Antimicrobial Resistance in Biofilms: Discussion

4.1 ARG Findings

After filtering by length and identity, 144 ARGs were detected in the full set of samples, and none occurred in any controls. Of the 144 biofilm samples that were successfully sequenced, 96 samples were positive for at least one ARG. In this context, when we refer to an ARG, it may more properly be thought of as a gene that could potentially be associated with antimicrobial resistance (see section 4.2.3). Specifically, the definition of an ARG in this project includes any gene that is annotated as an ARG in CARD reference database.

When we consider the distribution of ARG frequencies (i.e. how many different ARGs are found in samples from each plant) these do appear similar. No plant has a clearly higher or different ARG frequency distribution, although Factory B has a lower (though non-significant) median number of ARGs per sample than Factory A (p=0.07). There is also an apparently longer tail for samples from factory D – the four most ARG-heavy samples were from this factory. Furthermore, as well as having the four samples with the highest number of different ARGs in them, factory D also had the highest average incidence of the ARGs across its samples. This means that, overall, each ARG found in plant A was found more often within a sample than were ARGs in other plants. This could be due to the same ARGs being found in multiple different bacteria or mobile genetic elements within a sample, for example.

The picture is similar with the different meat types processed, with samples associated with “pork and chicken” (that is, plants or lines through which both pork and chicken are processed) having a similar distribution to samples from factory A. However, it is difficult or impossible to disentangle the effect of meat type from the effect of factory, due to the small number of factories, with most focussing on only one or two meat species. Interestingly, there is some indication that the processing machinery exposed to non-meat media, i.e. brine or dextrose water, have relatively high incidence, although low sample numbers for these sample types make it impossible to draw firm conclusions. Both dextrose water samples had non-zero ARG counts, as did five of the six brine samples, compared with 88 of the remaining 136 other samples. There is some evidence to suggest that the presence of dextrose can enhance biofilm formation [44].

Regarding taxonomic assignment, although some of the formal ARG names in the ontology include names of bacteria, others may occur in multiple species. An analysis of the taxonomic origins of the ARGs is beyond the scope of this report, and further analysis is of interest to determine any implications of the above results for the types of bacteria found (for example, halophiles).

Of the ARGs detected, *rsmA* was found in by far the highest number of samples (n=73), and had the highest mean TPM value across samples, suggesting that where it is found it constitutes a relatively high proportion of all ARG reads within a sample. *rsmA* regulates virulence in Pseudomonas aeruginosa [45], and *rsmA* homologs (*csrA*) can regulate a wide range of factors including biofilm formation [46]. *rsmA* is annotated as an ARG in this instance as it can also be involved with regulation of multidrug efflux pumps, with an *rsmA* mutant showing increased resistance to amikacin, nalidixic acid, trimethoprim, ceftazidime and gentamicin, and decreased resistance to polymyxin B and colistin [47]. The widespread detection of a P. aeruginosa-

associated gene is not surprising, as P. aeruginosa is known to be able to form biofilms on surfaces commonly used in food processing environments, including various plastics [48, 49] and stainless steel [49].

Other ARGs of note with high mean TPM values include the qac genes *qacE* [50], *cqaH* [51] and *qacL*. These genes encode small multidrug resistance (SMR) efflux pump genes carried on plasmids and transposons, which can confer resistance to quaternary ammonium compound (QAC) biocides [52]. QACs are widely used in industry as biocides, and biofilms have been demonstrated to increase resistance to QACs [53], so it is again unsurprising that qac genes are identified in biofilms in meat processing plants. While these genes can potentially confer resistance to QACs, they do not appear to confer resistance to antibiotics. However, they could still be of interest as they may promote the transfer of antibiotic resistance genes within the microbiome [54] and may mean that bacteria are more likely to survive and persist in biofilms if they are resistant to the use of QACs.

