J Antimicrob Chemother 2021; **76**: 101–109
doi:10.1093/jac/dkaa390 Advance Access publication 3 October 2020

Journal of Antimicrobial Chemotherapy

Detection of mobile genetic elements associated with antibiotic resistance in *Salmonella enterica* using a newly developed web tool: MobileElementFinder

Markus H. K. Johansson [1]*, Valeria Bortolaia[1], Supathep Tansirichaiya[2], Frank M. Aarestrup [1], Adam P. Roberts[3] and Thomas N. Petersen[1]

[1]*National Food Institute, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark;* [2]*Department of Clinical Dentistry, Faculty of Health Sciences, UiT the Arctic University of Norway, Tromsø, Norway;* [3]*Department of Tropical Disease Biology, Liverpool School of Tropical Medicine, Liverpool, UK*

*Corresponding author. E-mail: markjo@food.dtu.dk

Received 20 May 2020; accepted 19 August 2020

Objectives: Antimicrobial resistance (AMR) in clinically relevant bacteria is a growing threat to public health globally. In these bacteria, antimicrobial resistance genes are often associated with mobile genetic elements (MGEs), which promote their mobility, enabling them to rapidly spread throughout a bacterial community.

Methods: The tool MobileElementFinder was developed to enable rapid detection of MGEs and their genetic context in assembled sequence data. MGEs are detected based on sequence similarity to a database of 4452 known elements augmented with annotation of resistance genes, virulence factors and detection of plasmids.

Results: MobileElementFinder was applied to analyse the mobilome of 1725 sequenced *Salmonella enterica* isolates of animal origin from Denmark, Germany and the USA. We found that the MGEs were seemingly conserved according to multilocus ST and not restricted to either the host or the country of origin. Moreover, we identified putative translocatable units for specific aminoglycoside, sulphonamide and tetracycline genes. Several putative composite transposons were predicted that could mobilize, among others, AMR, metal resistance and phosphodiesterase genes associated with macrophage survivability. This is, to our knowledge, the first time the phosphodiesterase-like *pdeL* has been found to be potentially mobilized into *S. enterica*.

Conclusions: MobileElementFinder is a powerful tool to study the epidemiology of MGEs in a large number of genome sequences and to determine the potential for genomic plasticity of bacteria. This web service provides a convenient method of detecting MGEs in assembled sequence data. MobileElementFinder can be accessed at https://cge.cbs.dtu.dk/services/MobileElementFinder/.

Introduction

Antimicrobial resistance (AMR) is considered one of the biggest threats to human health.[1] Bacteria can acquire AMR either though mutations in the genome or through horizontal gene transfer (HGT) where HGT of AMR usually involves mobile genetic elements (MGEs).[2,3]

MGEs are discrete regions of DNA defined by their ability to move within and/or between bacterial cells. They are categorized into types based on their properties and their genetic layout.[4] Elements capable of integrating into the host DNA are referred to here as integrating MGEs (iMGEs).

Insertion sequences (ISs) are among the smallest types of iMGEs. They are often composed of a transposase gene flanked by two inverted repeats (IRs). They are notable for their ability to modulate gene expression and promote mobility by forming composite transposons (ComTns), translocatable units (TUs) and in the case of elements from the IS26 family pseudo-composite transposons (PCTs).[5] ComTns are formed when a transposase accidentally acts on the IR of a related MGE nearby and transposes the two elements with the intermediary region.[3,6] TUs are formed when one of the ISs in a ComTn is excised with adjacent DNA as a circular molecule.[7]

Unit transposons (Tns) are generally flanked by IRs and carry a transposase gene. They usually carry a resolvase gene, accessory genes and/or additional iMGEs.[3] Miniature Inverted Repeats (MITEs) are non-autonomous ISs or Tns that have undergone

© The Author(s) 2020. Published by Oxford University Press on behalf of the British Society for Antimicrobial Chemotherapy.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

deletions in their core genes but have retained the IR and can form ComTn-like structures.[3]

Integrative Conjugative Elements (ICEs), Cis-Mobilizable Elements (CIMEs) and Integrative Mobilizable Elements (IMEs) are larger iMGEs capable of conjugation. They can either conjugate independently or be co-mobilized by conjugation of other elements. These elements carry many accessory genes and other MGEs.[3,8–11]

MGEs interact with one another to form a complex network with the potential to recruit and disseminate genes throughout a bacterial population. Through this network, MGEs play a pivotal role in the spread of AMR. The ability to identify and characterize MGEs is crucial to elucidate AMR epidemiology.[3,12]

The rapid development of next-generation sequencing has made genomic analyses more available. A bottleneck has been the limited availability of user-friendly analysis tools. Many MGE detection tools require bioinformatics expertise to operate and/or are limited to a specific MGE type.[13–18]

Here we describe MobileElementFinder, a new user-friendly webserver that detects iMGEs in assembled sequence data and annotates their relationship to AMR, virulence genes and plasmids. The tool was applied to describe iMGEs and their association with AMR in 1725 zoonotic *Salmonella enterica* isolates. *S. enterica* is a Gram-negative human and animal pathogen that is commonly transmitted to humans through consumption of contaminated food. It is the leading cause of bacterial foodborne disease where the increased prevalence of MDR causes higher mortality and increased cost of treatment.[19,20] By studying the dynamics of the mobilome and its interaction with AMR, the importance of MGEs can be investigated in greater detail.

Methods

Development of MobileElementFinder

MobileElementFinder was written in Python v3.7 and can be installed from PyPi or accessed as a webserver. The tool includes a database of known MGEs built from public nomenclature and data repositories.[13,15,21] MobileElementFinder can detect the following types of MGEs: MITEs, ISs, ComTns, Tns, ICEs, IMEs and CIMEs.

Details of the tool development are described in the Supplementary data available at *JAC* Online.

