW, Andrews S. 2018. FastQ Screen: A tool for multi-genome
1110 mapping and quality control. F1000Research 7:1338-1338.
1111 51. Li H, Durbin R. 2009. Fast and accurate short read alignment with
1112 Burrows–Wheeler transform. Bioinformatics 25:1754-1760.
1113 52. Cheng L, Connor TR, Sirén J, Aanensen DM, Corander J. 2013. Hierarchical
1114 and Spatially Explicit Clustering of DNA Sequences with BAPS Software.
1115 Molecular Biology and Evolution 30:1224-1228.
1116 53. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M,
1117 Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote
1118 pan genome analysis. Bioinformatics 31:3691-3693.
1119 54. Ranwez V, Harispe S, Delsuc F, Douzery EJP. 2011. MACSE: Multiple
1120 Alignment of Coding SEquences Accounting for Frameshifts and Stop
1121 Codons. PLOS ONE 6:e22594.
1122 55. Team RC. 2013. R: A language and environment for statistical computing.
1123 56. Kimura M. 1980. A simple method for estimating evolutionary rates of
1124 base substitutions through comparative studies of nucleotide sequences.
1125 Journal of Molecular Evolution 16:111-120.
1126 57. Paradis E. 2011. Analysis of Phylogenetics and Evolution with R. Springer
1127 Science & Business Media.
1128 58. Paradis E, Schliep K. 2018. ape 5.0: an environment for modern
1129 phylogenetics and evolutionary analyses in R. Bioinformatics 35:526-528.
1130 59. Schliep KP. 2011. phangorn: phylogenetic analysis in R. Bioinformatics
1131 27:592-593.

