The *Klebsiella* pheV tRNA locus: a hotspot for integration of alien genomic islands

Jon Jurriaan van Aartsen*

Department of Infection Immunity and Inflammation, University of Leicester, Leicester, UK.

*Corresponding author: Lab 212, Department of Infection Immunity and Inflammation, Maurice Shock Medical Sciences Building, University of Leicester, University Rd, Leicester LE1 9HN. Tel: +44 (0)116 2523056. Email: jjv1@le.ac.uk

**Supervisor:** Kumar Rajakumar, Department of Infection Immunity and Inflammation, University of Leicester, Leicester, UK.

---

*Klebsiella* sp. cause a wide range of human infections, particularly nosocomial septicaemia, pneumonia and urinary tract infections. Like other *Enterobacteriaceae*, *Klebsiella* are likely to possess plastic genomes comprised of core regions interspersed with horizontally acquired genomic islands. As phenylalanine tRNA genes are known to be occupied by islands in other *Enterobacteriaceae*, we utilized PCR-based screening and chromosome walking techniques to examine the pheV locus in *Klebsiella* isolates from blood stream and urinary tract infections. We hypothesized that this gene was an integration hotspot that served as a repository for novel genetic material in *Klebsiella*. The pheV site in *Klebsiella* KR116 and KR164 harboured an islet encoding four genes, two with similarity to genes within an island downstream of pheR in *Salmonella enterica* serovar Typhi CT18. In KR173 the locus contained a larger, potentially intact version of this island and harboured an integrase gene similar to that in the S. Typhi CT18 island. However, the *Klebsiella* and *Salmonella* islands were clearly distinguishable by strain-specific segments and organizational variation. On the basis of available sequence and restriction fragment length polymorphism data, three other *Klebsiella* isolates were found to possess an entirely distinct entity that resembled a 12.6 kb *phe*V associated island in *K. pneumoniae* MGH78578. This island was predicted to encode a P pilus-like structure, a probable virulence factor on the basis of parallels with *E. coli*. A unique and intriguing feature of *Klebsiella pheV* loci was the presence of multiple tandem repeats of up to 163 bp immediately downstream of *pheV* and a truncated copy at the opposite end of the islands. The tRNA proximal repeats were variable in number and size between isolates, while the solitary downstream repeats varied in length. These elements may represent genetic debris of previous recombination events. In conclusion, the *pheV* locus of *Klebsiella* exhibited considerable variability between strains and harboured at least two distinct island types that could play important roles in adaptation and/or virulence. Functional characterization of this genetic armory will help unravel basic microbial and pathogenesis processes and may in time lead to improvements in the diagnosis, prevention and treatment of *Klebsiella* infections.

Key words: *Klebsiella*, phenylalanine, tRNA, genomic island, pathogenicity island, genome plasticity.

---

Introduction

The *Klebsiella* genus contains a diverse group of commensal and pathogenic species. *K. pneumoniae* and *K. oxytoca* are the most frequently implicated species in nosocomial and community acquired *Klebsiella* infections, which include pneumonia, septicaemia, urinary tract infections and wound infections.\(^1\), \(^2\)

Bacterial genomes consist of two parts: the core and flexible genome. The core genome is shared by nearly all strains of the same species and encodes proteins involved in basic cellular function. The remainder is the flexible genome, a highly variable complement of strain-specific genes.\(^3\)

Genome variability plays an important role in the evolution and adaptive ability of bacteria, allowing for loss and/or acquisition of functions via mutational changes or horizontal gene transfer.\(^4\), \(^5\) The flexible genome often harbours segments of recognizable mobile genetic elements, such as transposons, phages, plasmids and archetypal integrative genomic islands, that may confer enhanced antibiotic resistance, pathogenicity or ecological fitness.\(^6\)