Dataset selection

A dataset consisting of whole-genome-sequenced *S. enterica* isolates was generated from publicly available food surveillance data.

National Antimicrobial Resistance Monitoring System (NARMS) dataset

NARMS is a US domestic national surveillance programme for AMR resistance. For this study, a subset of the *S. enterica* caecal samples from pigs and broilers (study accession: SRP063697, SRP062916) was selected using the following criteria: (i) collected by US Department of Agriculture (USDA) between the years 2015 and 2019; (ii) whole-genome, paired-end shotgun sequenced on an Illumina platform; (iii) base count greater than 10× median genome size of all assembled *S. enterica* spp. on NCBI (4.81 Mb).[22]

From this, a subset of samples was selected by binning them on source, submission date and the state from which they were collected. Up to 10 samples were randomly selected, without duplicates, from each bin that contained more than 5 samples, leading to all samples being included from bins containing fewer than 11 samples. The final dataset contained 1543 isolates for which raw FASTQ files were downloaded in October 2019.

COMPARE dataset

A dataset consisting of *S. enterica* Typhimurium isolates originating from human and various different meat and environmental sources were collected from Denmark, France, Germany and the UK as a part of the COMPARE project.[23] The data originated from various surveillance programmes or larger studies conducted between 2010 and 2016.[10]

Isolates from Danish and German pork and chicken meat that fulfilled the previously described quality criteria (ii) and (iii) were used in this study (191 isolates in total). Isolates are denoted as originating from pig and chicken regardless of the exact meat product.

Read processing and assembly

Raw reads were trimmed with bbduk2 (part of BBmap v36.49), using score cut-off = 20 and removing reads shorter than 50 bp. Adapters were removed with bbduk2 by matching to an internal database.[24] Sequence quality was evaluated with FastQC v0.11.5 before and after quality processing. Trimmed reads were assembled with Spades v3.13.0 using error correction, coverage cut-off = 2 and the kmer sizes 21, 33, 55, 77, 99 and 127. Contigs shorter than 500 bases were discarded.[25,26] The quality of the *de novo* assembled contigs was assessed using Quast (v4.5).[27]

In silico prediction of AMR, MGEs and epidemiological typing

AMR genes were predicted using ResFinder and overlapping genes were filtered out keeping the gene with the highest coverage and sequence identity.[28] Plasmids were predicted using PlasmidFinder.[16] MLST was done using MLSTFinder with the *Salmonella enterica* PubMLST database.[29,30] See Table S1 for versions of tools and databases used.

Estimating clonality of samples

The diversity within different sets of isolates was estimated using their core-genome MLST profile, determined with cgMLSTFinder with the Enterobase scheme (Table S1).[31,32] The average pairwise core-genome allele difference between samples was used to estimate the diversity within given subsets of data depending on the application. If the average allele difference was equal to or lower than seven the selection was considered as clonal.[33,34]

Characterization of MGEs in S. enterica

iMGEs were predicted using MobileElementFinder (v1.02) using the method and thresholds described in the Supplementary data. The distribution of iMGEs throughout the *S. enterica* population was determined by clustering the samples on the predicted MGE profile, considering MGEs as either present or absent. Putative ComTns were not included in the MGE profile to avoid introducing bias from false-positive or false-negative predictions. Clustering was performed using the R package vegan using Jaccard distance and complete linkage.[35] The result was visualized using iTOL (v4) overlying country, ST (for STs occurring more than 20 times) and meat source.[36]

Additional accessory genes carried on detected ComTns, Tns, IMEs and predicted putative ComTns were predicted using Prokka v1.14.6 with the default parameters.[37]

Classification of mobile elements associated with AMR

Each resistance gene was classified as either being iMGE-associated, carried by an MGE or having an unknown association. The AMR was considered

associated if it was located within 31 kb of an iMGE. The threshold corresponds to the longest ComTn (Tn*6108*) from *S. enterica* in the database and is intended to reflect which genes have the potential to be mobilized by surrounding iMGEs.

The iMGE-associated AMR genes were grouped on MGE type and distance to the closest MGE. Groups with 10 or more members were investigated further as they could be putative TUs. The level of conservation of the sequence spanning between the iMGE and the associated AMR gene was estimated by calculating the average nucleotide identity (ANI) with FastANI (v1.3).[38] Translocatability was indicated by a particular MGE and AMR gene combination being located on multiple different plasmids across several unrelated isolates.

Integrons located in association with these putative TUs were detected using Integron Finder v2-2020-04-28 with the local-max option.[39]

Results

Characterization of MGEs

A dataset consisting of 1725 whole-genome-sequenced zoonotic *S. enterica* isolates from three countries was collected from public sources. The average isolate had 80.4× read coverage (range: 20.4–417.4) (Figure S1a). Isolates were *de novo* assembled, resulting in an average N50 of 308 kb (range: 14.7–2460 kb) (Figure S1b) and averaging a total assembly size of 4.85 Mb (range: 4.51– 5.29 Mb) (Figure S1c).

In total, MobileElementFinder predicted 12 056 iMGEs, of which the majority were either ISs (36.5%) or MITEs (62.6%), as shown in Table S2; IS*3* constituted ~40% (1662) of all predicted ISs. At the isolate level, there were on average more ISs (3.65 per genome) in chicken-origin isolates than in pig-origin ones (1.97 per genome) (Figure 1). The prevalence of ISs was highly variable within the dataset. Five IS families constituted ~80% of all detected ISs.

A total of 65 Tns were detected, of which the majority were located in American isolates, the exception being Tn*2*, which was also found in Danish chicken isolates (Table S2). Of the 65 detected Tns, 19 were predicted to be located on plasmids. Tn*2* was the most common element, identified on IncI1-Iγ and IncN plasmids. Tn*6024* and Tn*6196* were only found on IncHI2A-IncHI2 plasmids (Figure 2). The majority of the detected Tns are predicted to carry AMR or metal resistance genes (Table 1).