Some of the other ARGs identified highlight the difficulties associated with metagenomic sequencing. ARGs annotated as belonging to M. tuberculosis (*Mycobacterium tuberculosis* intrinsic *murA* conferring resistance to fosfomycin) and *K. pneumoniae* (*Klebsiella pneumoniae* KpnH) are unlikely to be associated with the taxa in question. Given the 80/80 percent identity/length these hits could well be to homologs of these genes in related taxa (for example, K. *oxytoca*) or even more distantly related species. This caveat is also relevant to the observations of P. aeruginosa *rsmA* discussed above, as upon further inspection, several of the identified ARGs appear to share a high percentage identity to *csrA*, and could therefore potentially be present in non-Pseudomonas species.

Other ARGs found at high mean TPMs illustrate the complex nature of other aspects of the analysis of biofilms undertaken in this project. Genes potentially involved in resistance to tetracycline (*tet*(H) and *tet*(K)) were found at relatively high mean TPMs, which does to some extent justify the selection of a tetracyline resistance gene as one of the qPCR targets. The specific gene chosen, *tet*(*B*), was not among the highest TPM ARGs, and neither was *sul1*, but the timescales of the project required that qPCR targets were identified ahead of the results of the sequence analysis being available. In this case, the results of FS301050 were used to select genes identified in cooked meats.

There are important difficulties in drawing firm conclusions about the ARG burden of biofilms, for example, the degree to which they contribute to the ARGs present on foods produced in biofilm-containing factories. This is principally due to lack of comparable samples. There is little published in the literature that applies the same techniques used here to relevant sample types (for example, secondary meat products), and no samples were taken in this project from intake carcasses/outflow products. That being the case, we have taken two published studies, using two different approaches, and attempted to compare our data to them.

One study is a small-scale study of ARG prevalence in chicken processed in the UK and Ireland [38]. In that study an array-based detection method was applied, which was deemed a relevant comparison as it was a molecular method, not targeting a single bacterial species, although in this instance it did follow an enrichment for Gram-negative bacteria. That being the case, we also confined our comparisons to biofilm samples which contained evidence of Gram-negative bacteria. Other limitations to the comparison included the limited number of ARGs tested for, the small number of samples tested by the array, and the fact that the samples were not contemporaneous with the biofilm samples. Overall observations of ARGs were observed in a smaller proportion of Gram-negative samples taken from biofilms than were seen in Gram-negative samples taken from chicken. There may be a number of different drivers for this observation. One possibility is that bacteria in factory biofilms actually contain fewer ARGs than bacteria in chicken. Another reason for this observation may be that the method used to test for ARGs in chicken, is more sensitive than the method of detection applied to factory biofilm. Hence,

it is difficult to tell whether looking for ARGs in chicken may be more informative than looking for ARGs in biofilms, or whether looking for ARGs using the method applied to chicken samples is more informative than looking for ARGs using the method applied to biofilm samples.

Even more difficulties are encountered when trying to compare our data with phenotypic results from isolates from retail meat surveys. In predicting which genes confer resistance to antibiotics and comparing the biofilms with retail meat surveys (Section 3.6.1) some simplifications were necessary. A major caveat is that the particular genes identified in our study do not necessarily confer resistance to all antibiotics within a particular class. However, the antibiotic class was the only annotation available from CARD/aro and it was therefore relatively simple to identify common classes. A more detailed investigation to check for resistance to individual antibiotics would involve a specific literature search and would be a much larger task, because all 26 antibiotic drug classes (Table 15) would need to be checked.

An assessment was also made to identify which of the genes found in our study are consistent with E. *coli*, though some of these genes may also be present (with exactly the same sequences) in other species. Multiple resistance mechanisms are observed, with some conferring resistance by their presence (as single gene system) and others as efflux pumps or parts of operon. Those that are not single gene systems do not necessarily imply phenotypic resistance, so it may be that they should be excluded form analysis. For example, the following ARGs were found in our biofilm data and are not single gene systems (number of samples shown in parentheses):

CRP (12), *emrR* (12), *msbA* (9), *Klebsiella pneumoniae KpnH* (8), *marA* (5), *baeR* (1), *Escherichia coli marR* mutant conferring antibiotic resistance (1), *Escherichia coli soxR* with mutation conferring antibiotic resistance (1), *Escherichia coli soxS* with mutation conferring antibiotic resistance (1), *PmrF* (1), Klebsiella pneumoniae *KpnF* (1).