tRNA loci commonly serve as insertion sites for mobile elements as they are highly conserved between bacteria and thus allow for a greater degree of promiscuous
movement.\textsuperscript{7, 8} The \textit{phe} tRNA genes in several members of the \textit{Enterobacteriaceae} family have been found to harbour pathogenicity islands, large 10 to 200 kb clusters of strain-specific genes some of which confer defined virulence traits.\textsuperscript{9, 10} The \textit{pheV} locus of uropathogenic \textit{E. coli} strain J96 contains a 170 kb island (PAI I\textit{pve}) that encodes a P pilus, an essential virulence factor in pyelonephritis.\textsuperscript{11, 12} In enteropathogenic \textit{E. coli} this same locus is occupied by the LEE pathogenicity island that is crucial for the attachment and effacing phenotype responsible for much of the resulting pathology.\textsuperscript{8} Similarly, the \textit{pheV} locus of \textit{S. flexneri} serotype 2a is occupied by the \textit{she} island which encodes multiple genes involved in \textit{Shigella} pathogenesis.\textsuperscript{13–15} Additionally, a large-scale analysis of sequenced genomes using the Islander algorithm, which searches the genome for potential islands next to tRNA sites bordered by direct repeats and containing an integrase gene, has also identified phenylalanine tRNA genes as insertion ‘hotspots’ in several other bacteria.\textsuperscript{16} These findings were confirmed by tRNAcc, an algorithm that evaluates the content and context of tRNA and tmRNA genes across two or more genomes by identifying the conserved segments that flank potential island integration sites.\textsuperscript{17}

Although much effort has been focused on identifying \textit{E. coli}, \textit{Salmonella} and \textit{Shigella} genomic islands, little is known about the \textit{Klebsiella} flexible genome and its constituent islands, as only one \textit{Klebsiella} genome, that of \textit{K. pneumoniae} MGH78578, has been sequenced completely to date. On the basis that \textit{pheV} is a known integration target in other \textit{Enterobacteriaceae} and assuming that genomic islands move at low frequency between bacteria sharing an ecological niche and that once acquired integrate at the genome for potential islands next to tRNA sites bordered by direct repeats and containing an integrase gene, has also identified phenylalanine tRNA genes as insertion ‘hotspots’ in several other bacteria.\textsuperscript{16} These findings were confirmed by tRNAcc, an algorithm that evaluates the content and context of tRNA and tmRNA genes across two or more genomes by identifying the conserved segments that flank potential island integration sites.\textsuperscript{17}

\section*{Material and methods}

\textbf{Bacterial strains, plasmids and media}

Bacterial strains and plasmids used in this study are listed in Table 1. Clinical \textit{Klebsiella} isolates were obtained from blood and urine cultures at Leicester Royal Infirmary and stored in $\text{-}20$ °C/–80 °C glycerol stocks. Strains were grown at 37 °C in LB medium or LB agar, supplemented with ampicillin (100 $\mu$g ml$^{-1}$) or kanamycin (50 $\mu$g ml$^{-1}$) when required.

\section*{Preparation and manipulation of DNA}

Genomic DNA was isolated by a modified phenol/chloroform extraction protocol.\textsuperscript{18} Plasmid DNA was prepared by standard alkaline lysis.\textsuperscript{18} Restriction enzymes (Roche Diagnostics) and T4 DNA ligase (Promega) were used according to manufacturer’s instructions. Genomic libraries were constructed by overnight ligation of digested genomic DNA to appropriately digested pBluescriptII KS + . Chemically competent \textit{E. coli} DH5$\alpha$ were prepared and transformed according to standard methods. Standard subcloning methods were utilized.\textsuperscript{18}

\textbf{tRNAcc analysis and PCR primer design}

tRNAcc was run using default parameters.\textsuperscript{17, 20} 87 tRNA loci from \textit{K. pneumoniae} MGH78578 were mapped to homologous tRNA loci in \textit{K. pneumoniae} Kp342, a partially sequenced genome. Subsequently, the tRNAcc subprogram ExtractFlank was used to obtain and align 2 kb upstream and downstream conserved flanking regions corresponding to these tRNA loci from both genomes. Genomic islands were identified as non-homologous regions lying between conserved upstream and downstream flanks. Primers to amplify across the \textit{pheV} locus, 55pheU (CGTGCCTTT TAGCGCAATGT) and 55pheD (GACATAACCATTTAC CCACTCGT), were designed using the upstream and downstream \textit{pheV} flanking consensus sequences, respectively (M. Patel and H.Y. Ou, personal communication).