Difference in MGEs between MLST types

The number of iMGEs per isolate varied depending on the MLST, e.g. STs 32 and 96 contained the highest variation in MGE abundance (SD: 1.63 and 1.4, respectively) and STs 34, 64 and 11 the lowest variation but with several extreme values (Figure 3a). The differences in MGE abundance were considered accurate due to the large sample size (Figure 3b). All of the 12 included STs were considered to be constituted by diverse samples since the average allele distance per ST was much larger than the clonality threshold of seven alleles (Figure 3c).

The impact of source and country on the distribution of iMGEs was analysed by clustering the isolates on their MGE profile and comparing clustering formation with the overlaying metadata (Figure 4). The formation of clusters corresponded well with ST and isolate source, thus indicating that samples with the same ST

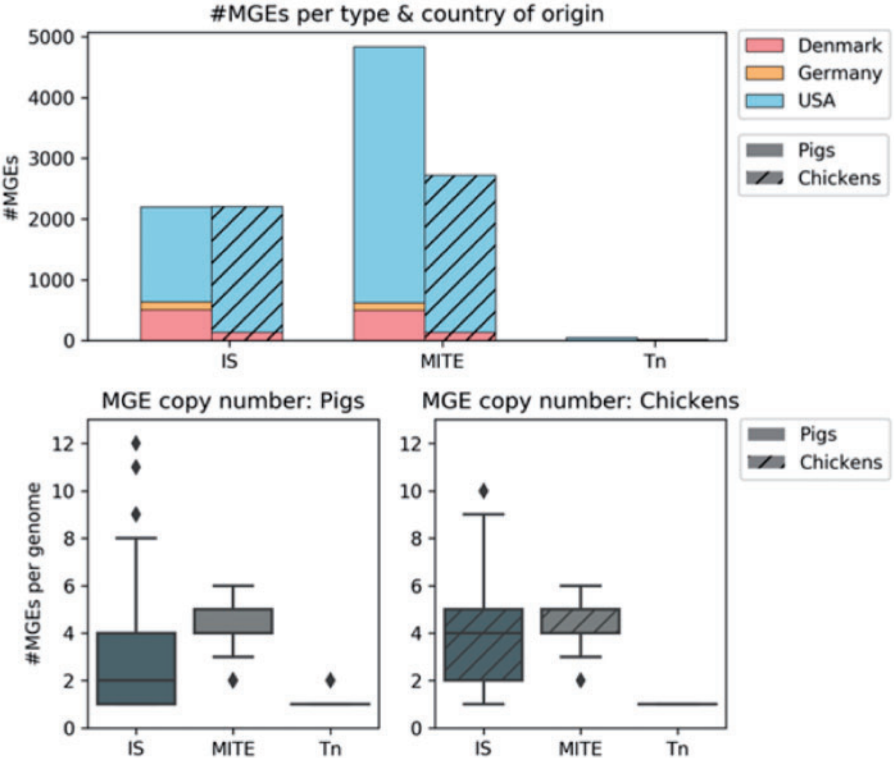

Figure 1. Total number of predicted MGEs and abundance of predicted MGEs per source and country of origin. Putative ComTns are not separated from ComTns.

tended to carry a similar MGE profile and samples of a given ST tended to originate from a given meat source. Some STs (for instance, STs 40, 471 and 96) tended to carry a highly diverse set of MGEs that were often more similar to other STs.

Associations of MGEs and antibiotic resistance

AMR genes either carried by iMGEs or located within 31 kb were classified as being associated with iMGEs. The total number of iMGEs associated with AMR genes was greater in isolates originating from chickens than in isolates from pigs. This was especially prominent for β-lactam resistance genes where 63.1% of the resistance genes in isolates from chickens were located near iMGEs, compared with 21.6% in the isolates from pigs. Aminoglycoside resistance genes tended to be more frequently associated with iMGEs in isolates originating from chickens (55.3%) than in those from pigs (24.4%). Tetracycline resistance genes were more frequently associated with MGEs in pig isolates (35.3%) than in chicken isolates (10.8%) (Figure S2).

We analysed whether the invariable iMGE, distance and AMR gene combinations could be explained by plasmid mobilization or by the isolates being clonal. The association with plasmid replicons was calculated for each combination observed at least four times. Most combinations were found on contigs with different plasmid replicons (Figure S3). The pairwise CG allele differences of samples carrying a given combination indicated that the samples were unrelated to one another (Figure S4). These invariable units will hereon be referred to as putative TUs.

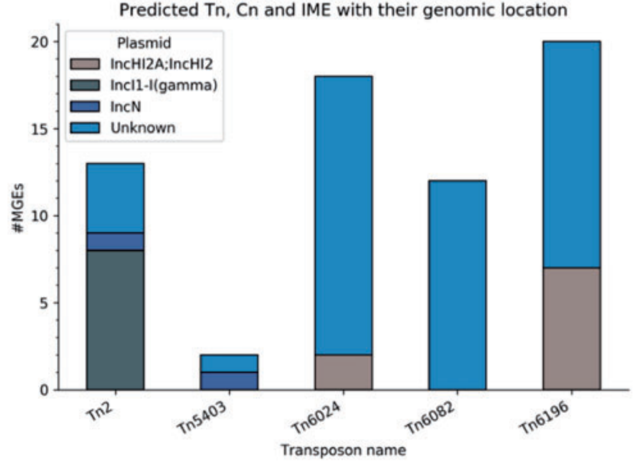

Figure 2. The number of iMGEs capable of carrying passenger genes and their genomic location.

