Ann Microbiol (2014) 64:887–890
DOI 10.1007/s13213-013-0727-5

SHORT COMMUNICATION

Comparative genomics of closely related strains of *Klebsiella pneumoniae* reveals genes possibly involved in colistin resistance

Davide Sassera · Francesco Comandatore · Paolo Gaibani · Giuseppe D'Auria · Mara Mariconti · Maria Paola Landini · Vittorio Sambri · Piero Marone

Received: 24 June 2013 / Accepted: 17 September 2013 / Published online: 2 October 2013

Abstract Strains of colistin-resistant *Klebsiella pneumoniae* are emerging worldwide, due to the increased use of this molecule in antibiotic-resistant nosocomial infections. Comparative genomics was performed on three closely related *K. pneumoniae* strains isolated from three patients in a single hospital in Bologna, Italy. Two of these isolates are colistin-resistant, while the third is sensitive to this antibiotic. The designed bioinformatic approach detected, among the three analyzed genomes, single nucleotide polymorphisms, insertions and deletions, specific patterns of gene presence and absence, in a total of 270 genes. These genes were analyzed by automatic and manual methods, to identify those potentially involved in colistin resistance, based on the data available in the literature and on the mechanism of action of colistin, the alteration of the outer membrane. Three of the identified genes (*waaL*, *rfbA*, *vacJ*), all presenting non-synonymous substitutions in the colistin resistant strains, resulted to be of special interest, due to the specific function of their protein products, involved in the biosynthesis of the outer bacterial membrane.

Keywords *Klebsiella pneumoniae* · Colistin resistance · *waaL* · *rfbA* · *vacJ*

D. Sassera (✉) · F. Comandatore
Dipartimento di Scienze Veterinarie e Sanità Pubblica, Facoltà di Medicina Veterinaria, Università degli Studi di Milano, Via Celoria 10, 20133 Milano, Italy
e-mail: davide.sassera@unimi.it

P. Gaibani · M. P. Landini · V. Sambri
Unit of Clinical Microbiology, St. Orsola University Hospital, Via Pietro Albertoni 15, 40138 Bologna, Italy

G. D'Auria
Cavanilles Institute for Biodiversity and Evolutionary Biology, University of Valencia, C/ Catedrático José Beltrán nº 2, 46980 Paterna, Valencia, Spain

G. D'Auria
CIBER en Epidemiología y Salud Pública (CIBEResp), Madrid, Spain

M. Mariconti · P. Marone
Fondazione IRCCS Policlinico San Matteo, Via Taramelli 5, 27100 Pavia, Italy

Findings

Klebsiella pneumoniae is a Gram-negative bacterium, member of the family Enterobacteriaceae. It is one of the most important causes of nosocomial infections, due to its characteristics of being an opportunistic pathogen, of being able to infect different tissues/organs, and of being able to develop resistance quickly to most antibiotic classes (Queenan and Bush 2007; Nordmann et al. 2011). The recent spread of *K. pneumoniae* strains resistant to most antibiotic families (Gaibani et al. 2011) has led to the renaissance of colistin (polymixin E) use for the treatment of these infections (Yahav et al. 2012). Colistin is a pentacationic cyclic lipodecapeptide antibiotic that acts by permeabilizing the outer membrane of Gram-negative bacteria. This renewed use of colistin has led to an increasing number of reports of colistin-resistant *K. pneumoniae* strains (Yahav et al. 2012). Studies have identified, in different bacterial species, a complex pattern of molecular mechanisms leading to different instances of colistin resistance emergence, with a number of genes possibly involved. These include single nucleotide polymorphisms (SNPs) in *Salmonella enterica* (Sun et al. 2009), nonsense mutations leading to complete loss of lipopolysaccharide production (Moffatt et al. 2010), movement of long insertion sequences (Moffatt et al. 2011) and variation in gene

