Diversity of antibiotic resistance integrative and conjugative elements among haemophili

Ioanna D. Dimopoulou,¹ Sofia I. Kartali,¹ Rosalind M. Harding,²
Tim E. A. Peto³ and Derrick W. Crook³

Correspondence

Derrick W. Crook

derrick.crook@ndcls.ox.ac.uk

¹Department of Microbiology – Infectious Diseases, University Hospital of Alexandroupolis, Democritus University of Thrace, Alexandroupolis 68100, Greece

²Department of Zoology, University of Oxford, Oxford, UK

³Infectious Diseases and Clinical Microbiology, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DU, UK

The objective of this study was to investigate the sequence diversity in a single country of a family of integrative and conjugative elements (ICEs) that are vectors of antibiotic resistance in *Haemophilus influenzae* and *Haemophilus parainfluenzae*, and test the hypothesis that they emerged from a single lineage. Sixty subjects aged 9 months – 13 years were recruited and oropharyngeal samples cultured. Up to 10 morphologically distinct *Pasteurellaceae* spp. were purified, and then the species were determined and differentiated by partial sequence analysis of 16S rDNA and *mdh*, respectively. ICEs were detected by PCR directed at five genes distributed evenly across the ICE. These amplicons were sequenced and aligned by the neighbour-joining algorithm. A total of 339 distinguishable isolates were cultured. ICEs with all 5 genes present were found in 9 of 110 (8%) *H. influenzae* and 21 of 211 (10%) *H. parainfluenzae*, respectively. ICEs were not detected among the other *Pasteurellaceae*. A total of 20 of 60 (33%) children carried at least 1 oropharyngeal isolate with an ICE possessing all 5 genes. One of the five genes, integrase, however, consisted of two lineages, one of which was highly associated with *H. influenzae*. The topology of neighbour-joining trees of the remaining four ICE genes was compared and showed a lack of congruence; though, the genes form a common pool among *H. influenzae* and *H. parainfluenzae*. This family of antibiotic resistance ICEs was prevalent among the children studied, was genetically diverse, formed a large gene pool, transferred between *H. influenzae* and *H. parainfluenzae*, lacked population structure and possessed features suggestive of panmixia, all indicating it has not recently emerged from a single source.

Received 18 December 2006

Accepted 28 February 2007

INTRODUCTION

Antibiotic resistance amongst major human pathogens is becoming an increasing public health problem (Finch, 1998; Williams, 1997). The prevalence of transferable resistance has increased among human bacterial pathogens with the increasing use of antibiotics (Davies, 1997). Despite these major epidemiological changes, a detailed understanding of how conjugative elements have evolved in pathogen populations is incomplete and only gradually emerging. A better understanding of the evolutionary process leading to conjugative resistance elements amongst pathogens will help predict the future spread of resistance.

The emergence of antibiotic resistance among clinical isolates of *Haemophilus influenzae* dates back to the early 1970s (Gunn *et al.*, 1974; Mathies, 1972). Following the first detection of

resistance in 1973 (Mathies, 1972), the prevalence of resistance to ampicillin, and increasingly to other antibiotics, rose rapidly (Syriopoulou *et al.*, 1978). It has recently been shown that the main vector of this antibiotic resistance is not a plasmid, but a large integrative and conjugative element (ICE) (Dimopoulou *et al.*, 2002; Mohd-Zain *et al.*, 2004). These ICEs likely form a related family first regarded as conjugative plasmids in the 1970s (Elwell *et al.*, 1975, 1977; Laufs & Kaulfers, 1977; Laufs *et al.*, 1981; Saunders & Sykes, 1977; Stuy, 1980). Recently, publication of the whole sequence of one of these ICEs, ICE_{Hin1056} (Mohd-Zain *et al.*, 2004), provided the information from which the sequence diversity among these ICEs could be studied and, thereby, make the investigation of the evolution and routes of transmission of these ICEs possible. This approach will allow the hypothesis that they spread from a single point source to be tested by performing a detailed ecological study in a single geographical region, Alexandroupolis, Greece.

Abbreviation: ICE, integrative and conjugative element.