Five of the putative TUs were associated with an array of AMR genes located on the same contig. The arrays contained a variable number of genes where isolates with shorter arrays often carried the same AMR genes as isolates with the longer arrays, but on different contigs (data not shown). The exception was IS*Ec59*; 368, where on one occasion an integron carrying *dfrA14* was inserted before *aph(3″)-Ia* (Figure 5). This integron was not predicted in isolates with longer versions of the array.

Putative ComTns were identified based the presence of ISs and the distance between them. In total, 38 putative ComTns were predicted in 38 different isolates in which the DNA was often mobilized by IS*Ec13* (11 times) and IS*26* (10 times). The putative ComTns varied in length, including ones with identical flanking sequences, where IS*Ecl10*-based sequences were considerably larger than other elements. Sequences with the same flanking iMGE tended to carry a similar set of genes and share synteny, indicating that they were variants of the same element and might originate from the same genomic context. Of 12 different putative ComTns, 3 were carrying either tetracycline [*tet*(B)] or aminoglycoside [*aph(3′)-Ia*] resistance genes. IS*26*-based elements were predicted to carry a mercury resistance gene and the longer IS*903*-based element carried genes related to arsenic resistance. Several putative ComTns carried toxin (*ccdB*) and/or antitoxin genes (*ccdA*, *yfjZ*, *higA1*) and some elements carried *pdeL*, which couples expression of other genes to cyclic di-guanylate monophosphate (c-di-GMP) (Table 2).

Discussion

MobileElementFinder was developed to be a user-friendly online tool to enable non-bioinformatically trained researchers to study MGEs. iMGEs are detected based on sequence similarity to sequences of known MGEs. Using this approach, the accuracy and completeness of the database are important factors for the tool's performance. To ensure accuracy, the database was built using information from well-annotated MGE and nomenclature databases. Partial MGEs were excluded since it would not be possible to assure the presence of the entire MGE. The final database consists of ∼4450 MGE sequences that originate from ∼1050 different species, which allows the tool to detect MGEs in many bacterial species including ones of clinical importance, as presented here.

MobileElementFinder was designed to detect iMGEs in assembled sequences as this allows study of the genetic context and association with nearby genes. This information can be used

Table 1. Number of detected iMGEs with the accessory genes they are predicted to carry. AMR and metal resistance genes are displayed in separate columns

Name	No. of MGEs	Type	AMR genes	Metal resistance genes	Additional genes
Tn2	13	Tn	*bla*$_{TEM-1B}$; *bla*$_{TEM-1C}$	—	—
Tn*5403*	2	Tn	—	—	*pinR*
Tn*6024*	18	Tn	—	*silE*; *copA*; *copB*; *copD*; *copR*	—
Tn*6082*	12	Tn	—	—	—
Tn*6196*	20	Tn	—	—	*hin*

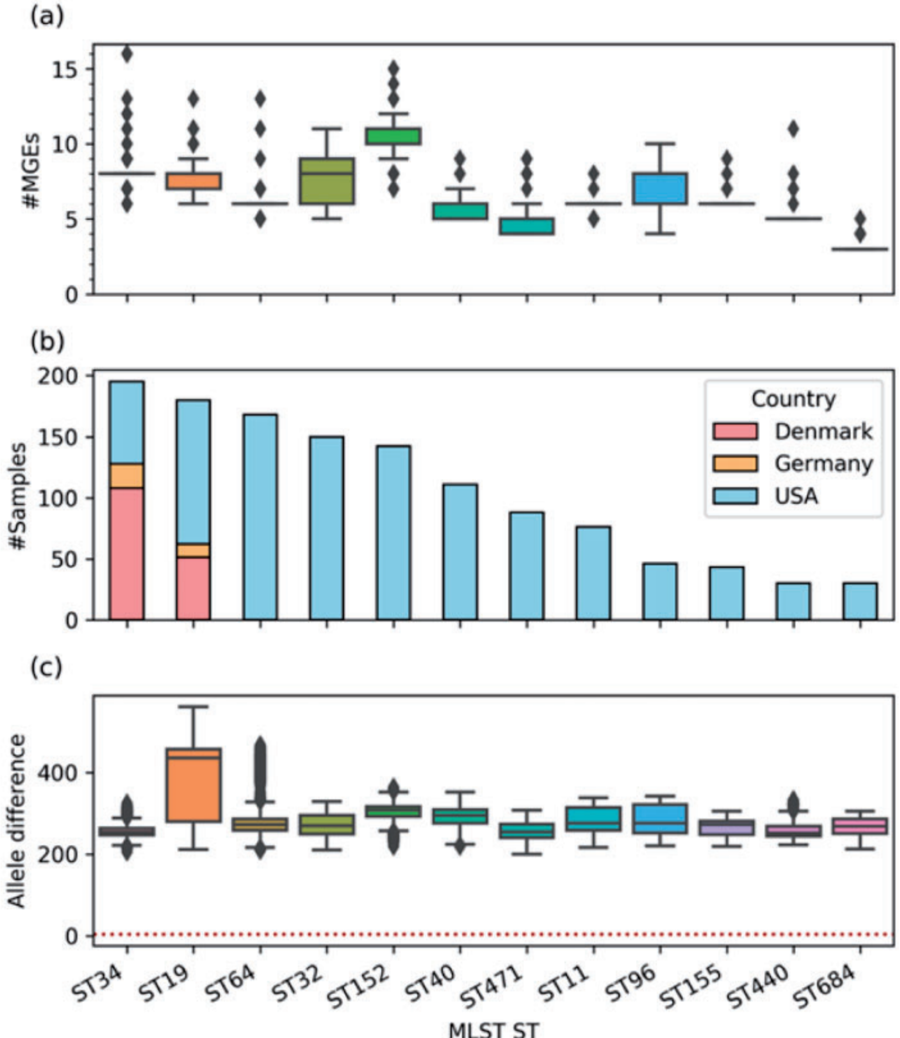

Figure 3. (a) Distributions of the number of predicted MGEs for samples of the 12 most common MLST STs. (b) The number of samples per country with a given MLST ST. (c) Distribution of pairwise core-genome allele differences for the 12 most common MLST STs. The dotted line indicates the threshold used to separate clonal *S. enterica*. MLST ST19 contained the most diverse isolates, with an average pairwise distance of 403 alleles.

to infer potential mobilization and regulatory aspects of nearby genes. In addition, the tool was designed to be user friendly and to be easy to integrate into analysis pipelines, thus enabling more researchers to routinely account for iMGEs in their analysis.