\textbf{tRIP PCR, SGSP PCR reactions and sequencing}

tRIP PCR (tRNA site interrogation for pathogenicity islands, prophages and other genomic islands PCR) reactions were performed in a volume of 20 $\mu$l using 1.25 U GoTaq® DNA polymerase (Promega), 0.4 $\mu$l of 10.0 mM dNTP, 20 pmol of primers 55pheU and 55pheD, and 10 ng of genomic DNA as template. Cycling conditions comprised 30 cycles of 30 s at 95.0°C, 30 s at 59.0°C and 3 min at 72.0°C. SGSP PCR (single genome specific primer PCR) was performed similarly, but used genomic libraries based on five distinct restriction enzymes (\textit{EcoRI}, \textit{BamHI}, \textit{PstI}, \textit{HindIII}, \textit{HincII}) as template instead. Additionally, either the primer 55pheU or 55pheD was used in conjunction with a vector specific primer (T3 or T7) to amplify the island extremities adjacent to the upstream and downstream conserved flanks, respectively. SGSP PCR cycling conditions when using T7 comprised an initial 10 cycles of 30 s at 95.0°C, 30 s at 67.4°C (decreased by 1°C each cycle) and 4 min at 72.0°C. This was followed by 20 cycles of 30 s at 95.0°C, 30 s at 57.4°C and 4 min at 72.0°C. When using T3 the annealing temperatures of 67.4°C and 57.4°C were decreased to 63.0°C and 53.0°C, respectively. PCR amplions were gel purified using the DNA Spin Gel Extraction
PCR DNA purification kit (Yorkshire Bioscience) and DNA sequencing was performed by MWG Biotech.

Sequence analysis
Both local and online databases were searched for nucleotide and amino acid similarities using Blastn, Blastp, Blastx, tBlastn and tBlastx. The MobilomeFINDER and Islander databases were explored to identify whether similarity hits occurred within known genomic islands. Protein coding sequence (CDS) prediction was performed using Glimmer 3.02 and CDD identified protein domains. Tandem Repeat Finder and Blastn were used to localize upstream and downstream repeats, respectively. Repeat nucleotide sequences were aligned using ClustalX with default parameters.

Results
Interrogation of pheV loci in KR116 and KR164 reveals a novel genomic islet
Ten of 16 Klebsiella strains produced pheV tRIP PCR amplicons of 0.5 kb, confirming these loci were unoccupied. Four strains (KR162, KR163, KR169 and KR173) produced no tRIP PCR amplicon and two (KR116 and KR164) yielded an ~3.7 kb product.
KR116 _pheV_, the amplicon corresponding to the KR116 _pheV_ site, was ligated into pCR4-TOPO® (Invitrogen) and subcloned into pWSK12926 for sequencing. Sequence analysis revealed that this segment harboured four predicted CDS and was novel to _Klebsiella_ (Fig. 1). KR116_pheV_1 coded for a 51 amino acid (aa) protein with high homology to a part of a non-functional putative transposase in _Yersinia pestis_. The second CDS, KR116_pheV_2, was predicted to encode a novel 181aa protein with no Blastn, tBlastn or Blastp matches and no conserved domains. The protein encoded by KR116_pheV_3 (174aa) strongly matched a putative acetyltransferase in _Salmonella enterica_ Typhi CT18, an association supported by the detection of a Gcn5 related N-acetyltransferase (GNAT) domain at the protein’s amino terminus. KR116_pheV_4 was predicted to encode a 79aa protein that harboured a truncated version of a domain of unknown function present in a _S_. Typhi CT18 hypothetical protein. Interestingly, the two corresponding _S_. Typhi CT18 genes were themselves located within a 133.7 kb _pheR_ associated genomic island (H.Y. Ou, personal communication). The matching region of the _Klebsiella_ islet had 85% nucleotide identity to the _S_. Typhi CT18 island. Sample sequencing of the KR164 _pheV_ islet revealed ≥96% nucleotide sequence identity to that of KR116, strongly suggesting that the two strains harboured near identical islets at this genomic location.