METHODS

Study subjects, sampling and culture. Sixty subjects aged 9 months – 13 years were recruited. They were selected from children admitted without infection to the Department of Paediatrics, University General Hospital of Alexandroupolis, during the periods October 2001 – December 2001 and April 2002 – July 2002. Children on or recently receiving antibiotic treatment (that is within a week) were excluded. After consenting, throat and dental swabs were collected. The swabs were placed in haemophilus transport medium (HTM) as previously described (Barbour *et al.*, 1993a). After bacteria were dispersed by vigorous vortexing, 50 µl HTM was cultured and the remainder of the medium was stored at -80°C for later culture.

Two sets of children were studied. The first set of 20 children was to determine differences between the floras of the throat versus the dental margin. From these 20 children throat and dental swabs were immediately subbed onto HTM (without antiserum) (Barbour *et al.*, 1993b) containing bacitracin (10 units ml^{-1}) alone or plus ampicillin ($2\text{ }\mu\text{g ml}^{-1}$) or tetracycline ($4\text{ }\mu\text{g ml}^{-1}$). The bacitracin alone plate allowed selection of *Haemophilus* spp. without selection against ampicillin or tetracycline susceptible isolates. Up to ten morphologically different colonies from the antibiotic containing media were purified onto the same media with the above concentrations of antibiotics, and tested for the presence of β -lactamase using intralactam strips (Mast Diagnostics). For the 40 remaining subjects, only throat swab samples from the frozen stock were cultured on the selective plates. Four, three and two different morphotypes from the bacitracin, ampicillin and tetracycline containing media, respectively, were purified.

Speciation and differentiation of isolates. In addition to standard phenotypic identification using X, V and X+V utilization, partial sequence of one of two regions of DNA encoding 16S rRNA was used for speciation. Differentiation of haemophili within an individual was based on sequence comparison of partial sequence of *mdh* (malate dehydrogenase), a gene used for *H. influenzae* multilocus sequence typing. Sequence difference in *mdh* of strains from the same individual defined distinguishable isolates. To design primers targeting sequence shared by different haemophili, the primers, MdhFP, MdhRP, MdhFS and MdhRS were designed by alignment of *mdh* from the whole sequences of *H. influenzae* KW20 (GenBank accession no. L42023) (Fleischmann *et al.*, 1995), *Haemophilus ducreyi* (GenBank accession no. NC_002940) and '*Haemophilus somnus*' (GenBank accession no. NC_008309), and choosing regions with conserved sequence. For differentiating the few *H. influenzae* and *Haemophilus parainfluenzae* isolates that failed to produce sequence from these primers, further internal primers were designed from alignment of the *mdh* sequences generated above, and further nested primer sets were designed targeting conserved sequences to optimize amplification and sequencing. This allowed differentiation of the few remaining strains.

The PCR amplicons were sequenced and read on an ABI PRISM 3700 DNA analyser (Applied Biosystems). The primers used for sequencing are shown in Table 1. Sequence data were assembled using the STADEN suite of computer programs. 16S rDNA sequence species identification was determined by DNA sequence alignment using the BLASTN algorithm available through the National Center for Biotechnology Information (NCBI). For comparison, *mdh* sequences from strains isolated from within an individual were aligned using both NRDB (written by Warren Gish, University of Washington, St Louis, MO, USA and available through <http://pubmlst.org/>) and ClustalW (Thompson *et al.*, 1994).

ICE detection. For ICE detection, five primer sets were designed from equally spaced core genes (these genes have been described by Mohd-Zain *et al.*, 2004) of ICEHin1056 (GenBank accession no.

AJ627386): two targeting a putative replicative region, namely *orf1* and *orf11* being homologues of *parA* and *topB*; two targeting a recently described putative conjugative region consisting of a novel type IV secretion system (Juhas *et al.*, 2007), namely *orf43* and *orf51*, genes *tfc14* and *tfc22*, respectively, with homology to *traB* and *pilT*, respectively; and lastly, *int*, referred to as *int1*, a homologue of *xerC/D*, a tyrosine integrase, at the 5' end of ICEHin1056. Because many ICEs in the study contained a divergent *xerC/D* homologue, a different primer set, *int2*, was designed from sequence of the 'cryptic' ICE present in the antibiotic susceptible *H. influenzae* 86-028NP, GenBank accession no. CP000057 (Harrison *et al.*, 2005) to detect and sequence this lineage of integrase. This lineage of *int*, *int2*, was typical of that found in the previously described *H. influenzae* type b isolate, 299, from Athens, Greece (Dimopoulou *et al.*, 1997). In addition, PCR products were sequenced, in preference, using nested sequencing primers where suitably available (Table 1).