Characterization of MGEs in S. enterica

MobileElementFinder was used to analyse the mobilome of zoonotic *S. enterica* from pig and chicken meat. In the 1725 analysed isolates, on average, eight iMGEs were detected per isolate, where the smaller MITEs and ISs were more abundant than Tns. While the differences in abundance between types was expected, it was unexpected to find that MGEs rarely existed in more than one copy. This was especially the case for elements similar to IS*26*, which are known to occur in arrays of repeated iMGEs in Gram-negative bacteria.[7] This discrepancy might be due to difficulty assembling repeated sequences, which results in repeated elements being merged into a single copy.[40] Another discrepancy was that no *Salmonella* Genomic Islands (SGIs) were identified. SGIs are a group of IMEs that are common in several *S. enterica* serovars and carry antimicrobial or heavy metal resistance genes.[41,42] There were several alignments to SGI reference sequences but fragmented over many contigs despite the good assembly quality (Figure S1b). This shows that the current prediction algorithm has limitations in predicting conjugative Tns from short-read assemblies. This could be mitigated by making hybrid assemblies with long-read sequences.

There was a clear difference in the number and variability of iMGEs carried by different strains of *S. enterica*, where ST152 isolates carried more MGEs than other isolates and ST684 isolates carried fewer (Figure 3a). The samples were controlled for clonality and these results are therefore unlikely to be an artefact of a homogeneous dataset. Similar strain-dependent differences in the abundance of ISs have previously been observed in *Acinetobacter baumannii* and *Klebsiella pneumoniae*.[43]

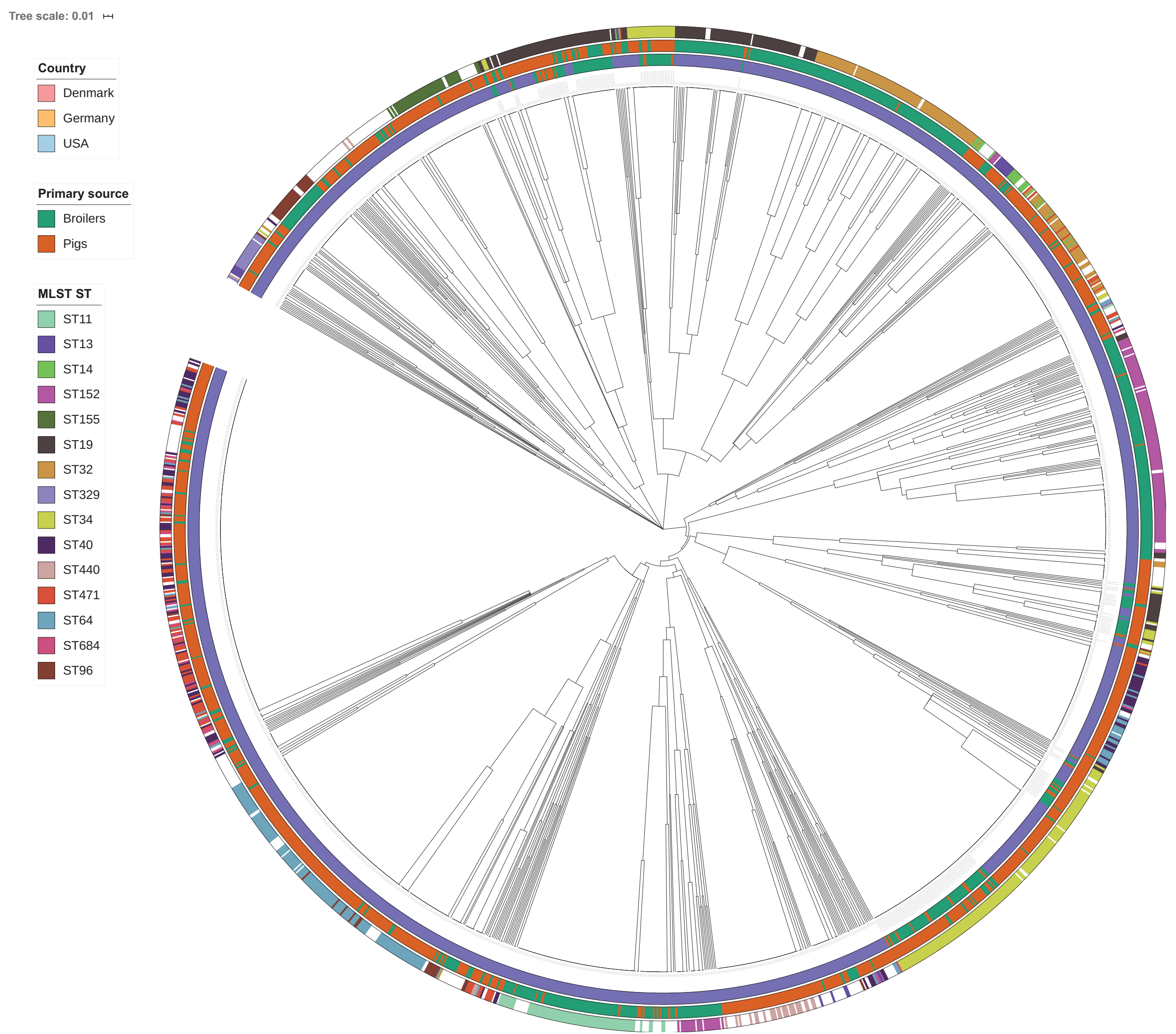

Figure 4. Samples clustered on the predicted MGE profile using Jaccard distance overlaying meat source, predicted MLST ST and country as overlay. MLST STs occurring less than 12 times were excluded.