**KR173 harbours a large integrase bearing element within the _pheV_ locus**

SGSP PCR amplicons produced using as template EcoRI (~2.7 kb), HindIII (~1.5 kb) and BamHI (~1.3 kb) genomic libraries of KR173 were selectively sequenced to chromosome walk into the putative island integrated into the KR173 _pheV_ gene (Fig. 2). Blastn revealed that the defined portions of the upstream arm (UA) of the KR173 island had high homology to corresponding regions of the _pheR_ island in _S_. Typhi CT18, the same island mentioned previously in relation to the KR116 and KR164 islets. This region of the _Salmonella_ island encoded a P4 integrase that had 93–99% aa identity to the matching predicted KR173 protein. However, the minor size discrepancy between these regions suggested a possible short insertion within the KR173 integrase gene (Fig. 2). Additionally, in KR173 there was a small 270 bp UA segment which did not match the _Salmonella_ island and was predicted to encode a hypothetical protein that lacked Blastp homologs. The last 172 bp of the KR173 UA matched the _S_. Typhi CT18 island ~2 kb further downstream of the integrase bearing Blastn hit.

The first ~1.4 kb of the KR173 _pheV_ island DA matched an equivalent region in KR116; approximately 900 bp of this common region, which encoded two hypothetical proteins, exhibited strong similarity to a portion of island DNA 12.1 kb downstream of _pheR_ in _S_. Typhi CT18. However, unlike KR116_pheV_4, the slightly larger KR173 homolog, like that of _Salmonella_, was predicted to encode the full domain of unknown function. Nucleotide sequence from further within the KR173 island did not exhibit DNA matches to KR116_pheV, _S_. Typhi CT18 or other _Genbank _entries. The sequence was predicted to code for a 303aa hypothetical protein with a very low homology Blastp hit to a _Lactococcus lactis cremoris_ MG1363 protein; there were no conserved domains identified.

**Restriction pattern similarities at three loci: a prevalent _Klebsiella_ island coding for type 1 pili?**

Three TRIP PCR negative strains (KR162, KR163 and KR169) that potentially harboured elements at the _pheV_ site produced comparable restriction patterns with SGSP PCR. Chromosome walking into the UA produced ~3.5 kb and ~2.1 kb fragments with _BamHI_ and _PstI_ genomic libraries, respectively. Similarly, DA analysis using _PstI_ libraries produced ~1.8 kb fragments, while the use of _EcoRI_ libraries generated ~4.0 kb products. *In silico* SGSP PCR analysis on the _K. pneumoniae_ MGH78578 genome revealed the amplicon sizes from the three test strains matched those expected from _K. pneumoniae_ MGH78578, which harboured a 12.6 kb genomic island immediately downstream of the

![Figure 1](https://academic.oup.com/biohorizons/article-abstract/1/1/51/233304/54)

Figure 1. An outline of the novel ~3.7 kb _Klebsiella_ genomic islet present at the KR116 and KR164 _pheV_ loci with predicted CDS and Blastp similarity hits. The grey rectangle highlights the full conserved domain of unknown function in the _S_. Typhi CT18 hypothetical protein homolog of KR116_pheV_4. UF, conserved upstream flank; DF, conserved downstream flank; DR, direct repeat; I, identity; E, expect value.
The sequence data and restriction patterns confirmed that at least 7.5 kb of the 12.6 kb MGH78578 island was present in all three test strains (Fig. 3 and Table 2). KPN_03400 putatively coded for a full length (397aa) P4 integrase, with 67% amino acid identity to an Enterobacter sp. 638 gene. In total, ten SGSP PCR products representative of the three test strains were sequenced and confirmed to have very high nucleotide homology (≥95%) to that of the MGH78578 island. The only major discrepancy was an extra 163 bp repeat in MGH78578 (see in what follows).
integrate. KPN_03401 to KPN_03403 and KPN_03406 had
low identity (39–53%) to protein homologs in
*Yersinia frederiksenii* ATCC29909, ATCC33641, and
MGH78578. Additionally, these genes
were located ~5 kb downstream of a lambda like integrase,
suggesting that the *Yersinia* counterparts were also part of
a larger acquired island. *KP_03407* putatively encoded a
hypothetical protein. The presence of a similar gene in a
*Y. intermedia* strain supported the hypothesis that these genes code for a novel
*Klebsiella* fimbrial appendage.