The sequences were trimmed and the individual sequences of each gene and/or the concatenated sequences were aligned using the neighbour-joining algorithm available in MEGA, version 3.1 (Kumar *et al.*, 2004). Bootstrap values were used to judge the reliability of clusters. These values on the neighbour-joining tree are percentages of 100 computer-generated trees produced by randomly sampling the sequences and are shown at branch nodes.

Detecting antibiotic resistance genes by PCR. For all the 41 ICE containing isolates the presence of the following resistance genes were assayed using PCR: TEM-1 and ROB-1 β -lactamase or *tet(B)* and the other associated Tn10 genes *tetC* and *tetD*. For details of the primers used see Table 1.

Statistical analysis. Standard statistics were used for frequency analyses.

RESULTS AND DISCUSSION

Culture, speciation and differentiation of isolates

Five hundred and eighty strains were isolated and studied further. These isolates consisted mostly of strains picked from the bacitracin plate; a minority of resistant isolates were picked from the bacitracin plus ampicillin or tetracycline plate. Within individuals, only isolates distinguishable by 16S rDNA and/or *mdh* sequence analysis (i.e. unique sequence types for a child) were included for further analysis. There were 339 such isolates, 266 were isolated from throat swabs and 73 were isolated from dental margin cultures. A wide diversity of *Haemophilus* spp. isolates (up to eight distinguishable types) were observed for isolates cultured from the throat or dental margin, see Fig. 1. Furthermore, of the 20 people in whom both throat and dental margin cultures were available, 2–12 distinguishable strains were present within an individual (including both sites), 9 shared only 1 strain and 2 shared 2 strains at both swabbing sites. Consequently, most isolates cultured from each geographical site were unique with limited overlap.

This ecological study indicates a diverse population of haemophili much in line with the work of Kilian & Schiott (1975). The simple culture technique used in this study is inevitably biased away from detecting minority populations of susceptible haemophili. Notwithstanding this

Table 1. Primer name, sequence and co-ordinates

Primer	Nucleotide sequence	Site
16SF1	AGAGTTTGATCATGGCTCAG	8–27*
16SR1	GTATTACCGCGGCTGCTG	519–536c*
16SF2	CAAACAGGATTAGATACCCTGG	779–800*
16SR2	GTGACGGGCGGTGTGTAC	1392–1409c*
16SF2_seq	GATTAGATACCTGGTAGTCC	786–806*
16SR2_seq	CCCGGGAACGTATTCACCG	1369–1387c*
MdhFP	CACGTAAACCTGGTATGGATC	†
MdhRP	TCTTCTACCCCTTCTTTACC	†
MdhFS	CCTGGTATGGATCGTTCAGA	†
MdhRS	ACGGGTTGAGCGAAGAAACG	†
H.flu_mdhF	TCGTTCAGATTTATTCAATATTAATGC	†
H.flu_mdhR	CCTTCAACATAAGTACATTCAAC	†
H.para_mdhF	TAATTGAACATGTTGCAAAAACCTTG	†
H.para_mdhR	CCACTTAAGCCCTTCACTAATG	†
orf51F	CGGTTCCAGTGTTATATTCACG	45681–45702‡
orf51R	GAATGTGATCGGTGAGAAGC	46176–46195c‡
orf51_seqF	CCAGTGTTATATTCACGTAGCTC	45686–45708‡
orf51_seqR	GATCGGTGAGAAGCAATTAAC	46169–46189c‡
orf11F	GAGACTCACAAAGCGACAACC	10427–10447‡
orf11R	GGCTTGAGGTGCGTCATCATCTTC	10838–10861c‡
orf1F	CTCCCGGTGGTATTATATGAG	328–347‡
orf1R	CTATATGCCACATGCGCAGG	825–844c‡
orf1_seqF	GGTATTTATGAGTTTTTAACCCGTCG	336–361‡
orf1_seqR	CACATGCGCAGGGATTG	820–836c‡
orf43F	GATCAGGTTGTACAAGCTGG	37213–37232‡
orf43R	GCTGTTCCAGAGACAATTGC	37693–37712c‡
1056int_PCRF	CTTGGAATAGTTATGTTCCGCC	57595–57615‡
1056int_PCRR	GGGTTCTTCATTAAGTTGGTTGC	58173–58195c‡
299 int_PCRF	GGAACAGCTATATTCGTCAGC	146595–146615§
299 int_PCRR	CAGGTTTTTTCATTAGATTGTTGC	147170–147194§
tet(B)F	CGATTCCGACCTCATTAAGCAG	22749–22770‡
tet(B)R	GAGGGTCATTATATTTGCGCG	24129–24149c‡
tetCF	GTGGGATCGCTTTACCTGCA	23828–23847‡
tetCR/tetDR	GCCGATAATAACTTGGTGTAAC	25256–25277c‡
tetDF	CGTGCTGAGTTAACCAGTGATTG	24741–24763‡
TEM-1F	CCCGTCGTGTAGATAACTACGA	51058–51079‡
TEM-1R	ACAGCGGTAAGATCCTTGAGAG	51676–51697c‡
ROB_1L	CCCATACCAATCGGCTAGAA	823–842
ROB_2R	TTGCGTTATTGTCCAACCAA	971–990c

