Classifying mobile genetic elements and their interactions from sequence data: the importance of existing biological knowledge.

Sally R. Partridge[a,*], Virve I. Enne[b], Elisabeth Grohmann[c], Ruth M. Hall[d], Jonathan R. Iredell[a], Julian I. Rood[e], Paul H. Roy[f,g], Christopher M. Thomas[h], Neville Firth[d,*].

[a]Centre for Infectious Diseases and Microbiology, The Westmead Institute for Medical Research, The University of Sydney and Westmead Hospital, Westmead, NSW 2145, Australia.

[b]Centre for Clinical Microbiology, University College London, Royal Free Hospital, Rowland Hill Street, London NW3 2PF, UK.

[c]School of Life Sciences and Technology, Beuth University of Applied Sciences, Berlin, 13347 Berlin, Germany.

[d]School of Life and Environmental Sciences, University of Sydney, Sydney, NSW 2006, Australia

[e]Infection and Immunity Program and Department of Microbiology, Biomedicine Discovery Institute, Monash University, Clayton, Vic 3800, Australia.

[f]Centre de Recherche en Infectiologie, Université Laval, Québec, G1V 4G2, Canada.

[g]Département de Biochimie, de Microbiologie et de Bio-informatique, Université Laval, Québec, G1V 0A6, Canada.

[h]Institute of Microbiology and Infection, School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK.

Author contributions

S.R.P. and N.F. analyzed data and wrote the paper, with advice and approval from all other authors (listed alphabetically).

*To whom correspondence may be addressed.

Sally Partridge, CIDM, The Westmead Institute for Medical Research, 176 Hawkesbury Road, Westmead, NSW 2415; Australia. E-mail: sally.partridge@health.nsw.gov.au; Phone +61286273412.

Neville Firth, School of Life and Environmental Sciences, University of Sydney, Sydney, NSW 2006, Australia. E-mail: neville.firth@sydney.edu.au; Phone +61293513369.

Keywords: Plasmid, insertion sequence, antibiotic resistance

We agree with Che et al. (1) that understanding how mobile genetic elements (MGE) spread antimicrobial resistance (AMR) is important. However, decades of research have already characterized a diverse toolbox of MGE involved in emergence of AMR, including conjugative plasmids and insertion sequences (IS), and their interactions. 'Mobile' AMR generally arises following rare capture of a chromosomal gene by a particular 'intracellularly-mobile' MGE (IS; transposon, Tn; integron, In) and translocation to 'intercellularly-mobile' MGE (plasmids, phage, integrative elements; (2, 3)). Tools for systematic analysis of AMR gene-MGE interactions in mounting sequence data are needed but incorporating existing biological knowledge into their design is vital, as are representative benchmarking datasets and recognition of data biases and resulting limitations.

Che et al. screened available bacterial plasmid sequences for markers to classify them as non-mobilizable (no relaxase), mobilizable (relaxase) or conjugative (relaxase+). Relying purely on the absence of a predicted feature is risky, and Plascad's poor sensitivity in many species was not recognized due to a taxonomically inadequate benchmark dataset (different from the source cited). Numerous known conjugative plasmids, and close relatives, from low GC Gram-positive organisms (Dataset S1; e.g., *Staphylococcus*, pSK41, pWBG4, pWBG749; *Clostridium*, pCW3, pCP13; *Enterococcus/Streptococcus*, pAMβ1, pCF10, pRE25, pMG1, pIP501) (3-5) are misclassified so that no conjugative plasmids were identified in these genera. Additionally, mobilizable plasmids exploiting *trans*-acting relaxases from co-resident MGE (6, 7) were not considered, impacting both sensitivity and specificity. Misclassification within such a core parameter undermines dependent quantitative estimates (Figs. 1-5, S2-5, S17) and likely explains the supposedly high GC content of conjugative plasmids, while their larger size reflects the requirement for 10-30 conjugation genes; no "differences in the evolutionary process" need be envisioned.