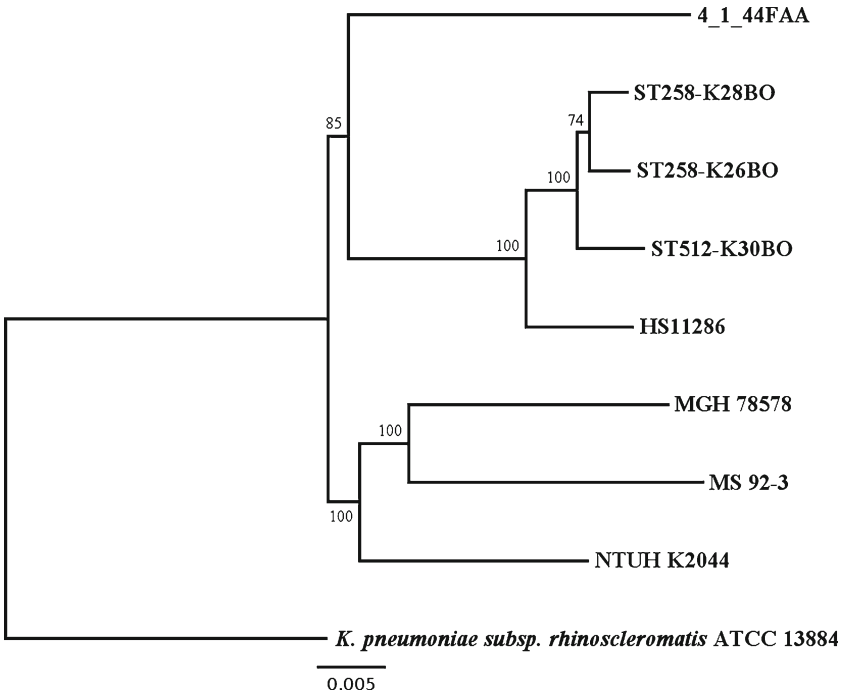

Fig. 1 Maximum likelihood phylogenetic comparison of 9 strains of *Klebsiella pneumoniae*, based on a concatenated alignment of 910624 aminoacids from 3155 genes. Bootstrap values are shown on each node

expression (Park et al. 2011, Henry et al. 2012) in *Acinetobacter baumannii*, while operon *dlt* (five enzymes for D-alanine esterification of lipoteichoic acid and wall teichoic acid) was shown to be required for resistance in *Bacillus cereus* (Abi Khattar et al. 2009). A recent comparative genomics study (Snitkin et al. 2012) presented the first eight mutations associated with colistin resistance in strains of *K. pneumoniae*. These include seven SNPs (four coding) and a 2-bp insertion. Four of the eight mutations found are in putative membrane proteins, confirming the importance of this group of proteins as targets for mutations associated with colistin resistance. Recently we published the genome sequences of three strains of *K. pneumoniae* isolated in Italy (Comandatore et al. 2013a, b). These three isolates (ST258-K26BO, ST258-K28BO, ST512-K30BO, hereafter respectively abbreviated as K26, K28 and K30) are supposed to be genetically very similar, as they were isolated in the same hospital in a very limited timespan and they belong to closely related strain types, (i.e. ST258 and ST512). All three isolates are resistant to all β-lactams, fluoroquinolones, macrolides, and aminoglycosides, while only K26 and K28 are resistant to colistin (Eucast 2013). In particular, all of the three isolates were identified as KPC-producers both by phenotypic and genotypic confirmatory assays (Ambretti et al. 2013). Here we present an in silico analysis aimed at uncovering the genetic variations possibly related to the emergence of colistin resistance.

Phylogenomic analysis was performed on the genomes of the three closely related strains of *K. pneumoniae* described above, plus six further published genomes of strains of this bacterial species (See Fig. 1 for the names of the nine isolates). The analysis was carried out using an in-house pipeline to select conserved genes through bidirectional best hit Blast, to align them using Muscle (Edgar 2004) and to trim them with Gblocks (Castresana 2000). The trimmed alignment (910624 aminoacids from 3155 genes) was subjected to phylogenetic analyses using PhyML version 3.0 (Guindon et al. 2010). Mapping, SNPs and short (<6nt) insertion-deletion (indel) calling analysis was performed with an in-house pipeline. In details, Illumina paired-ends reads of K26 and K28 where aligned to the K30 genome using SMALT (http://www.sanger.ac.uk/resources/software/smalt/); in addition, Illumina paired-end reads of K30

Table 1 Number of genes mutated or present/absent in the different *K. pneumoniae* genomes

Genomes	Differential presence pattern (OrthoMCL)	Non-synonymous SNPs	Insertions/deletions
K26 only	2	11	7
K28 only	3	19	3
K26 and K28	128	61	0
K30 and NTUH	36	N.A.	N.A.