c, Reverse complement.

*Nucleotide positions as found on the 16S rDNA of *Escherichia coli*.

†As described in the text.

‡Nucleotide positions found on ICEHin1056, GenBank accession no. AJ627386.

§Nucleotide positions found on *H. influenzae* 86–028NP, GenBank accession no. CP000057.||Nucleotide positions found on *H. influenzae* plasmid RRob β -lactamase ROB-1 (*bla*) gene, GenBank accession no. AF022114.

limitation, a large number of distinguishable (unique) *Haemophilus* isolates were recovered per individual with little overlap between the two sampling sites; therefore, this suggests that the diversity is not being exhausted by the sampling. This also suggests there are even more strains present in people's throats than found in this study. This consequently points to a very large pool of host haemophili

per individual that can potentially support the family of *Haemophilus* ICEs.

ICE detection and characterization

Of the 339 distinguishable isolates, 30 contained complete ICEs (namely, all 5 target genes, i.e. *orf1*, *orf11*, *orf43*, *orf51*

Fig. 1. Charts indicating the number of distinguishable isolates of *Haemophilus* spp. cultured from either the throat (a) or the dental margin (b) in each child. Grey bars, the number of morphologically distinguishable isolates; ▲, unique isolates identified by sequence analysis.

and *int*, being present). Five of these ICEs were from dental cultures. A total of 11 incomplete ICEs (i.e. ≤ 4 target genes) were found, of which 4 were from dental cultures. Those elements designated as incomplete are based purely on failing to detect a PCR product following amplification using the sets of primers described in this study. Further characterization of these elements has not been done to determine whether they are in fact genetically coherent using other methods such as Southern blotting or whether they are functional, that is able to conjugate. Further studies need to be done to answer these questions.

The distribution of antibiotic resistance phenotypes and resistance genes was investigated among all the 41 ICE positive (both complete and incomplete) isolates, see Table 2. Of the 339 isolates, 66 expressed β -lactamase

Table 2. Distribution of ICE sequences among resistance phenotypes

Ap^R, ampicillin resistant; β -lac⁺, β -lactam positive; Tet^R, tetracycline resistant; β -lac⁻, β -lactam negative.

Phenotype	ICE (both complete and incomplete)		Total
	Positive	Negative	
Ap ^R β -lac ⁺	19	15	34
Ap ^R β -lac ⁺ , Tet ^R	20	6	26
Tet ^R	2	4	6
Ap ^R β -lac ⁻	0	30	30
Susceptible	0	243	243
Total	41	298	339