Plasmid misclassification and database biases likely also contribute to dominance of Gram-negative species in "IS-associated transfer events" (Fig. 5). In these species, known MGE-AMR gene combinations typically cluster in large (tens of kb), complex multi-resistance regions (MRR; Fig. 12 in

(2)). Thus, proximity of an IS to an AMR gene is a flawed indicator of a role in movement of that gene. Che et al. (Discussion) acknowledged this limitation but nonetheless simply identified and counted (Figs. 4-6, S7-S16) IS, including fragments (non-functional), within an arbitrary 5 kb of each AMR gene. This may exclude relevant associations and/or include irrelevant ones (Table 1; Fig. 1). Also, 100% identity between AMR genes and the same IS-AMR gene distance in sequences from different species is dubious evidence for "recent" (not defined) horizontal transfer. AMR gene sequences are highly conserved (over decades) and analysis of all MGE types and signatures of insertion (flanking target site duplications, TSD) in entire MRR is usually required to determine what has moved and how (Fig. 1; (2)). Dependence on counts derived from Figs. 6, S7-S16 means that the discovered "massive IS-associated AMR gene transfer network" (Fig. 5B) and numerical details (Abstract) misrepresent the true picture.

We highlight the issues above in the hope that this will lead to improved methods for analyses of MGE in genomic data.

References

1. Y. Che *et al.*, Conjugative plasmids interact with insertion sequences to shape the horizontal transfer of antimicrobial resistance genes. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2008731118 (2021).
2. S. R. Partridge, Analysis of antibiotic resistance regions in Gram-negative bacteria. *FEMS Microbiol. Rev.* **35**, 820-855. (2011).
3. S. R. Partridge, S. M. Kwong, N. Firth, S. O. Jensen, Mobile genetic elements associated with antimicrobial resistance. *Clin. Microbiol. Rev.* **31**, e00088-17 (2018).
4. N. Goessweiner-Mohr, K. Arends, W. Keller, E. Grohmann, Conjugation in Gram-positive bacteria. *Microbiol. Spectr.* **2**, PLAS-0004-2013 (2014).
5. S. A. Revitt-Mills *et al.*, Virulence plasmids of the pathogenic Clostridia. *Microbiol. Spectr.* **7**, GPP3-0034-2018 (2019).
6. J. P. Ramsay, N. Firth, Diverse mobilization strategies facilitate transfer of non-conjugative mobile genetic elements. *Curr. Opin. Microbiol.* **38**, 1-9 (2017).
7. G. A. Blackwell, R. M. Hall, Mobilisation of a small *Acinetobacter* plasmid carrying an *oriT* transfer origin by conjugative RepAci6 plasmids. *Plasmid* **103**, 36-44 (2019).
8. S. R. Partridge, G. Tsafnat, E. Coiera, J. R. Iredell, Gene cassettes and cassette arrays in mobile resistance integrons. *FEMS Microbiol. Rev.* **33**, 757-784 (2009).
9. S. R. Partridge, J. R. Iredell, Genetic contexts of $bla_{\text{NDM-1}}$. *Antimicrob. Agents Chemother.* **56**, 6065-6067 (2012).
10. P. Siguier *et al.*, ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* **34**, D32-6 (2006).

Table 1. IS involved in "transfer events" with "abundance" >1 (blue bars in Che at al. Fig. 5A)