It also appears that isolates from the same ST tended to carry similar MGE profiles (Figure 4) and that some MGEs were only detected in specific genetic backgrounds. There was still considerable variation in how well the clusters corresponded to the ST and the diversity within each cluster, for instance ST34, ST11 and ST64 had low iMGE diversity whereas ST40 and ST96 were diverse. It is thus unlikely that this was caused by biases in the isolate selection since the STs were represented by isolates unrelated to one another. iMGE diversity did not correspond to either host specificity or prevalence, as ST34, ST11 and ST40 belonged to prevalent serovars (Typhimurium, Enteritidis and Derby) with broad host range.[20,44] Instead the diversity might reflect differences in genome plasticity since ST96 isolates are from a lineage with higher homologous recombination frequency than ST11.[45] However, further research is needed to fully explain these patterns.

Of the 38 putative ComTns predicted by MobileElementFinder, 14 were predicted to carry AMR genes, or genes associated with metal resistance. Some elements carried type II or type IV antitoxin genes, which are often located on larger MGEs such as plasmids or ICEs; this suggests that these putative ComTns are located on larger elements.[3,9] Some putative ComTns were predicted to carry the phosphodiesterase gene *pdeL*, which, in *Escherichia coli*, acts as a sensor that up-regulates transcription of associated genes in cells with low c-di-GMP concentrations and in *S.* Typhimurium is important for macrophage survivability and virulence.[46,47] This is, to the best of our knowledge, the first time this gene has been described as being potentially mobilized. By mobilizing this gene, the bacteria might have greater flexibility to respond to external stresses. However further research is required to prove actual mobility of these elements.

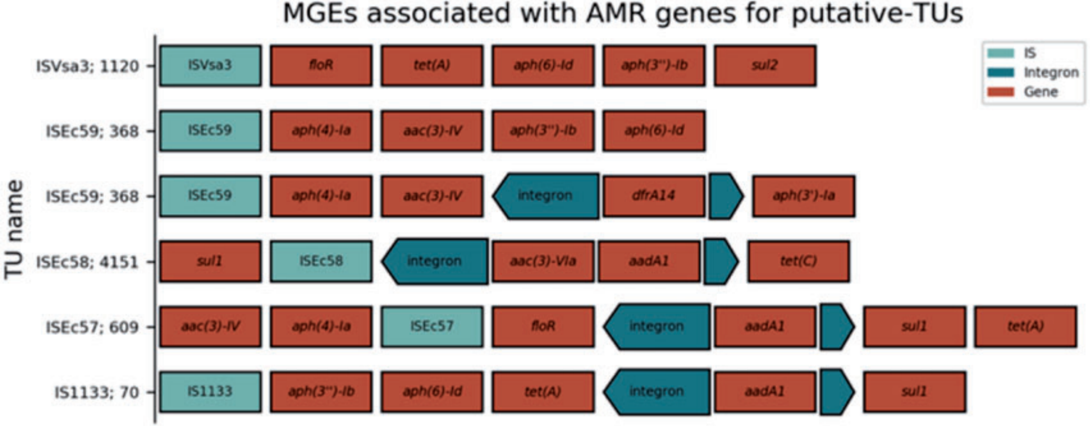

Figure 5. The association of AMR genes with MGEs for five putative TUs. Genes carried on MGEs are located in the relevant element. The figure represents the synteny, not orientation and scale of elements.

Table 2. Number of detected putative ComTns and their predicted accessory genes. AMR and metal resistance genes are displayed in separate columns

Name	No. of MGEs	AMR genes	Metal resistance genes	Additional genes
cn_19285_ISEc13	11			*pdeL*; *higA1*
cn_19309_ISEc13	2			*pdeL*; *higA1*
cn_15656_ISEc13	1			*ccdA*; *ccdb*; *tsr*
cn_20262_IS26	10	*tet*(B)	*merT*; *merP*; *merC*; *merA*	*gltS*; *dcm*; *tet*(A); *tet*(C); *tet*(R)
cn_20097_IS26	1	*tet*(B)	*merT*; *merP*; *merC*; *merA*	*gltS*; *dcm*; *tet*(A); *tet*(C); *tet*(R)
cn_17717_IS26	1		*merT*; *merP*; *merC*; *merA*	*gltS*; *dcm*
cn_4114_IS102	5			
cn_10380_IS102	2			*tet*(R); *tet*(A)
cn_45790_ISEcl10	2			*yidZ*; *ligB*; *gmk*; *ropZ*; *spoT*; *trmH*; *recG*; *gltS*; *xanP*; *yicI*; *yicJ*; *intS*; *yfjZ*; *recF*
cn_33512_ISEcl10	1			*yidZ*; *ligB*; *gmk*; *ropZ*; *spoT*; *trmH*; *recG*; *gltS*; *xanP*; *yicI*; *yicJ*; *intS*
cn_11943_IS903	1		*arsC*; *arsB*; *arsH*	
cn_3064_IS903	1	*aph(3′)-Ia*		