and/or were tetracycline resistant, and 41 of these were found to contain ICE sequences; therefore, ICE sequences are significantly associated with this phenotype ($\chi^2=193$, $P\leq 0.001$). This confirms in another geographical location the finding that ICE sequences associate with antibiotic resistance, as reported by Leaves *et al.* (2000). Of the 39 ICE positive isolates that were Ap^R and β -lactamase positive, 31 yielded PCR amplicons for TEM-1 and none yielded amplicons for ROB-1. Eight failed to amplify with either of the sets of primers TEM-1 or ROB-1. Determination of the likely genetic basis for β -lactamase activity in these eight strains is beyond the scope of this study and deserves further investigation. Of the 22 Tc^R resistant ICE containing isolates 17 yielded PCR amplicons to *tet*(B). All, but two yielded PCR products containing *tet*C or *tet*D, therefore suggesting the Tc^R resistance phenotype is associated with Tn10. The mechanism of resistance for the five Tc^R isolates, for which no PCR products were produced, is unknown and is also beyond the scope of this work. These observations suggest that ICEs associate largely with Tn10 encoded Tc^R and TEM-1 β -lactamase-encoded ampicillin resistance. Whether other β -lactamase genes or tetracycline resistance genes are found on ICEs needs to be investigated.

Among the 339 distinguishable isolates cultured in this study, 8 species were identified (7 of these being *Haemophilus* spp. and 1 *Actinobacillus actinomycetemcomitans*), as listed in Table 3. ICEs were only found in a similar proportion among the two species *H. influenzae* and *H. parainfluenzae* as follows: complete ICEs in 8.2 and 9.95 %, and incomplete ICEs in 1.8 and 4.3 %, respectively. The possibility that ICEs may occur among species other than *H. influenzae* and *H. parainfluenzae* cannot be ruled out as

Table 3. 16S rDNA speciation and ICE distribution

16S speciation	No.	Percentage	ICE (all loci)*		ICE (some loci)†	
			No.	Percentage	No.	Percentage
<i>Actinobacillus actinomycetemcomitans</i>	2	0.59	0	0	0	0
<i>Haemophilus aphrophilus</i>	2	0.59	0	0	0	0
<i>Haemophilus influenzae</i> ‡	110	32.45	9	8.2	2	1.8
<i>Haemophilus parahaemolyticus</i>	2	0.59	0	0	0	0
<i>Haemophilus parainfluenzae</i>	211	62.24	21	9.95	9	4.3
<i>Haemophilus paraphrohaemolyticus</i>	1	0.295	0	0	0	0
<i>Haemophilus paraphrophilus</i>	1	0.295	0	0	0	0
<i>Haemophilus segnis</i>	10	2.95	0	0	0	0
Total	339	100	30	—	11	—

*Complete ICE = all loci, all the 5 gene targets, present.
†Incomplete ICE = <5 gene targets detected.
‡Includes six *Haemophilus haemolyticus*, two with ICE.

Fig. 2. Neighbour-joining tree of *int* nucleotide sequences indicating the two lineages differing by 26%. Underlined strains indicate *int* from ICEs found in *H. influenzae*, including *H. haemolyticus*. *int1* refers to the ICE*Hin*1056 lineage and *int2* refers to the ICE*Hin*299 lineage. Bootstrap values are indicated on the branch nodes.

Fig. 3. Neighbour-joining trees produced from nucleotide sequence alignments of *orf1* (a), *orf11* (b), *orf43* (c) or *orf51* (d). Strain HT16B2, which lacked detectable *int*, is indicated with an asterisk. Underlined strains indicate ICEs found in *H. influenzae* including *H. haemolyticus*. Bootstrap values are indicated on the branch nodes. ▲, Strains associated with the ICE_{Hin1056} *int1* lineage, *int1*.

this study was underpowered for detection among the other species.

The occurrence of ICE containing haemophili among study subjects was as follows: 20 children (33 %) contained at least one complete ICE and 11 children (18 %) at least one incomplete ICE; therefore, 31 children (51 %) carried ICE sequences among culturable haemophili. In three of the children, both complete and incomplete ICEs were observed. Three different study subjects yielded strains with indistinguishable ICEs present in two genotypically different host haemophili, consistent with recent horizontal transfer of the ICE between haemophili in the nasopharynx. Two other genotypically indistinguishable ICEs were found in distinguishable host haemophili, though in different children, also suggesting horizontal transfer of these ICEs,

though, more distantly in the past. These data suggest there is frequent transfer of ICEs and this is consistent with the rapid flux of ICE genes, and indicates that these elements may rapidly spread through populations. This may indeed be what happened in the 1970s when resistance rapidly rose from low to high frequency among *H. influenzae* (Syriopoulou *et al.*, 1978) over a relatively short time.