**Variable tandem repeats downstream of pheV tRNA genes**

Analyses of DNA sequences from various *K. pneumoniae*
*tRNA-phe* loci revealed multiple repeats existed immediately
downstream of the tRNA gene in the majority of strains ana-
lysed and that a solitary repeat sequence often lay within the
opposite end of the island as well (Fig. 4).

The typical full length repeat unit was identified as a
163 bp sequence that started with 18 bp of the 3' terminus of
*pheV*. ClustalX alignment of the full set of identified
 repeats revealed that, with the exception of Kp342_UR3, all UA
repeat segments in this six strain panel exhibited
limited nucleotide variation and were highly homologous
(Fig. 5). KR310 possessed an empty *pheV* site and only
had a single 81 bp truncated repeat sequence. In contrast,
Kp342 also had an empty site but possessed a three repeat
sequence of 386 bp in total. Furthermore, even though
KR173 harboured the large novel island identified in this
study it lacked any recognizable upstream repeats, perhaps
with the exception of the 18 bp 3' terminus of its *pheV*
gene. These observations suggested that the UA repeat
number did not correlate with the size or type of island inte-
grated at the *pheV* locus.

The solitary DA repeats that were found also possessed
strong nucleotide similarity to the 163 bp consensus
sequence, although they were much shorter than their UA
counterparts. In KR116 and KR173, the downstream
repeats contained 17 bp of the 3' end of *pheV* and continued
with a further 63–69 bp of repeat sequence; those in KR162
and MGH78578 did not contain any *pheV* sequence at all
and were even shorter in length. KR310 and Kp342 did
not possess any downstream repeats. Additionally, none of
the DA repeats lay at the very extremity of the island arm,
as identified based on the boundary with highly conserved
downstream flanking core genome. Between 67 and 74 bp

---

**Table 2. Genes present on an island downstream of pheV in *K. pneumoniae* MGH78578**

<table>
<thead>
<tr>
<th>GenBank</th>
<th>Function</th>
<th>Coordinates†</th>
<th>O</th>
<th>L</th>
<th>Best homolog</th>
</tr>
</thead>
<tbody>
<tr>
<td>KPN_03399</td>
<td>tRNA-phe†</td>
<td>1</td>
<td>78</td>
<td>78</td>
<td><em>Klebsiella pneumoniae</em> W70-tRNA-phe</td>
</tr>
<tr>
<td>KPN_03400</td>
<td>Putative phage P4-integrase</td>
<td>873</td>
<td>1736</td>
<td>288</td>
<td><em>Enterobacter</em> sp. 638-P4 phage integrase</td>
</tr>
<tr>
<td>KPN_03401</td>
<td>Putative fimbrial usher protein</td>
<td>4576</td>
<td>7185</td>
<td>870</td>
<td><em>Y. frederiksenii</em> ATCC33641-PapC</td>
</tr>
<tr>
<td>KPN_03402</td>
<td>Putative P pilus assembly protein</td>
<td>7217</td>
<td>7927</td>
<td>237</td>
<td><em>Y. frederiksenii</em> ATCC33641-PapD</td>
</tr>
<tr>
<td>KPN_03403</td>
<td>Putative P pilus assembly protein</td>
<td>7990</td>
<td>8784</td>
<td>265</td>
<td><em>Y. frederiksenii</em> ATCC33641-PapD</td>
</tr>
<tr>
<td>KPN_03404</td>
<td>Putative fimbrial protein</td>
<td>8777</td>
<td>9355</td>
<td>193</td>
<td><em>Y. intermedia</em> ATCC29909-FimA</td>
</tr>
<tr>
<td>KPN_03405</td>
<td>Putative fimbrial protein</td>
<td>9374</td>
<td>9877</td>
<td>168</td>
<td><em>Y. intermedia</em> ATCC29909-FimA</td>
</tr>
<tr>
<td>KPN_03406</td>
<td>Putative P pilus assembly protein</td>
<td>9906</td>
<td>10637</td>
<td>244</td>
<td><em>Y. frederiksenii</em> ATCC33641-PapD</td>
</tr>
<tr>
<td>KPN_03407</td>
<td>Putative fimbrial associated protein</td>
<td>10688</td>
<td>11203</td>
<td>172</td>
<td>Hypothetical protein</td>
</tr>
</tbody>
</table>