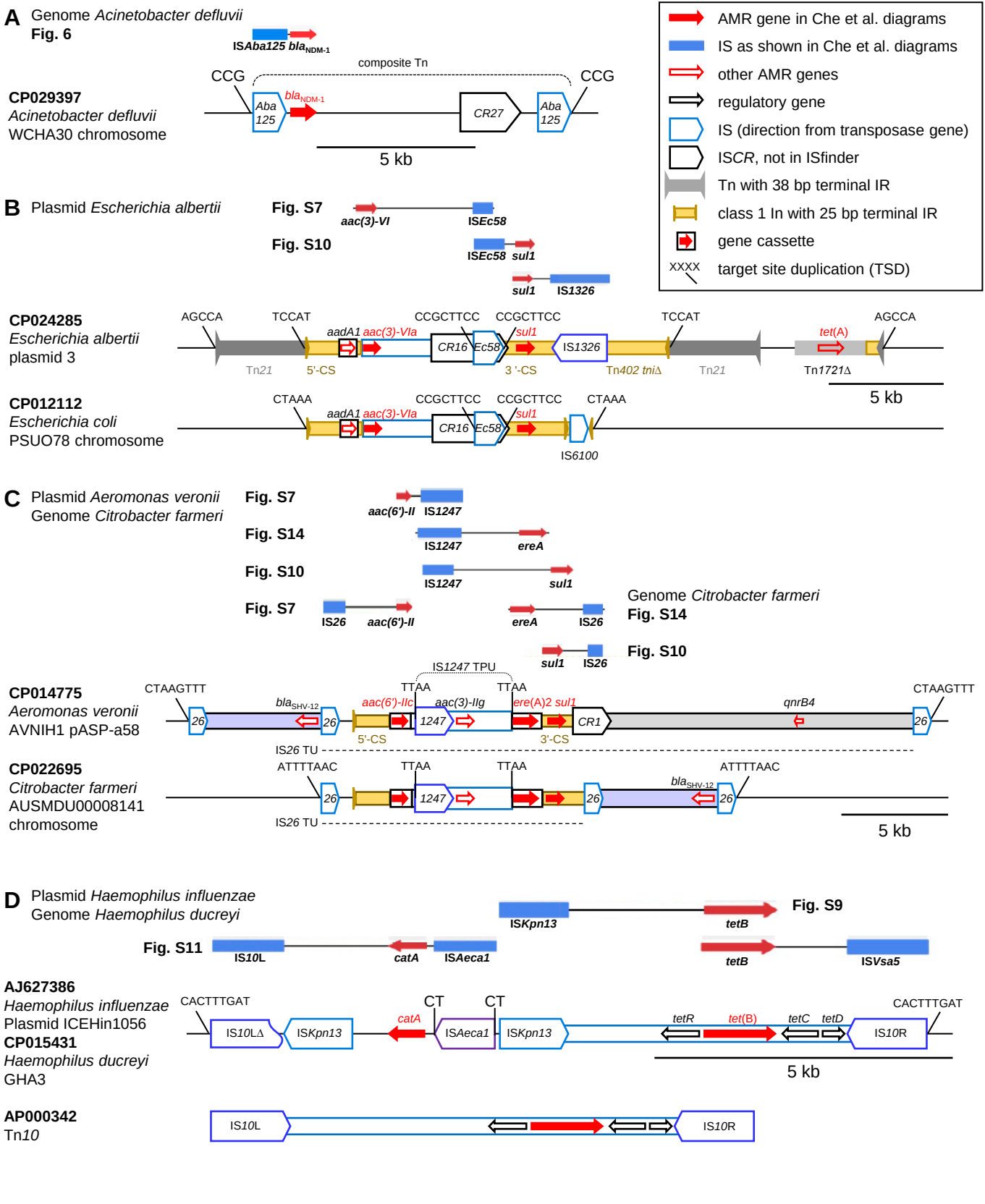
IS	Notes[a]	Reference(s)[b]
26	Dominant AMR-associated IS in Gram-negative clinical isolates. A single copy can move AMR genes as 'translocatable units' (TU) that insert adjacent to other copies of IS*26*.	(3)
6100	In In4-type class 1 In, explaining link to *sul1* and cassette-borne AMR genes. Also linked to *mph*(A). Rarely involved in transfer of AMR genes.	Fig. 1B (2, 8)
Ecp1	Moves segments (can be >5 kb) adjacent to right end as transposition units (TPU).	(2, 3)
Shes11	Corresponds to a Tn*3* family Tn, mainly small fragments in Che et al. Figs. S7-S16.	
15	Isoform of IS*26*.	
Pa40	Corresponds to a Tn*3* family Tn, mainly small fragments in Che et al. Figs. S7-S16.	
1133	Found inserted in Tn*5393*, which carries *aph(3'')-Ib* (*strA*) and *aph(6)-Id* (*strB*).	(2)
1R	Isoform of IS*1*, which has known associations with several AMR genes.	(2, 3)
Vsa5	Identical to IS*10R*.	
903B	Part of IS*Ecp1*-$bla_{\text{CTX-M-9}}$ group TPU, but not involved in their movement.	(2)
Pa38	Corresponds to a Tn*3* family Tn, mainly small fragments in Che et al. Figs. S7-S16.	
*10*L	Found in composite Tn*10* containing *tet*(B) (plus *tetRCD*, not considered AMR genes).	Fig. 1D (2, 3)
Aba125	Forms composite Tn e.g., Tn*125* (>5 kb, bla_{NDM}), Tn*aphA6* (*aph(3')-VIa*/*aphA6*).	Fig. 1A (7, 9)
5075	Targets 38 bp inverted repeats of Tn*21*-like Tn, likely prevents movement of Tn.	(2, 9)
1326	In In2-type class 1 In, so linked to *sul1* and cassette-borne genes.	(2, 8)