N.A. stands for not applicable

Table 2 Genes of interest and respective mutations in the different *K. pneumoniae* genomes

Genes of interest	Respective mutations
Murein transglycosylase A—COG2821	Absent in K26 and K28
Metallopeptidase, M23B family—COG0793	Absent in K30 and NTUH-K2044
Colicin immunity protein E6—COG0859	Absent in K30 and NTUH-K2044
Cell division protein FtsQ—COG1589	C>A translated to P>T at position 682 in K26 and K28
Potassium efflux system KefA protein—COG3264	T>G translated to S>A at position 1837 in K26 and K28
glucarate dehydratase—COG4948	C>G translated to S>T at position 653 in K26 and K28
Lipoprotein VacJ—COG2853	C>T translated to A>V at position 20 in K26
dTDP-glucose pyrophosphorylase *rfbA*—COG1209	T>G translated to S>A at position 205 in K28
O-antigen ligase *waaL*/*rfaL*—COG3307	G>T translated to R>L at position 1100 in K28

were also aligned to the K30 genome, as an internal control. The Samtools suite (Li et al. 2009) was then used for format conversions. SNPs and indels where called using VarScan2 (Koboldt et al. 2012) accepting variant with p-values lower than 0.01 and results were parsed using in-house Perl scripts. High quality SNPs and indels were selected as those present in more than 75 % of the reads. Long indels (>5nt) were detected on orthologous groups (identified based on Blast results on the COG database) using an in-house Perl script pipeline and manually curated. Orthology analysis was performed using OrthoMCL (Li et al. 2003) comparing the genomes of the three closely related strains (K26, K28 and K30) and of the *K. pneumoniae* strain HS11286 (Liua et al. 2012). Genes found by orthoMCL to be present in K26 only, K28 only, K26 and K28, K30 and HS11286, as well as genes presenting non synonymous SNPs or indels, were inserted into a database, subjected to BLAST analyses (using the NR, Kegg, COG databases), sorted by COG (Cluster of Orthologous Groups) then manually analyzed.

As previously described, we isolated three strains (i.e. K26, K28 and K30) of *K. pneumoniae* obtained from patients from an Italian hospital and sequenced their genomes (Comandatore et al. 2013a and 2012b). K26 and K28 present high resistance to colistin (MIC >128 μg/ml), while K30 was found to be susceptible to this molecule. MLST profiles indicated the three strains to be of the highly similar profiles ST258 (K26 and K28) and ST512 (K30). A phylogenomic approach was implemented to investigate further the relatedness of the three strains, comparing them to six *K. pneumoniae* strain genomes available in the database. The resulting phylogenetic tree (Fig. 1) shows that K26 and K28 are the most closely related, and they are sister group to K30, confirming the high genetic relatedness of the three analyzed strains. In order to highlight genes possibly involved in colistin resistance we designed a comparative genomic approach (between K26, K28, K30, adding the phylogenetically closest HS11286 for orthology analysis) to find four types of genetic variations: SNPs, short indels (1-5nt), long indels (>5nt), presence/absence of orthologous genes. All the genes that exhibited a specific OrthoMCL presence pattern (169 genes) or that contained non-synonymous high quality SNPs (91 genes) or high-quality indels (10 genes), were collected (Table 1). The entire database of these 270 genes was manually analyzed based on annotation, COG category and Kegg classification, giving special consideration to nine genes found to belong to the M COG category (Cell wall/membrane/envelope biogenesis), as colistin acts by permeabilizing the outer membrane (see Table 2 for a description of the genes and of the respective mutations). We compared the database with a list of genes reported in the literature to be possibly implicated in colistin resistance in *K. pneumoniae* and in other bacteria, without finding any overlap.