There were two distinguishable lineages of *int*, one lineage, *int1*, similar to ICE_{Hin1056}, and the other, *int2*, which diverged by 26 % at the nucleotide level (as determined by NCBI BL2SEQ), see Fig. 2. These two lineages of *int* distributed unequally between *H. influenzae* and *H. parainfluenzae*; *int1* of the ICE_{Hin1056} lineage being significantly associated with *H. influenzae* ($\chi^2=19.95$, $P<0.001$). This finding is discussed further below.

Fig. 4. Neighbour-joining tree of the concatenated nucleotide sequence for all four loci *orf1*, *orf11*, *orf43* and *orf51*. No species-specific clustering of ICE sequences from *H. influenzae* or *H. parainfluenzae* is observed. The sequences of both *int* lineages are distributed across the diversity of the concatenated sequence. Underlined strains indicate *H. influenzae* including *H. haemolyticus*. The ICE in strain HT16B2 without *int* is indicated with an asterisk. The ICE_{Hin1056} and ICE_{Hin299}, both found in *H. influenzae* type b strains isolated from cases of meningitis, are labelled as 1056 and 299, respectively. Bootstrap values are indicated on the branch nodes. ▲, Strains associated with the ICE_{Hin1056} *int* lineage, *int1*.

Each of the partial gene sequences *orf1*, *orf11*, *orf43* and *orf51* present in the complete ICEs were aligned and the resultant trees compared as represented in Fig. 3. The topology of each of these trees differed markedly. There was no evidence that specific sequence types associated with either of the species *H. influenzae* or *H. parainfluenzae*. These features would be consistent with a single gene pool circulating between the two *Haemophilus* spp. and undergoing frequent recombination. The tree generated by alignment of the concatenated sequences, including those of ICE_{Hin299} (a previously described *H. influenzae* type b isolate from Athens, Greece) (Dimopoulou *et al.*, 1997) and ICE_{Hin1056}, both found in *H. influenzae* type b hosts isolated from cases of meningitis, clustered with those found among *H. parainfluenzae* (see Fig. 4). This suggests that the antibiotic resistance ICEs found in *H. influenzae* type b share the same gene pool as those occurring in the commensal *H. parainfluenzae* flora.

This family of antibiotic resistance ICEs have been shown to be prevalent among the set of Greek children studied, are genetically diverse, form a large common gene pool, lack population structure and possess features suggestive of panmixia. As such, these data are inconsistent with the recent emergence (in the past 30 years) of a single lineage of ICE transferring to *H. influenzae*, for example, from a *H. parainfluenzae*, as a rare founding event in Greece and then spreading horizontally among these Greek *H. influenzae*. Rather the sequence diversity observed among the ICEs suggests a long-standing and shared evolutionary relationship between these ICEs and the host bacteria *H. influenzae* and *H. parainfluenzae*. Alternatively, the data are consistent with multiple acquisitions from a diverse source population other than the two species *H. influenzae* and *H. parainfluenzae*. The present study did not find evidence of such a third population, though it was insufficiently powered to exclude one of the other oropharyngeal *Haemophilus* spp. being the source.

Though these ICEs lack population structure, curiously, these elements consist of a region at the 3' end represented by *int*, which can be either of two different lineages that diverge by 26 % at the nucleotide level. The remaining ICE sequences, represented by *orf1*, *orf11*, *orf43* and *orf51*, by contrast, differ by as little as 3 % at the nucleotide level, similar to the extent of nucleotide divergence observed in housekeeping genes of *H. influenzae* (Meats *et al.*, 2003). In this sample from Greece, the *int* lineage, *int1*, similar to that of ICE_{Hin1056} strongly associates with *H. influenzae* and is not found in *H. parainfluenzae*. However, the associated four non-*int* genes (*orf1*, *orf11*, *orf43* and *orf51*) do not consist of two divergent lineages and appear to be part of the same gene pool undergoing frequent recombination, including with the two distinct *int* lineages. This is inferred from the lack of congruence in the topology of the neighbour-joining trees for each of the individual gene alignments, see Fig. 3. The relevance of the *int1* lineage associating with *H. influenzae* is unclear and a better understanding may arise from population studies of larger collections of isolates from divergent geographical regions to see if this finding is largely unique to these Greek isolates. Another explanatory possibility that could be investigated in laboratory studies is a possible fitness advantage *int1* possessing ICEs have for *H. influenzae* over *H. parainfluenzae*.