The assessment could be repeated with these genes excluded. This kind of assessment is feasible when small numbers of metagenomically-identified ARGs are concerned, but less so for large scale studies.

4.2 Technical Considerations

4.2.1 Sampling Strategy

The sampling of biofilms was limited to only a small number of factories. However, as shown in Section 3.6.3, and particularly summarised in Table 17, the main types of secondary meat products consumed within the UK are represented by processing within these factories (chicken products, pork sausages and bacon). Furthermore, in Section 3.6.4 the representativeness of the sampled production areas within these factories was considered, comparing the types of meat processed there with the main consumed types. We believe that the samples provide a good coverage of the most commonly consumed items. The only category that is less well represented is minced beef or beef burgers.

Data were not available to quantify the amounts of processed meat purchased from these factories as a proportion of the total UK purchase and consumption. As a result, only limited information is available about the between-factory variation in ARGs. A quantitative estimate of the proportion of consumers represented by the survey would require information about all processors within the meat supply chain.

With regards to the actual sampling implementation, due to covid restrictions we were unable to be physically present for any of the sampling locations or times. We were reliant on the goodwill and expertise of the factory staff to take samples, for which we are very grateful. However, this does introduce another source of variability.

4.2.2 SOP Development

After recommendation by experts in food production hygiene, the use of the Biofinder spray was trialled for identifying biofilms. However, as the Biofinder spray contains hydrogen peroxide, there was a concern that it may degrade any DNA present prior to sequencing. After sequencing the test material, the Biofinder + MGW rinse and the non-Biofinder treatments generated similar percentages of both bacteria present. The non-Biofinder method detected more *Pseudomonas*, though all methods detected comparatively little *Pseudomonas*. This was interesting as twice as much *Pseudomonas* was inoculated onto the slides as E. coli. The reasons for this are unclear. Potentially the E. coli was able to replicate faster than the *Pseudomonas* from the same initial inoculum, as has been shown previously [55], or was able to form a stronger biofilm by attaching and proliferating more readily. Unfortunately, a detailed investigation of this was beyond the scope and budget of the current project.

4.2.3 Sequencing considerations

Challenges which affected some samples were low DNA yields, and possible contamination.

The factories reported that Biofinder foaming was not frequent, and so factories were required to sample even where it did not foam, but in places where we anticipated likely biofilm formation. Even where Biofinder was negative, samples corresponded to the definition of biofilms used in this report for example, bacteria and other microorganisms that remain adhered to a surface (presumably by extracellular secreted substances) immediately after surface rinsing. However, this potentially would have led to some of the lower DNA yields. Although only two samples generated unusably small numbers of DNA sequence reads, sampling from heavier biofilms (if present) may have generated higher DNA yields overall.

Regarding the source of contamination, the sampling protocol was shown to work with the test biofilms, which suggests that factories did not generally harbour strong biofilms, at least in the samples tested. However, there was a significant amount of DNA sequence obtained in the controls. This is itself is not surprising, as the volume of control samples incorporated into the sequencing pool is the same as the average volume pooled across all samples. Given the high depth of sequencing undertaken in this research project, even relatively low amounts of DNA obtained from the controls would be expected to generate sequence data. Nevertheless, the amounts of sequence generated for some of these controls is very high. This suggests at least two phenomena – the relatively low yields of DNA found in general among the biofilm samples, and the presence of contaminant ‘kitome’ DNA in the sampling and/or DNA extraction kits.

Increased depth of sequencing compared to previous studies (FS301050) improved our ability to assemble the DNA reads into contigs and meant that we could use assemblies rather than short reads. However, the type of sample also assisted in this regard (the previous study involved obtaining bacterial DNA from the surface of or within host organisms, increasing problems of non-bacterial contamination). It might also be expected that the bacterial populations of the biofilms are less diverse, generally providing greater coverage for any given genome segment.