In order to detect genetic differences in closely related *K. pneumoniae* isolates we performed a specifically designed comparative genomics approach. We obtained a database of 270 genes exhibiting variation. These genes were analyzed by automatic and manual methods, and three of them, all presenting non-synonymous substitutions, were found to be of special interest, due to the specific function of their protein products, as indicated by genomic annotation and Blast analyses. The gene *waaL* (COG3307), also known in other organisms as *rfaL*, presents a single non-synonymous substitution in the genome of the colistin-resistant strain K28. The product of this gene is an O-antigen ligase, which is part of a group of bacterial proteins involved in the synthesis of the O-antigen, a lipopolysaccharide component present in the outer membrane of Gram-negative bacteria (Pérez et al. 2008). This gene has been shown to be fundamental in the infection mechanisms of enterohemorrhagic *Escherichia coli* (EHEC) (Miyashita et al. 2012). The gene *rfbA* (COG1209), sometimes annotated as *rffH*, also presents a single non-synonymous substitution in the K28 strain. This mutation could be implicated in the mechanism of resistance to colistin, as this gene encodes a dTDP-glucose pyrophosphorylase, which is, similarly to *rfaL*, necessary for the O-antigen biosynthesis (Köplin et al. 1997). The gene *vacJ* (COG2853) of K26 also presents a single non-synonymous substitution. The membrane-associated lipoprotein product of this gene has been shown to be fundamental for intercellular spreading in *Shigella flexneri* (Suzuki et al. 1994) and was later identified as one of the mutated genes responsible for the emergence of the antimicrobial surfactant-resistant mutant *E. coli* OW66 (Nakata et al. 2010).

The discovery of mutations in three genes of interest suggests that emergence of single non-synonymous mutations

can be the mechanism, or one of the mechanisms, by which colistin-resistance can evolve in *K. pneumoniae*. These results are coherent with a previous study in indicating SNPs as possible cause of colistin resistance emergence (Köplin et al. 1997). However, for each of the four strains analyzed (two from the previous and two from this study), unique mutations were detected, cementing the hypothesis that colistin resistance is a phenotypic trait that could be determined by mutations in different genes, in some way involved in the composition of the outer envelope. Should the result here reported that resistance to colistin may derive from mutations in a gene involved in surfactant resistance be confirmed (i.e. the v*acJ* gene), this could imply that molecules that are widely (and perhaps improperly) used for cleaning, both in laboratories and hospitals, might lead to the selection of colistin-resistant bacteria.

Acknowledgments We thank Claudio Bandi and Simone Ambretti for laboratory help and useful discussion.

References

Abi Khattar Z, Rejasse A, Destoumieux-Garzón D, Escoubas JM, Sanchis V, Lereclus D, Givaudan A, Kallassy M, Nielsen-Leroux C, Gaudriault S (2009) The dlt operon of *Bacillus cereus* is required for resistance to cationic antimicrobial peptides and for virulence in insects. J Bacteriol 191:7063–7073

Ambretti S, Gaibani P, Berlingeri A, Cordovana M, Tamburini MV, Bua G, Landini MP, Sambri V (2013) Evaluation of phenotypic and genotypic approaches for the detection of class A and class B carbapenemases in Enterobacteriaceae. Microb Drug Resist. doi: 10.1089/mdr.2012.0165

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17:540–552

Comandatore F, Gaibani P, Ambretti S, Landini MP, Daffonchio D, Marone P, Sambri V, Bandi C, Sassera D (2013a) Draft genome of *Klebsiella pneumoniae* sequence type 512, a multi-drug resistant strain isolated during a recent KPC outbreak in Italy. Genome Announc 1:1

Comandatore F, Sassera D, Ambretti S, Landini MP, Daffonchio D, Marone P, Sambri V, Bandi C, Gaibani P (2013b) Draft genome sequences of two multi-drug resistant (MDR) *Klebsiella pneumoniae* ST258 isolates resistant to colistin. Genome Announc 1:1

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797

European Committee on Antimicrobial Susceptibility Testing (EUCAST) (2013) Breakpoint tables for interpretation of MICs and zone diameters Version 3.1, valid from 2013-02-11 [Accessed 15 February 2013] Available at: http://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Breakpoint_tables/Breakpoint_table_v_3.1.pdf

Gaibani P, Ambretti S, Berlingeri A, Gelsomino F, Bielli A, Landini MP, Sambri V (2011) Rapid increase of carbapenemase-producing *Klebsiella pneumoniae* strains in a large Italian hospital: surveillance period 1 March—30 September 2010. Euro Surveill 16:pii=19800