Given the sequence diversity found among these Greek isolates, how did resistance emerge in the 1970s? The findings reported here are consistent with the hypothesis suggested by Laufs that ICEs existed as cryptic elements in the pre-antibiotic era and then became 'visible' in the clinical laboratory when they acquired transposons (e.g. Tn3 and Tn10) that encoded antibiotic resistance (Jahn *et al.*, 1979; Laufs *et al.*, 1981). Whether this occurred independently in different parts of the world or whether there has been pandemic spread of ICEs worldwide from a few common sources can now be addressed. The former possibility would be apparent were ICE DNA sequence diversity geographically structured in the same way as found for *Helicobacter pylori* (Falush *et al.*, 2003); while the latter would be associated with an even distribution of ICE sequence types worldwide, namely, a lack of geographical structure. This question can now be answered on collections of isolates from around the world using the techniques described in this report combined with more advanced phylogenetic analysis.

As these *Haemophilus* resistance ICEs have recently been shown to be related to a large family of syntenic genomic islands with deep evolutionary origins found among Proteobacteria (Mohd-Zain *et al.*, 2004), these findings emphasize the association between antibiotic resistance genes or structures that evolved well before the antibiotic era. Though this has been recognized for resistance genes themselves (D'Costa *et al.*, 2006; Levy, 2006), it has not been so apparent for the vectors. The family of *Haemophilus* ICEs clearly have distant evolutionary origins and this reinforces the deep origins of bacterial genes encoding and spreading antibiotic resistance.

ACKNOWLEDGEMENTS

We thank Dr Basiliki Arvanitidou, Dr Diogenis Spathopoulos and Dr Anastasia Grapsa for helping collect the strains. I. D. D. was funded by the Democritus University of Thrace and D. W. C. was on a Wellcome Trust leave fellowship for part of the work.

REFERENCES

- Barbour, M. L., Booy, R., Crook, D. W., Griffiths, H., Chapel, H. M., Moxon, E. R. & Mayon-White, D. (1993a). *Haemophilus influenzae* type b carriage and immunity four years after receiving the *Haemophilus influenzae* oligosaccharide - CRM197 (HbOC) conjugate vaccine. *Pediatr Infect Dis J* 12, 478–484.
- Barbour, M. L., Crook, D. W. & Mayon-White, R. T. (1993b). An improved antiserum agar method for detecting carriage of *Haemophilus influenzae* type b. *Eur J Clin Microbiol Infect Dis* 12, 215–217.
- Davies, J. E. (1997). Origins, acquisition and dissemination of antibiotic resistance determinants. *Ciba Found Symp* 207, 15–27 (discussion 27–35).
- D'Costa, V. M., McGrann, K. M., Hughes, D. W. & Wright, G. D. (2006). Sampling the antibiotic resistome. *Science* 311, 374–377.
- Dimopoulou, I. D., Jordens, J. Z., Legakis, N. J. & Crook, D. W. (1997). A molecular analysis of Greek and UK *Haemophilus influenzae* conjugative resistance plasmids. *J Antimicrob Chemother* 39, 303–307.
- Dimopoulou, I. D., Russell, J. E., Mohd-Zain, Z., Herbert, R. & Crook, D. W. (2002). Site-specific recombination with the chromosomal *tRNA^{Leu}* gene by the large conjugative *Haemophilus* resistance plasmid. *Antimicrob Agents Chemother* 46, 1602–1603.
- Elwell, L. P., De Graaff, J., Seibert, D. & Falkow, S. (1975). Plasmid-linked ampicillin resistance in *Haemophilus influenzae* type b. *Infect Immun* 12, 404–410.
- Elwell, L. P., Saunders, J. R., Richmond, M. H. & Falkow, S. (1977). Relationships among some R plasmids found in *Haemophilus influenzae*. *J Bacteriol* 131, 356–362.
- Falush, D., Wirth, T., Linz, B. & other authors (2003). Traces of human migrations in *Helicobacter pylori* populations. *Science* 299, 1582–1585.
- Finch, R. G. (1998). Antibiotic resistance. *J Antimicrob Chemother* 42, 125–128.
- Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, F. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A. & other authors (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269, 496–512.
- Gunn, B. A., Woodall, J. B., Jones, J. F. & Thornsberry, C. (1974). Ampicillin-resistant *Haemophilus influenzae*. *Lancet* 2, 845.
- Harrison, A., Dyer, D. W., Gillaspay, A., Ray, W. C., Mungur, R., Matthew, B. C., Zhong, H., Gipson, J., Gipson, M. & other authors (2005). Genomic sequence of an otitis media isolate of nontypeable *Haemophilus influenzae*: comparative study with *H. influenzae* serotype d, strain KW20. *J Bacteriol* 187, 4627–4636.
- Jahn, G., Laufs, R., Kaulfers, P. M. & Kolenda, H. (1979). Molecular nature of two *Haemophilus influenzae* R factors containing resistances and the multiple integration of drug resistance transposons. *J Bacteriol* 138, 584–597.
- Juhas, M., Crook, D. W., Dimopoulou, I. D., Lunter, G., Harding, R. M., Ferguson, D. J. & Hood, D. W. (2007). Novel type IV secretion system involved in propagation of genomic islands. *J Bacteriol* 189, 761–771.
- Kilian, M. & Schiott, C. R. (1975). Haemophili and related bacteria in the human oral cavity. *Arch Oral Biol* 20, 791–796.