1X2	Isoform of IS*1*, which has known associations with several AMR genes.	(2, 3)
Cfr1	Found near *aac(3)-IId*, probably direct insertion not involved in movement.	(2)
Vsa3	Corresponds to part of IS*CR2* (not included in ISfinder), does not create TSD.	(2)
1247	Has inserted a TPU carrying *aac(3)-IIg* (previously *-IIf*) in the *ere(A)2* gene cassette.	Fig. 1C (2, 3)
Ec28	Truncated copy found upstream of *armA* in IS*26*-bounded region designated Tn*1548*.	(9)
Kpn13	Flanks *catA* in Tn*10* variant in *Haemophilus* ICEHin1056.	Fig 1D
*15*ΔII	Isoform of IS*26*.	
Ec59	IS*6*-family. Associated with *aac(3)-IVa* and *aph(4)-1a*. Role in transfer uncertain.	
1353	In In2-type class 1 In, so linked to *sul1* and cassette-borne genes.	Fig. 1B (2, 8)
Aba1	Moves $bla_{\text{OXA-23}}$ in composite Tn (e.g., Tn*2006*).	(2, 3, 7)
Kpn6	Flanks bla_{KPC} genes but within larger, mobile Tn*4401* (see Che et al. Discussion).	(2, 3)
Kpn7	Flanks bla_{KPC} genes but within larger, mobile Tn*4401* (see Che et al. Discussion).	(2, 3)
Aeca1	Inserted upstream of *catA4* in Tn*10* variant in *Haemophilus* ICEHin1056.	Fig 1D
*15*ΔI	Isoform of IS*26*.	
Aba14	Found upstream of *aph(3')-VIa* (*aphA6*) and $bla_{\text{NDM.}}$	(9)
UnCu1	Targets *attC* sites of gene cassettes, more likely involved in deletions.	(8)
Cco2	Has internal 'passenger' genes conferring aminoglycoside and streptothricin resistance.	
Ec58	Evidence of direct insertion between *aac(3)-VIa* and *sul1* (flanking 8 bp TSD).	Fig. 1B
4321R	Targets 38 bp inverted repeats of Tn*21*-like Tn, likely prevents movement of Tn.	(2, 9)