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59:307–321

Henry R, Vithanage N, Harrison P, Seemann T, Coutts S, Moffatt JH, Nation RL, Li J, Harper M, Adler B, Boyce JD (2012) Colistin-resistant, lipopolysaccharide-deficient *Acinetobacter baumannii* responds to lipopolysaccharide loss through increased expression of genes involved in the synthesis and transport of lipoproteins, phospholipids, and poly-β-1,6-N-acetylglucosamine. Antimicrob Agents Chemother 56:59–69

Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 22:568–576

Köplin R, Brisson JR, Whitfield C (1997) UDP-galactofuranose precursor required for formation of the lipopolysaccharide O antigen of *Klebsiella pneumoniae* serotype O1 is synthesized by the product of the rfbDKPO1 gene. J Biol Chem 272:4121–4128

Li L, Stoeckert CJ Jr, Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13:2178–2189

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079

Liua P, Li P, Jiang X, Bi D, Xie Y, Tai C, Deng Z, Rajakumar K, Ou HY (2012) Complete genome sequence of *Klebsiella pneumoniae* subsp *pneumoniae* HS11286, a multidrug-resistant strain isolated from human sputum. J Bacteriol 194:1841–1842

Miyashita A, Iyoda S, Ishii K, Hamamoto H, Sekimizu K, Kaito C (2012) Lipopolysaccharide O-antigen of enterohemorrhagic *Escherichia coli* O157:H7 is required for killing both insects and mammals. FEMS Microbiol Lett 333:59–68

Moffatt JH, Harper M, Adler B, Nation RL, Li J, Boyce JD (2011) Insertion sequence ISAba11 is involved in colistin resistance and loss of lipopolysaccharide in *Acinetobacter baumannii*. Antimicrob Agents Chemother 55:3022–3024

Moffatt JH, Harper M, Harrison P, Hale JD, Vinogradov E, Seemann T, Henry R, Crane B, St Michael F, Cox AD, Adler B, Nation RL, Li J, Boyce JD (2010) Colistin resistance in *Acinetobacter baumannii* is mediated by complete loss of lipopolysaccharide production. Antimicrob Agents Chemother 54:4971–4977

Nakata K, Koh MM, Tsuchido T, Matsumura Y (2010) All genomic mutations in the antimicrobial surfactant-resistant mutant, *Escherichia coli* OW66, are involved in cell resistance to surfactant. Appl Microbiol Biotechnol 87:1895–1905

Nordmann P, Naas T, Poirel L (2011) Global spread of Carbapenemase-producing Enterobacteriaceae. Emerg Infect Dis 17:1791–1798

Park YK, Choi JY, Shin D, Ko KS (2011) Correlation between overexpression and amino acid substitution of the PmrAB locus and colistin resistance in *Acinetobacter baumannii*. Int J Antimicrob Agents 37:525–530

Pérez JM, McGarry MA, Marolda CL, Valvano MA (2008) Functional analysis of the large periplasmic loop of the *Escherichia coli* K-12 WaaL O-antigen ligase. Mol Microbiol 270:1424–1440

Queenan AM, Bush K (2007) Carbapenemases: the versatile beta-lactamases. Clin Microbiol Rev 20:440–458

Snitkin ES, Zelazny AM, Thomas PJ, Stock F, NISC Comparative Sequencing Program Group, Henderson DK, Palmore TN, Segre JA (2012) Tracking a hospital outbreak of carbapenem-resistant *Klebsiella pneumoniae* with whole-genome sequencing. Sci Transl Med 4:148

Sun S, Negrea A, Rhen M, Andersson DI (2009) Genetic analysis of colistin resistance in *Salmonella enterica* serovar Typhimurium. Antimicrob Agents Chemother 53:2298–2305

Suzuki T, Murai T, Fukuda I, Tobe T, Yoshikawa M, Sasakawa C (1994) Identification and characterization of a chromosomal virulence gene, *vacJ*, required for intercellular spreading of *Shigella flexneri*. Mol Microbiol 11:31–41

Yahav D, Farbman L, Leibovici L, Paul M (2012) Colistin: new lessons on an old antibiotic. Clin Microbiol Infect 18:18–29