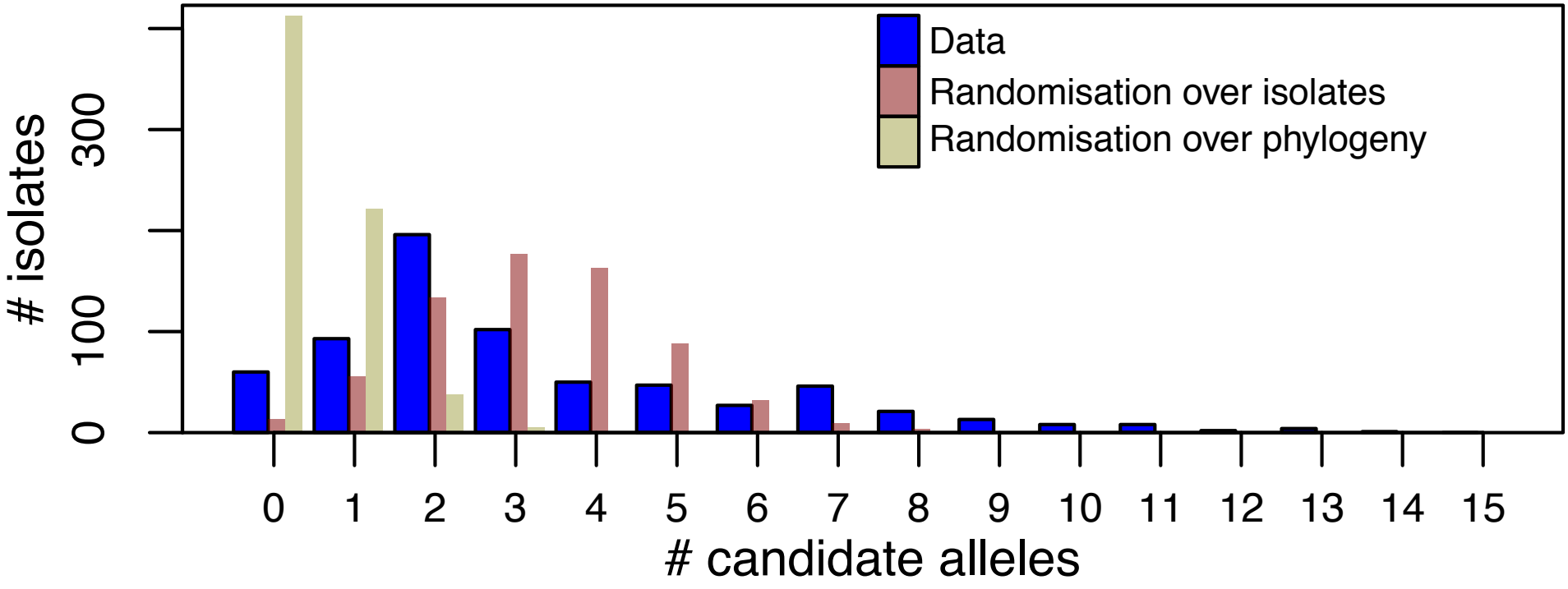
1132 60. Hunt M, Mather AE, Sánchez-Busó L, Page AJ, Parkhill J, Keane JA, Harris
1133 SR. 2017. ARIBA: rapid antimicrobial resistance genotyping directly from
1134 sequencing reads. Microbial genomics 3:e000131.
1135 61. Lopez P, Espinosa M, Greenberg B, Lacks St. 1987. Sulfonamide resistance
1136 in *Streptococcus pneumoniae*: DNA sequence of the gene encoding
1137 dihydropteroate synthase and characterization of the enzyme. Journal of
1138 Bacteriology 169:4320-4326.
1139 62. Adrian PV, Klugman KP. 1997. Mutations in the dihydrofolate reductase
1140 gene of trimethoprim-resistant isolates of *Streptococcus pneumoniae*.
1141 Antimicrobial Agents and Chemotherapy 41:2406-13.
1142 63. Hakenbeck R, Grebe T, Zähner D, Stock JB. 1999. β-Lactam resistance in
1143 *Streptococcus pneumoniae*: penicillin-binding proteins and non-penicillin-
1144 binding proteins. Molecular Microbiology 33:673-678.
1145 64. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high
1146 accuracy and high throughput. Nucleic acids research 32:1792-1797.
1147 65. Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: A multiplatform
1148 graphical user interface for sequence alignment and phylogenetic tree
1149 building. Molecular Biology and Evolution 27:221-4.
1150 66. Contreras-Martel C, Dahout-Gonzalez C, Martins Ados S, Kotnik M, Dessen
1151 A. 2009. PBP active site flexibility as the key mechanism for beta-lactam
1152 resistance in pneumococci. Journal of Molecular Biology 387:899-909.
1153 67. Fuursted K, Stegger M, Hoffmann S, Lambertsen L, Andersen PS, Deleuran
1154 M, Thomsen MK. 2016. Description and characterization of a penicillin-
1155 resistant *Streptococcus dysgalactiae* subsp. *equisimilis* clone isolated from
1156 blood in three epidemiologically linked patients. Journal of Antimicrobial
1157 Chemotherapy 71:3376-3380.
1158 68. Haroche J, Allignet J, Buchrieser C, El Solh N. 2000. Characterization of a
1159 Variant of vga (A) Conferring Resistance to Streptogramin A and Related
1160 Compounds. Antimicrobial Agents and Chemotherapy 44:2271.
1161 69. Pikis A, Donkersloot JA, Rodriguez WJ, Keith JM. 1998. A conservative
1162 amino acid mutation in the chromosome-encoded dihydrofolate
1163 reductase confers trimethoprim resistance in *Streptococcus pneumoniae*.
1164 Journal of Infectious Diseases 178:700-6.
1165 70. Bergmann R, van der Linden M, Chhatwal GS, Nitsche-Schmitz DP. 2014.
1166 Factors that cause trimethoprim resistance in *Streptococcus pyogenes*.
1167 Antimicrobial Agents and Chemotherapy 58:2281-8.
1168 71. Athey TBT, Teatero S, Lacouture S, Takamatsu D, Gottschalk M, Fittipaldi
1169 N. 2016. Determining *Streptococcus suis* serotype from short-read whole-
1170 genome sequencing data. BMC microbiology 16:162-162.
1171 72. Pinheiro J, Bates D, DebRoy S, Sarkar D, Heisterkamp S, Van Willigen B,
1172 Maintainer R. 2019. Package 'nlme'.
1173 73. Palmieri C, Varaldo PE, Facinelli B. 2011. *Streptococcus suis*, an Emerging
1174 Drug-Resistant Animal and Human Pathogen. Frontiers in Microbiology
1175 2:235.
1176 74. Huang K, Zhang Q, Song Y, Zhang Z, Zhang A, Xiao J, Jin M. 2016.
1177 Characterization of Spectinomycin Resistance in *Streptococcus suis* Leads
1178 to Two Novel Insights into Drug Resistance Formation and Dissemination
1179 Mechanism. Antimicrobial Agents and Chemotherapy 60:6390-2.