[a]Genes listed as *aac*, *aad*, *aph* types (encoding aminoglycoside modifying enzymes) in Che et al. Figs. 1C, 5A, S7 etc. mostly correspond to groups of several different genes with varying levels of identity. Specific names are given here and in Fig. 1. Simpler names used in an alternative nomenclature scheme are also given here (in parentheses).

[b]For IS with no reference listed refer to ISfinder (10) for additional information.

.

Fig. 1. Conserved IS-AMR gene distance/sequence identity is not sufficient to infer a role in mobility. Diagrams of IS-AMR gene pairs from Che et al. Figs. 6, S7-16 are shown with our analysis of the same sequences (identified by cross-referencing BLASTn searches with Tables S4, S7). (*A*) Analysing 5 kb upstream/downstream of an AMR gene may not be sufficient. Fig. 6 links $bla_{\text{NDM-1}}$ to IS*Aba125*, but another IS*Aba125* downstream of $bla_{\text{NDM-1}}$ forms a composite Tn with TSD. (*B*) Other MGE need to be considered. Supplementary figures link *aac(3)-VI* to IS*Ec58* and *sul1* to IS*Ec58* and IS*1326* in an *E. albertii* plasmid. Here, a class 1 In is inserted in a Tn*21*-derivative, both with TSD. In the only *E. coli* chromosome in Table S4 with *aac(3)-VIa* and IS*Ec58* (CP012112) these genes lie within a related class 1 In. *aac(3)-VIa* was apparently inserted by IS*CR16*, with TSD flanking IS*Ec58* indicating insertion. *sul1* is part of the 3'-conserved segment (CS) of class 1 In and IS*1326* is found in In2-type In. (*C*) Differences in IS mechanisms must be considered. Supplementary figures link *aac(6')-II*, *ereA* and *sul1* to IS*1247* and *aac(6')-II* to IS*26* in an *A. veronii* plasmid and a *C. farmeri* chromosome and *ereA* and *sul1* to IS*26* in the *C. farmeri* chromosome. IS*1247* has inserted *aac(3)-IIg* (not detected by Che et al.?) in a TPU (dotted line) with TSD. *aac(6')-IIc* (*aacA27*), *ere*(A)2 and *sul1* lie within different IS*26* TU (dotted lines) in the two sequences. (*D*) MRR can move as part of larger regions. Figs. S9 and S11 link *catA* to IS*10*L and IS*Aeca1* and *tetB* to IS*Kpn13* and IS*Vsa5* (=IS*10*R). The *catA* gene and *tet*(B) are within a variant of Tn*10* (Tn*10* itself shown below) with deletions and containing an IS*Kpn13*-bounded region into which IS*Aeca1* is inserted, with TSD. The Tn*10* derivative is in the same context in both *Haemophilus* sequences, flanked by the same TSD and the “plasmid” (AJ627386) is actually an integrative conjugative element (ICEHin1056).


A
Genome Acinetobacter defluvii
Fig. 6
ISAba125 blaNDM-1
composite Tn
CCG
CCG
CP029397
Acinetobacter defluvii
WCHA30 chromosome
blaNDM-1
Aba 125
CR27
Aba 125
5 kb
AMR gene in Che et al. diagrams
IS as shown in Che et al. diagrams
other AMR genes
regulatory gene
IS (direction from transposase gene)
ISCR, not in ISfinder
Tn with 38 bp terminal IR
class 1 In with 25 bp terminal IR
gene cassette
XXXX
target site duplication (TSD)
B
Plasmid Escherichia albertii
Fig. S7
aac(3)-VI
ISEc58
Fig. S10
ISEc58 sul1
sul1
IS1326
CP024285
Escherichia albertii
plasmid 3
AGCCA
TCCAT
aadA1 aac(3)-VIa
CCGCTTCC
CCGCTTCC
CR16
Ec58
sul1
IS1326
TCCAT
tet(A)
AGCCA
Tn21
5'-CS
3 '-CS
Tn402 tniΔ
Tn21
Tn1721Δ
5 kb
CP012112
Escherichia coli
PSUO78 chromosome
CTAAA
aadA1 aac(3)-VIa
CCGCTTCC
CCGCTTCC
CR16
Ec58
sul1
CTAAA
IS6100
C
Plasmid Aeromonas veronii
Genome Citrobacter farmeri
Fig. S7
aac(6')-II
IS1247
Fig. S14
IS1247
ereA
Fig. S10
IS1247
sul1
Fig. S7
IS26
aac(6')-II
ereA
IS26
Genome Citrobacter farmeri
Fig. S14
Fig. S10
sul1
IS26
IS1247 TPU
CP014775
Aeromonas veronii
AVNIH1 pASP-a58
CTAAGTTT
TTAA
TTAA
CTAAGTTT
blaSHV-12
aac(6')-Ilc
aac(3)-IIg
ere(A)2 sul1
qnrB4
26
26
1247
CR1
26
5'-CS
3'-CS
IS26 TU
CP022695
Citrobacter farmeri
AUSMDU00008141
chromosome
ATTTTAAC
TTAA
TTAA
ATTTTAAC
blaSHV-12
26
1247
26
26
IS26 TU
5 kb
D
Plasmid Haemophilus influenzae
Genome Haemophilus ducreyi
ISKpn13
tetB
Fig. S9
Fig. S11
IS10L
catA
ISAeca1
tetB
ISVsa5
AJ627386
Haemophilus influenzae
Plasmid ICEHin1056
CP015431
Haemophilus ducreyi
GHA3
CACTTTGAT
CT
CT
CACTTTGAT
IS10LΔ
ISKpn13
catA
ISAeca1
ISKpn13
tetR
tet(B)
tetC
tetD
IS10R
5 kb
AP000342
Tn10
IS10L
IS10R