O, orientation; L, length in amino acids; S, score; E, expect value; I, percentage identity.
†Coordinates are given in base pair values relative to the start of *pheV*.
†Data for tRNA-phe comprises length in nucleotides, and best homolog by Blastn analysis and its associated GenBank entry.
separated the single DA repeat and the predicted core genome. However, this could have been an artefact resulting from incorrect island/flank boundary determination by comparative genome analysis. This is known to occur particularly when only two input genomes are used for tRNAcc analysis, as was the case in this study.17

Discussion

Genome plasticity is of major importance in bacterial evolution and has frequently been associated with intraspecies phenotype variation.4 This paper has provided strong evidence that the Klebsiella pheV tRNA gene is an island integration hotspot. The presence of these elements gives rise to a variable genomic region that serves as a repository for novel genes. Of the 16 strains investigated, six were found to harbour one of three different islands within their pheV locus. Two entities, one potentially being a markedly truncated version of the other, were entirely novel to Klebsiella. The discovery of several islands at this locus was in perfect agreement with our hypothesis and with the findings of Germon et al. who reported frequent occupation of the E. coli pheV locus.9 Similarly, the presence within these elements of previously undiscovered DNA sequences was consistent with the report by Hsiao et al.29 who observed that novel bacterial genes were commonly localized to genomic islands.

KR116_pheV_3 was predicted to code for a hypothetical protein harbouring a GNAT domain, a signature motif of an enzyme superfamily widely distributed in all kingdoms.30 One group of GNAT proteins, the aminoglycoside transferases, chemically modify aminoglycosides, thus resulting in resistance to members of this class of antimicrobials.30 However, preliminary antibiotic susceptibility assays performed on KR116 have failed to show an increase in resistance to gentamicin or kanamycin (J.J.v.A., unpublished data) alluding to a possible alternative function for KR116_pheV_3. A second group of GNAT enzymes, represented by the E. coli GlmU protein, function as glucosamine-6-phosphate N-acetyltransferases. GlmU produces UDP-N-acetylglucosamine, an essential precursor of two biofilm components: peptidoglycans and lipopolysaccharides.31 In Klebsiella urinary tract and respiratory tract infections, the ability to form biofilms on abiotic surfaces is of recognized importance.32 Inhibition of GlmU reduces biofilm formation and bacterial colonization, and hence pathogenicity,31 raising the intriguing possibility that KR116_pheV_3 may play a direct role in virulence. Nevertheless, bioinformatics analysis alone could not attribute a specific function to KR116_pheV_3.
that this gene is located within the potentially unstable flexible genome of multiple strains, it is likely that it is being retained as a consequence of selection pressure that arises from its yet to be defined biological role.

The KR173 pheV island is likely to be an example of a cross genus lateral transfer event. Both the *Salmonella* and *Klebsiella* islands encoded a putative P4-related integrase, a type of enzyme which has been experimentally determined to control island excision and integration. Island encoded integrases are widely regarded as playing an essential role in horizontal island transfer events and genomic island evolution. The high level of nucleotide similarity of the terminal regions of both islands and their presence within phenylalanine tRNA genes in two distinct hosts, further substantiates the idea of inter genus movement. Additionally, the *Klebsiella* pheV and *Salmonella* pheR genes possess an identical nucleotide sequence, thus providing an identical substrate for the two integrases during site-specific recombination. We hypothesize that these elements arose from a common island ancestor present within the wider gene pool shared by *Salmonella* and *Klebsiella*. Over time, each island has differentiated and evolved independently as a function of the different environmental and host-derived selection pressures that are exerted on these two types of bacteria. The available evidence also suggests that the KR116/KR164 islet represents a remnant of the larger KR173 pheV island that has resulted from an imprecise excision or deletion event; similar phenomena affecting other genomic islands have been reported previously.