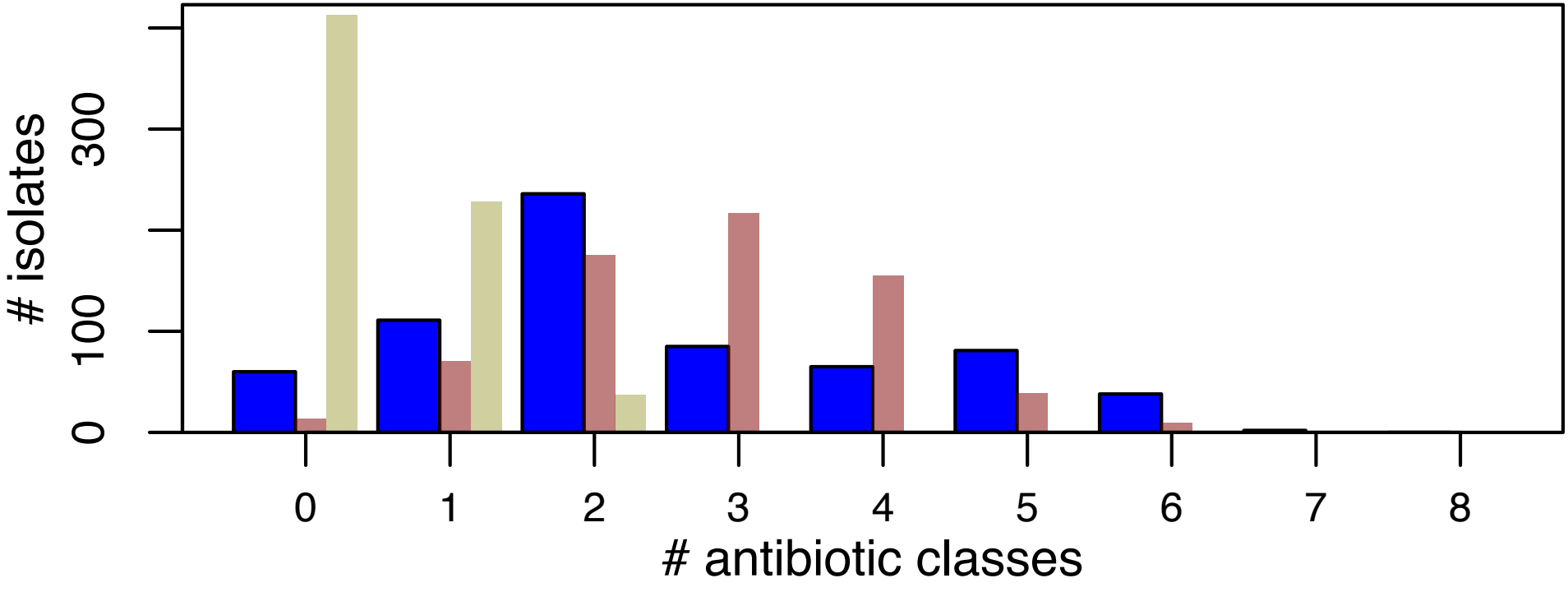
1180 75. Chang B, Wada A, Ikebe T, Ohnishi M, Mita K, Endo M, Matsuo H, Asatuma
1181 Y, Kuramoto S, Sekiguchi H, Yamazaki M, Yoshikawa H, Watabe N, Yamada
1182 H, Kurita S, Imai Y, Watanabe H. 2006. Characteristics of *Streptococcus*
1183 *suis* isolated from patients in Japan. Japanese Journal of Infectious
1184 Diseases 59:397-9.
1185 76. Bojarska A, Molska E, Janas K, Skoczynska A, Stefaniuk E, Hryniewicz W,
1186 Sadowy E. 2016. *Streptococcus suis* in invasive human infections in
1187 Poland: clonality and determinants of virulence and antimicrobial
1188 resistance. European Journal of Clinical Microbiology and Infectious
1189 Diseases 35:917-25.
1190 77. Athey TB, Auger JP, Teatero S, Dumesnil A, Takamatsu D, Wasserscheid J,
1191 Dewar K, Gottschalk M, Fittipaldi N. 2015. Complex Population Structure
1192 and Virulence Differences among Serotype 2 *Streptococcus suis* Strains
1193 Belonging to Sequence Type 28. PLoS One 10:e0137760.
1194


Beta
MLSB
Tetra
Fluoro
Amino
Pleuro
TMS
Phen
Amoxicillin
Cefquinome
Ceftiofur
Penicillin
Erythromycin
Lincomycin
Tilmicosin
Tylosin
Doxycycline
Tetracycline
Enrofloxacin
Marbofloxacin
Spectinomycin
Tiamulin
TMP/SMX
Florifenicol
log(MIC)
high
low
Amoxicillin r2=0.158
[5] (488−/190+)
Cefquinome r2=0.271
Ceftiofur r2=0.481
Penicillin r2=0.533
Erythromy. r2=0.802
[8] (283−/395+)
Lincomycin r2=0.78
Tilmicosin r2=0.751
Tylosin r2=0.789
Doxycycline r2=0.875
[11] (104−/574+)
Tetracycline r2=0.837
Enrofloxa. r2=0.002
[3] (662−/16+)
Marbofloxa. r2=0.003
Spectinomy. r2=0.371
[8] (550−/128+)
Tiamulin r2=0.642
[2] (485−/193+)
TMP/SMX r2=0.538
[4] (432−/246+)
Florifenicol r2=0
[1] (675−/3+)
N
A


#isolates
1000
100
10
1
log(MIC)
Erythromycin
Lincomycin
5
0
−5
mel
mefA
lnuC
lnuB
linB
ermT
ermG
ermB
Genotypes


amx
cef
ctf
pcn
ery
lcm
til
tyl
dox
tcn
enr
mar
spc
tia
tmp
flo
Beta
MLSB
Tetra
Fluoro
Amino
Pleuro
TMS
Phen
$\rho_{log(MIC)}$
$\rho_{allele\ pres/abs}$
0.00
0.50
1.00
0.00
0.25
0.50
0.75
1.00


Data
Randomisation over isolates
Randomisation over phylogeny
isolates
0
100
300
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
candidate alleles


isolates
0
100
300
0 1 2 3 4 5 6 7 8
antibiotic classes