This is beneficial because we are more likely to obtain longer and more accurate hits to ARGs which extend beyond the 250bp length of a single read. It also allows us to capture regions which may be unique to particular taxon, thus taxonomic identification of ARGs could be easier and more accurate. Furthermore, we benefit from error checks that happen during the assembly process, so have more confidence in the contigs produced. The potential disadvantage of analysing assemblies is that some genes/fragments represented by very few reads would not be incorporated into any contigs, effectively giving a false negative. This may affect samples with particularly low read counts. With short-read analysis, we previously showed that it is possible identify some individual reads as almost certainly originating from a longer ARG fragment (FS301050). However, in the general case, analysis of unassembled short reads presents far

greater complications in filtering out false positives, and therefore our view is that ARG sequence detection is better performed on assemblies.

Long-read sequencing appeared to perform better than in previous studies (FS301050), which is perhaps surprising due to the generally low amounts of DNA obtained in this study. This may reflect advances in ONT library preparation and sequencing chemistry. The average read length was still slightly lower than the 1-2 kb which would be ideal for the assembly software (Flye). However, hybrid-assembly with the Illumina short reads was very successful, though not in every case were the N50 and L50 scores better than those in the Illumina-only counterparts.

It is more likely that we would obtain structurally accurate contigs with hybrid assemblies, providing more accurate co-location information, in addition to assessing the possible impact of any MGEs. Furthermore, it resulted in an increase in the length of the longest contigs in each sample, increasing the chances of us obtaining a complete genome in one contig.

These improved hybrid assemblies enabled us to detect colocating ARGs. The MAGs identified in the hybrid assemblies, for example P. *lavamentivorans*, also highlight the issue of the kitome. This bacterium has been identified in a large number of the samples sequenced in this study, including extraction blanks. As such, it is likely that this bacterium is part of the “kitome” – the microbiome associated with sequencing and laboratory equipment, that may be amplified in the presence of little other DNA. The single contig from sample 087 was identified as *Chryseobacterium carnipullorum*. This bacterium has been previously isolated from raw chicken, from a poultry processing plant [56]. The 7 contigs from sample 053 were putatively identified as *Zymoseptoria tritici*, although further investigative blast searches suggest that this may more likely be of the genus *Ramularia*.

4.2.4 ARG-detection in assembled metagenomic sequences

The RGI software, which uses the CARD database as a reference [26], was originally developed for analysis of genomic sequences. Metagenomic data presents more challenges to any method for finding ARG sequences, since even high-depth sequence data will result in incomplete assemblies. While the most abundant genomes present may even be completely assembled, the lower abundance organisms and their genes will be represented in shorter assembled fragments, which may not always include full-length genes.

RGI has a mode for analysing such data, which we used. A result of this is that not unexpectedly, a high proportion of matches have a low coverage of the reference ARG sequence length. We therefore imposed a filter on the RGI ARG matches, discarding anywhere the matching segment was less than 80% of the length of the reference sequence. We also discarded any matches where the sequence identity of the match was less than 80%. However, the average sequence identity of the matches in all samples was in any case very high (for most of the samples it was > 97%). However, even sequences with 80% identity are expected to encode proteins with very high identity and with the same function. The CARD reference database is intensively curated with sequences supported by experimental evidence, but cannot be expected to be comprehensive. Consequently, discarding matches at 90% or 95% identity, for example, risks introducing significant false negatives. The important exceptions to this are the proteins where the resistance phenotype is conferred by a small number of mutations. However, RGI/CARD explicitly treats these cases accordingly, and does not treat them as positives by homology alone. As a result, instances of the non-resistant versions of these genes would generally not be returned as positive by RGI, and so would be unaffected by our filter.

For the analysis of the metagenomic data in comparison to qPCR results, we also applied in *silico* PCR to identify instances of two ARGs in the assembled metagenomes. One of these was essentially negative in any case, but the other identified sequences of *sul*(I) which we confirmed are near-identical to the reference *sul*(I) in CARD. We found that our filtered RGI results for this

gene was a subset (7 out of 11) of those confirmed *sul*(I) sequences. This indicates that our RGI filter is appropriate and is unlikely to be too liberal, and if anything may be on the conservative side. The likely reason for 4 of the sequences not being found by RGI is that while the primer-matching and intervening gene fragment is complete in the assemblies, this does not represent the whole gene, which could thus be present only as a fragment of < 80% of the full length.