Association of MGEs and AMR genes

iMGEs were classified as being associated with AMR genes depending on their relative location. Most AMR genes were not carried by Tns but instead located in proximity to one or more iMGEs, usually ISs (Figure S2). These iMGEs tended to be located at specific distances from the gene, with the exception of IS*26*, which was found to have a higher variability (Figure S5). This likely reflects its tendency to form IS arrays, which have been important for disseminating AMR genes in Gram-negative bacteria.[3] IS*Ec9* and IS*102* were always located in proximity to resistance genes; IS*Ec9* was located 118 nt upstream of $bla_{CMY\text{-}2}$ and IS*102* was 4 nt downstream of $bla_{CTX\text{-}M\text{-}65}$ and 1432 nt from *tet*(B). Both are known to form ComTns and it is likely that these ISs have mobilized these AMR genes.[48]

Several conserved AMR–MGE combinations could be TUs as they were located at fixed distances from one another and existed across unrelated isolates and located on different plasmids (Figure S3, Figure S5). Many combinations consisted of IS*26*, IS*Ec59* and IS*Ec58*, which are either known, or related to, elements capable of forming ComTns, although not always in *S. enterica*.[3,49]

Some of these putative TUs were associated with an array of AMR genes, some of which were carried by integrons (Figure 5). While most of the variability in the number of AMR genes in a given array was likely caused by the variation in contig length, there were some arrays exhibiting patterns that could be explained by intergenome mobility. For example, the array IS*Ec59*; 368 where one version had acquired an integron carrying *dfrA14* that was not found in the other isolates. However, determining whether the AMR genes are located on a TU requires experimental verification.[50]

Conclusions

We have presented MobileElementFinder, a novel tool for detecting iMGEs including MITEs, ISs, ComTns, PCTs, Tns and conjugative Tns (ICEs, IMEs and CIMEs) from assembled sequence data.

MobileElementFinder is available both as a webservice and as installable CLI software that could be integrated into existing analysis pipelines.

The tool was used to characterize iMGEs and their association with AMR genes in zoonotic *S. enterica*. We found a considerable diversity in the number and combination of MGEs, which was primarily dependent on ST rather than isolate source and geographic origin. In addition, several putative ComTns and TUs were identified that are likely to mobilize, among others, AMR and metal resistance genes. Additionally, to our knowledge, we have described for the first time a putative ComTn carrying a phosphodiesterase in *S. enterica*.

Using MobileElementFinder simplifies detection and characterization of MGEs and their relationship to other genes for large datasets, bringing a deeper understanding of the plasticity of the bacterial pan-genome.

Limitations and future prospects

The sensitivity for detecting large iMGEs, such as conjugative Tns, is limited by the quality of the assembled contigs as they tend to be fragmented into multiple short contigs. Long-read sequencing and hybrid assembly offers the potential to resolve these issues.

The ability of MobileElementFinder to predict novel iMGEs is likely limited to those elements homologous to ones in the database. As predictions are based on alignments with previously known MGEs, susceptibility prediction is dependent on the database, which will be updated when needed.

Acknowledgements

This publication made use of the cgMLST.org website (http://www.cgMLST.org/) that is developed and maintained by Ridom GmbH.

Funding

This work was supported by The Novo Nordisk Foundation (NNF16OC0021856) - Global Surveillance of Antimicrobial Resistance.

Transparency declarations

None to declare.

Supplementary data

Tables S1 and S2 and Figures S1 to S5 and additional information on the development of MobileElementFinder are available as Supplementary data at *JAC* Online

References

1 WHO. Antimicrobial Resistance: Global Report on Surveillance. https://apps.who.int/iris/bitstream/10665/112642/1/9789241564748_eng.pdf.

2 Peterson E, Kaur P. Antibiotic resistance mechanisms in bacteria: relationships between resistance determinants of antibiotic producers, environmental bacteria, and clinical pathogens. *Front Microbiol* 2018; **9**: 2928.

3 Partridge SR, Kwong SM, Firth N *et al.* Mobile genetic elements associated with antimicrobial resistance. *Clin Microbiol Rev* 2018; **31**: e00088–17.

4 van Hoek AHAM, Mevius D, Guerra B *et al.* Acquired antibiotic resistance genes: an overview. *Front Microbiol* 2011; **2**: 203.

5 Harmer C J, Pong C H, Hall R M. Structures bounded by directly-oriented members of the IS26 family are pseudo-compound transposons. *Plasmid* 2020; **111**: 102530.

6 Vandecraen J, Chandler M, Aertsen A *et al.* The impact of insertion sequences on bacterial genome plasticity and adaptability. *Crit Rev Microbiol* 2017; **43**: 709–30.

7 Harmer CJ, Moran RA, Hall RM. Movement of IS*26*-associated antibiotic resistance genes occurs via a translocatable unit that includes a single IS*26* and preferentially inserts adjacent to another IS*26*. *mBio* 2014; **5**: e01801–14.

8 Wozniak RAF, Waldor MK. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat Rev Microbiol* 2010; **8**: 552–63.

9 Carraro N, Burrus V. The dualistic nature of integrative and conjugative elements. *Mob Genet Elements* 2015; **5**: 98–102.

10 Wright LD, Grossman AD. Autonomous replication of the conjugative transposon Tn*916*. *J Bacteriol* 2016; **198**: 3355–66.

11 Delavat F, Moritz R, van der Meer JR. Transient replication in specialized cells favors transfer of an integrative and conjugative element. *mBio* 2019; **10**: e01133–19.

12 Frost LS, Leplae R, Summers AO *et al.* Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol* 2005; **3**: 722–32.

13 Liu M, Li X, Xie Y *et al.* ICEberg 2.0: an updated database of bacterial integrative and conjugative elements. *Nucleic Acids Res* 2019; **47**: D660–5.

14 Varani AM, Siguier P, Gourbeyre E *et al.* ISsaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. *Genome Biol* 2011; **12**: R30.

15 Siguier P, Perochon J, Lestrade L *et al.* ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* 2006; **34**: D32–6.

16 Carattoli A, Zankari E, García-Fernández A *et al.* In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 2014; **58**: 3895–903.