A 12.6 kb pheV island was identified in *K. pneumoniae* MGH78578 and although it harboured a large cluster of genes encoding P pilus subunits, including PapD, PapC and FimA, it did not appear to encode all subunits. P pili are essential virulence factors in *E. coli* that cause pyelonephritis, raising the prospect that the island may contribute to disease causation. The consequences of an incomplete set of P pilus subunits on the bacterial phenotype remains to be tested, but it must be noted that the ‘missing’ subunits...
may be coded for elsewhere on the genome. Additionally, multiple genes putatively encoded the same or closely related subunits. The incentive for maintaining this level of apparent redundancy is unknown, but may involve increased ability to evade the immune system by varying antigens presented on the cell surface. Apart from an extra 163 bp repeat, the extremities of the MGH78578 phev island were almost identical to phev associated islands in KR162, KR163 and KR169. However, it is quite possible that repeat structure variation may be indicative of significant divergence further within a genomic island, an area hidden from our current interrogation tools. Future studies using a yeast recombinational island capture system will be invaluable in addressing this matter.

A unique and intriguing feature of Klebsiella phev loci was the presence of multiple tandem repeats. With the exception of tandem insertion sequence elements, the E. coli LEE pathogenicity island is the only other island described as harbouring terminally located, large, directly oriented repeats (136 bp). The presence of other repeat motifs elsewhere in the Klebsiella genomes remains to be determined, though the phev 163 bp repeat sequence is specific to this particular locus. The differing numbers of UA repeats are reminiscent of many other variable number tandem repeat (VNTR) loci that have been described in bacteria; these are thought to originate from slipped strand mispairing or recombination events. It is widely accepted that some VNTR loci can affect phase variation, either by modifying promoters that modulate transcription or by altering protein amino acid sequence. Generally, however, variations in a large proportion of VNTR do not cause observable phenotype changes and are believed to simply constitute scars of erroneous DNA replication. Given that the phev repeats do not lie within or in close proximity to recognizable CDS, these repeats could well fall into the latter group. Alternatively, similar to the attL/attR sites that flank integrated prophages, the repeats may have arisen from a succession of previous imprecise island insertion and/or deletion events. This is supported by the finding that most of the Klebsiella phev repeats that we found exhibited high level similarity to the 3' end of phev, the target sequence for island insertion itself. The Klebsiella phev repeats appear to be an isolated phenomenon in the broader genomic island story and may provide new clues about island development, evolution and/or functional orchestration. Alternatively, these repeats may serve as a novel target for enhanced Klebsiella molecular typing.

The data presented here contributes to the complex and evolving story of bacterial genome plasticity and clearly places Klebsiella among the growing list of bacteria exhibiting a high degree of intra species genome diversity. Given parallels with fellow Enterobacteriaceae, E. coli, Shigella, Salmonella and Yersinia, it is very likely that the discovery and characterization of Klebsiella islands will be central to defining the detailed mechanisms of pathogenesis utilized by this important human pathogen. In light of increasing antimicrobial resistance, a comprehensive understanding of the genes, proteins and molecular events involved in pathogen survival, transmission and disease causation is essential if we are to develop a new generation of therapeutic, prognostic and diagnostic tools. Indeed this is likely to be crucial if we are to continue to successfully manage both community-acquired and nosocomial infections well into the future.

Acknowledgements

We thank Hong-Yu Ou for his assistance and support with tRNAac. We thank Dr Peter Munthali for providing the clinical Klebsiella strains from the Leicester Royal Infirmary, and James Lonnen, Ewan Harrison and Mansi Patel for laying the foundations on which this work was started. We also thank the Institute for Genomic Research (TIGR) and Washington University in St Louis for their policy of making preliminary sequence data publicly available and acknowledge the use in this study of unpublished genome data corresponding to Kp342 and MGH78578, respectively.

Funding

The author was supported by a Wolfson Foundation Award Intercalated BSc Studentship.

References