It is important to note that the positive predictions of ARG sequences present indicate only that the matching DNA was present in the sample at the sequencing stage (potentially, they could be contaminants such as from the kitome, section 4.2.3). They do not necessarily indicate the presence of live or dead bacterial cells, nor that such gene sequences would be expressed to result in a resistance phenotype.

4.2.5 Estimating abundance of ARGs and other genes from metagenomic data

involved than simply assessing proportions of reads which match each gene sequence in the assemblies. Software methods are based on sophisticated models which take account of the process of sampling DNA fragments of a wide range of abundances. Very low abundance genes may be subject to more random sampling effects than those of higher abundance even at the shotgun sequencing level. In the context of metagenomic assemblies, they may be disproportionately absent from or rare in the assembled fragments. There is also a wide range of gene lengths, which has consequences for frequencies of the sampled reads.

Some methods originally developed for transcriptomics are now more widely used for metagenomics, in which the same principles generally apply. We used a popular method, KALLISTO [27] to estimate proportional abundances of each gene (in this context, each predicted open reading frame (ORF), which results from one step of the RGI analysis prior to assessment of the ORFs for ARG matches). For each ORF, the resulting abundance metric specifies the number of times it occurs in every million ORFs. In practice, some of these are ORF fragments rather than complete ORFs, and we also included 16S rRNA gene sequences (determined by analysis of the metagenome assemblies by qPCR) in the calculations. Due to the transcriptomics legacy, the abundance units are called 'transcripts per million' (TPM), which although a misnomer for metagenomics, serves the same purpose.

For the total TPM of all ARGs collectively, we found a wide range of relative abundances in each factory and in each meat type/non-meat control. Comparing the overall distribution of TPM values between factories, these largely overlapped, and indeed two factories have a very similar median to each other, while the other two also have a similar median. There was more difference between the TPM values of meat types, with the lowest median in pork. The chicken-associated samples had a higher median TPM, and the samples corresponding to the processing of both chicken and pork had an intermediate value. The highest median TPM was in the no-meat category (samples taken from non-meat-processing parts of the site).

Due to the compositional nature of DNA sequencing (the numbers of reads indicate proportions of DNA fragments rather than absolute quantities), the TPM values in turn indicate relative abundances and thus comparisons between samples should not be used to draw conclusions about absolute abundance. Two samples could have identical ARG relative abundances, but if the overall concentration of DNA is much higher in one sample, then that sample would have a much higher absolute abundance of ARGs.

For this reason, we attempted to use the qPCR data (estimated copies/µl) on the 16S rRNA gene abundances to calibrate the relative abundance values, for the 118 samples on which the qPCR was performed. We used *in silico* PCR to determine the 16S sequences present in the metagenome assemblies and calculated their relative abundances (TPM). This enabled the calculation of a simple normalisation factor for each sample, which is the copy number per TPM

for 16S. In theory, this scale factor should apply to all genes in the sample. We therefore normalised the total-ARG TPM values to estimate a notional number of the copies of ARGs generally, in each sample. In principle these total-ARG copy numbers (normalised abundances) can be compared between samples.

We found that the normalised abundances correlated loosely with the unnormalised relative abundances. This indicates that the scale factors do not vary hugely between samples (i.e. across many orders of magnitude), which may simply be reflective of the range of absolute bacterial abundances and the range of the proportion of bacterial genes which are ARGs not varying to a huge degree.

Comparing the factories in terms of the normalised total-ARG values, the distributions are generally more overlapping than the unnormalised equivalents. The median of one factory is however considerably higher than the factory to which it is most similar for unnormalised abundances. This means that there may be a very uneven distribution of the quantity of arg-containing DNA between factories, which may in turn mean that there is a very uneven-distribution of the quantity arg-containing bacteria. However, because we only looked at four factories, we can't draw any general inferences from these observations. But we have demonstrated that there are some different methods that we can apply to try to get quantitative information about ARG presence.