17 Sheppard AE, Stoesser N, German-Mesner I *et al.* TETyper: a bioinformatic pipeline for classifying variation and genetic contexts of transposable elements from short-read whole-genome sequencing data. *Microb Genom* 2018; **4**: e000232.

18 Durrant MG, Li MM, Siranosian BA *et al.* A bioinformatic analysis of integrative mobile genetic elements highlights their role in bacterial adaptation. *Cell Host Microbe* 2020; **27**: 140–53.e9.

19 Chen S, Zhao S, White DG *et al.* Characterization of multiple-antimicrobial-resistant *Salmonella* serovars isolated from retail meats. *Appl Environ Microbiol* 2004; **70**: 1–7.

20 Jajere SM. A review of *Salmonella enterica* with particular focus on the pathogenicity and virulence factors, host specificity and antimicrobial resistance including multidrug resistance. *Vet World* 2019; **12**: 504–21.

21 Tansirichaiya S, Rahman MA, Roberts AP. The Transposon Registry. *Mob DNA* 2019; **10**: 40.

22 Karp BE, Tate H, Plumblee JR *et al.* National antimicrobial resistance monitoring system: two decades of advancing public health through integrated surveillance of antimicrobial resistance. *Foodborne Pathog Dis* 2017; **14**: 545–57.

23 Munck N, Leekitcharoenphon P, Litrup E *et al.* Four European *Salmonella* Typhimurium datasets collected to develop WGS-based source attribution methods. *Sci Data* 2020; **7**: 75.

24 Center for Genomic Epidemiology. FoodQCPipeline. https://bitbucket.org/genomicepidemiology/foodqcpipeline/src/master/.

25 Andrews S. FastQC: a Quality Control Tool for High Throughput Sequence Data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

26 Bankevich A, Nurk S, Antipov D *et al*. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012; **19**: 455–77.

27 Gurevich A, Saveliev V, Vyahhi N *et al*. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013; **29**: 1072–5.

28 Zankari E, Hasman H, Cosentino S *et al*. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 2012; **67**: 2640–4.

29 Larsen MV, Cosentino S, Rasmussen S *et al*. Multilocus sequence typing of total-genome-sequenced bacteria. *J Clin Microbiol* 2012; **50**: 1355–61.

30 Jolley KA, Bray JE, Maiden MCJJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res* 2018; **3**: 124.

31 Center for Genomic Epidemiology. cgMLSTfinder. https://bitbucket.org/genomicepidemiology/cgmlstfinder/src/master/.

32 Alikhan N-F, Zhou Z, Sergeant MJ *et al*. A genomic overview of the population structure of *Salmonella*. *PLoS Genet* 2018; **14**: e1007261.

33 Dangel A, Berger A, Messelhäußer U *et al*. Genetic diversity and delineation of *Salmonella agona* outbreak strains by next generation sequencing, Bavaria, Germany, 1993 to 2018. *Euro Surveill* 2019; **24**: 1800303.

34 Ridom GmbH. *Salmonella enterica* cgMLST Scheme. https://www.cgmlst.org/ncs/schema/4792159/.

35 Oksanen J, Blanchet F, Friendly M *et al*. vegan: Community Ecology Package. https://cran.r-project.org/package=vegan.

36 Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* 2019; **47**: W256–9.

37 Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014; **30**: 2068–9.

38 Jain C, Rodriguez-R LM, Phillippy AM *et al*. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 2018; **9**: 5114.

39 Cury J, Jové T, Touchon M *et al*. Identification and analysis of integrons and cassette arrays in bacterial genomes. *Nucleic Acids Res* 2016; **44**: 4539–50.

40 Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet* 2012; **13**: 36–46.

41 Levings RS, Djordjevic SP, Hall RM. SGI2, a relative of *Salmonella* genomic island SGI1 with an independent origin. *Antimicrob Agents Chemother* 2008; **52**: 2529–37.

42 Arai N, Sekizuka T, Tamamura Y *et al*. *Salmonella* genomic island 3 is an integrative and conjugative element and contributes to copper and arsenic tolerance of *Salmonella enterica*. *Antimicrob Agents Chemother* 2019; **63**: e00429–19.

43 Adams MD, Bishop B, Wright MS. Quantitative assessment of insertion sequence impact on bacterial genome architecture. *Microb Genom* 2016; **2**: e000062.

44 Foley SL, Lynne AM, Nayak R. *Salmonella* challenges: prevalence in swine and poultry and potential pathogenicity of such isolates. *J Anim Sci* 2008; **86**: E149–62.

45 Didelot X, Bowden R, Street T *et al*. Recombination and population structure in *Salmonella enterica*. *PLoS Genet* 2011; **7**: e1002191.

46 Reinders A, Hee CS, Ozaki S *et al*. Expression and genetic activation of cyclic di-GMP-specific phosphodiesterases in *Escherichia coli*. *J Bacteriol* 2016; **198**: 448–62.

47 Petersen E, Mills E, Miller SI. Cyclic-di-GMP regulation promotes survival of a slow-replicating subpopulation of intracellular *Salmonella* Typhimurium. *Proc Natl Acad Sci USA* 2019; **116**: 6335–40.

48 Machida Y, Machida C, Ohtsubo E. A novel type of transposon generated by insertion element IS102 present in a pSC101 derivative. *Cell* 1982; **30**: 29–36.

49 Prudhomme M, Turlan C, Claverys J-PP *et al*. Diversity of Tn*4001* transposition products: the flanking IS*256* elements can form tandem dimers and IS circles. *J Bacteriol* 2002; **184**: 433–43.

50 Tansirichaiya S, Mullany P, Roberts AP. PCR-based detection of composite transposons and translocatable units from oral metagenomic DNA. *FEMS Microbiol Lett* 2016; **363**: fnw195.