- Kumar, S., Tamura, K. & Nei, M. (2004).** MEGA3: integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* **5**, 150–163.
- Laufs, R. & Kaulfers, P. M. (1977).** Molecular characterization of a plasmid specifying ampicillin resistance and its relationship to other R factors from *Haemophilus influenzae*. *J Gen Microbiol* **103**, 277–286.
- Laufs, R., Riess, F. C., Jahn, G., Fock, R. & Kaulfers, P. M. (1981).** Origin of *Haemophilus influenzae* R factors. *J Bacteriol* **147**, 563–568.
- Leaves, N. I., Dimopoulou, I., Hayes, I. & other authors (2000).** Epidemiological studies of large resistance plasmids in *Haemophilus*. *J Antimicrob Chemother* **45**, 599–604.
- Levy, S. B. (2006).** Mechanisms for resistance in soil. *Science* **312**, 529.
- Mathies, A. W., Jr (1972).** Penicillins in the treatment of bacterial meningitis. *J R Coll Physicians Lond* **6**, 139–146.
- Meats, E., Feil, E. J., Stringer, S., Cody, A. J., Goldstein, R., Kroll, J. S., Popovic, T. & Spratt, B. G. (2003).** Characterization of encapsulated and noncapsulated *Haemophilus influenzae* and determination of phylogenetic relationships by multilocus sequence typing. *J Clin Microbiol* **41**, 1623–1636.
- Mohd-Zain, Z., Turner, S. L., Cerdeno-Tarraga, A. M., Lilley, A. K., Inzana, T. J., Duncan, A. J., Harding, R. M., Hood, D. W., Peto, T. E. & Crook, D. W. (2004).** Transferable antibiotic resistance elements in *Haemophilus influenzae* share a common evolutionary origin with a diverse family of syntenic genomic islands. *J Bacteriol* **186**, 8114–8122.
- Saunders, J. R. & Sykes, R. B. (1977).** Transfer of a plasmid-specified beta-lactamase gene from *Haemophilus influenzae*. *Antimicrob Agents Chemother* **11**, 339–344.
- Stuy, J. H. (1980).** Chromosomally integrated conjugative plasmids are common in antibiotic-resistant *Haemophilus influenzae*. *J Bacteriol* **142**, 925–930.
- Syriopoulou, V., Scheifele, D., Smith, A. L., Perry, P. M. & Howie, V. (1978).** Increasing incidence of ampicillin resistance in *Hemophilus influenzae*. *J Pediatr* **92**, 889–892.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994).** CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions – specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.
- Williams, R. (1997).** Developments in monitoring of antimicrobial resistance. *WHO Drug Inf* **11**, 248–249.