The effect on the comparison of meat types is similar, with the distributions being more similar, than with the unnormalised data. The interquartile range of pork now overlaps with that of chicken, chicken/pork and the non-meat samples, which was not the case for the unnormalised values. The non-meat type still has the highest median.

4.2.6 qPCR

We applied qPCR to estimate number of copies/µl of two specific ARGs and of the 16S rRNA gene generally, in a subset of the samples for which sufficient DNA was available. We obtained positives at generally high copy numbers for 16S in all cases, albeit one sample may have failed.

For the *sul*(I) gene, we obtained copy numbers above the threshold of 500 in only 8 samples. This threshold represents the lower limit at which the results should be treated as quantitative rather than qualitative (limit of quantification, LOQ). We also noticed that the *sul*(I) copy numbers broadly correlate with those of 16S and speculate that the;assay generates low-abundance, non-specific amplicon products which may have been detected. We therefore view the eight samples which exceeded the LOQ as possibly the only genuine cases.

Two of those eight, and one other slightly below the LOQ, correspond to samples positive for an *in silico sul*(I) amplicon. It is unclear why the other 6 higher copy number *sul*(I) samples were negative by the *in silico* analysis. Possibly, other sequences are present with more than the permitted three primer base mismatches. Lack of metagenomic sequencing depth may be more likely; the sample with the highest copy number had one of the lowest DNA read counts. Potentially, some samples may have reasonable read counts but a high proportion of non-bacterial DNA.

Notwithstanding the above observations, we also found that the relationship between the calculated relative abundances of the *sul*(I) amplicon sequences to the qPCR copy numbers, was not consistent with the same relationship for 16S. In theory, the number of copies per TPM should be similar for all genes. However, given the very small number of samples involved, with only three where qPCR and TPM could be compared, it is difficult to draw firm conclusions from this inconsistency.

The *tet*(*B*) qPCR assay produced positives in 11 samples, all well below the LOQ. The metagenomic sequence data was generally consistent with this, in fact being negative for all samples. The single *in silico* PCR amplicon sequence appeared to be an erroneous portion of an assembly, despite the flanking primer-matches. The RGI/CARD analysis produced only poor coverage matches in a few samples, with none passing the filter.

4.2.7 Uncertainties associated with UK dietary burden

As stated in Section 3.6.5, the overall burden of AMR due to biofilms in meat processing plants could not be estimated with an adequate degree of certainty based on the available data. The main limitations are:

- limited sample sizes and processing plants represented. Effectively, we assume that for each food type the sampled processing plants and measured samples are representative of all processing types, and the influence of biofilms is equal across all other processing plants from which the UK population's meat is obtained. UK consumption will include meat from various UK and overseas sources.
- some commonly consumed processed meat (for example, burgers and beef mince) are not as well represented as chicken and pork, so we may miss some ARGs as a result.
- linking of processing lines and other biofilm sample locations to food types is poorly characterised in the data. Modelling burden using these data would require simple assumptions with an unknown degree of conservatism. For example, in a plant that processes multiple food types, if a biofilm sample is taken from 'bowl chopper 3 – floor' would we assume the measured ARGs would also be present in all the food types processed in that plant?
- many consumed meat items do not directly link to the processed food types represented in the samples. A roast chicken may or may not be 'value-added' and chicken curry may be produced with chicken pieces or diced chicken. Again, a model calculation of burden could make simple assumptions and subjective judgments regarding the type of meat used in individual consumed items. The level of conservatism is unknown.
- the effects of cooking on the burden are not accounted for.

A standard approach to dealing with uncertainty when there is insufficient data to produce a quantitative measure is to assume a worst-case scenario for all sources of uncertainty. Given the issues listed above, this would lead to unrealistic estimates of the proportion of meat consumers exposed to the ARGs found in the samples (in some cases likely to be close to 100%). The summary of ARGs in the samples themselves is a more useful measure, but we must also consider the difference between the presence of an ARG and the AMR risk in a cooked product as consumed.