International Journal of Antimicrobial Agents 50 (2017) 210–218

Contents lists available at ScienceDirect

International Journal of Antimicrobial Agents

journal homepage: www.elsevier.com/locate/ijantimicag

Correlation between phenotypic antibiotic susceptibility and the resistome in *Pseudomonas aeruginosa*

Magali Jaillard [a,b], Alex van Belkum [a,*], Kyle C. Cady [c], David Creely [a], Dee Shortridge [a], Bernadette Blanc [a], E. Magda Barbu [c], W. Michael Dunne Jr [a], Gilles Zambardi [a], Mark Enright [d], Nathalie Mugnier [a], Christophe Le Priol [a], Stéphane Schicklin [a], Ghislaine Guigon [a], Jean-Baptiste Veyrieras [a]

[a] *bioMérieux SA, Marcy-l'Étoile, France*
[b] *University Lyon 1, Villeurbanne, France*
[c] *Synthetic Genomics, Inc., La Jolla, CA, USA*
[d] *Manchester Metropolitan University, Manchester, UK*

ARTICLE INFO

Article history:
Received 16 January 2017
Accepted 22 February 2017

Keywords:
Antibiotic resistance
Pseudomonas aeruginosa
Resistome-wide association study
RWAS
Integrons
Resistance gene mapping

ABSTRACT

Genetic determinants of antibiotic resistance (AR) have been extensively investigated. High-throughput sequencing allows for the assessment of the relationship between genotype and phenotype. A panel of 672 *Pseudomonas aeruginosa* strains was analysed, including representatives of globally disseminated multidrug-resistant and extensively drug-resistant clones; genomes and multiple antibiograms were available. This panel was annotated for AR gene presence and polymorphism, defining a resistome in which integrons were included. Integrons were present in >70 distinct cassettes, with In5 being the most prevalent. Some cassettes closely associated with clonal complexes, whereas others spread across the phylogenetic diversity, highlighting the importance of horizontal transfer. A resistome-wide association study (RWAS) was performed for clinically relevant antibiotics by correlating the variability in minimum inhibitory concentration (MIC) values with resistome data. Resistome annotation identified 147 loci associated with AR. These loci consisted mainly of acquired genomic elements and intrinsic genes. The RWAS allowed for correct identification of resistance mechanisms for meropenem, amikacin, levofloxacin and cefepime, and added 46 novel mutations. Among these, 29 were variants of the *oprD* gene associated with variation in meropenem MIC. Using genomic and MIC data, phenotypic AR was successfully correlated with molecular determinants at the whole-genome sequence level.

1. Introduction

Pseudomonas aeruginosa easily integrates exogenous DNA and, under selective antibiotic pressure, efficiently manifests resistance traits [1]. The emergence of resistance most likely occurs in isolates with enhanced virulence, during infection and treatment, allowing *P. aeruginosa* to evolve resistance to antibiotic-mediated killing [2]. Despite significant genetic variability within the species, antibiotic-resistant clones that spread globally have been identified [3]. Among them, multidrug-resistant (MDR) and extensively drug-resistant (XDR) clones are common, a phenotype that often relates to integron expansion. These pandemic clones include the highly prevalent sequence types ST235, ST111, ST348 and ST175 [4].

Progress is needed in the field of antibiotic resistance (AR) and antibiotic susceptibility testing [5]. Genome sequencing of *P. aeruginosa* has identified molecular markers for resistance to amikacin, meropenem and levofloxacin [6]. In addition, the International *Pseudomonas aeruginosa* Consortium has published a resistome recognising 73 AR genes in 389 isolates and highlighting the importance of the accessory genome [7]. CRISP-Cas-mediated immunity does not appear to be directly blocking acquisition of resistance elements [8]. Hence, additional phenotype–genotype association studies for *P. aeruginosa* are needed.

In other pathogens, studies have suggested that genomic antibiograms are as good as phenotypic ones. The first study involving mixed bacterial species documented 99.7% concordance between genotypes and phenotypes [9]. Work focusing on *Staphylococcus aureus*, *Escherichia coli*, *Klebsiella pneumoniae* and *Campylobacter* spp. has expanded these findings [10–12]. Optimum concordance was observed for clonal bacterial species such as *Mycobacterium tuberculosis* [13]. A recent resistome analysis of *P. aeruginosa* correlated meropenem resistance with outer membrane protein OprD

* Corresponding author. bioMerieux, 3 Route de Port Michaud, 38390 La Balme Les Grottes, France. Fax: +33474952599.
E-mail address: alex.vanbelkum@biomerieux.com (A. van Belkum).

http://dx.doi.org/10.1016/j.ijantimicag.2017.02.026
0924-8579/

polymorphism [14]. Genomic approaches hold promise for the development of future antibiotic susceptibility testing systems for routine use in clinical microbiology laboratories, although major knowledge gaps still need to be filled.

Here we present a resistome characterisation, integron dynamics and resistome-wide association study (RWAS) for 672 *P. aeruginosa* strains with complete genomes and minimum inhibitory concentration (MIC) values for various anti-*P. aeruginosa* antibiotics.

2. Materials and methods

2.1. Description of the strain panel

Genomic sequences of 672 *P. aeruginosa* strains were analysed. Strains were obtained from three collections: the bioMérieux collection ($n = 219$) [8]; the Kos collection ($n = 390$) [6]; and the Pirnay collection ($n = 63$) [14]. Genomic multilocus sequence typing (gMLST) and phylogenetic comparisons were performed.

Reference AR data based upon broth dilution assays and VITEK®2 tests were available for several antibiotics (Supplementary Table S1). More detailed information on all strains can be obtained from previously published work [8]. The genome sequences and antibiograms of 390 publicly available *P. aeruginosa* isolates (Kos collection) were added [6]. In this panel, resistance data covered meropenem, amikacin, levofloxacin and colistin.

2.2. Resistome

The resistome of the 672 *P. aeruginosa* isolates was obtained by annotating each genome assembly using an in-house, pan-bacterial AR sequence database. This database contains 2545 relevant, non-redundant reference sequences categorised into 569 AR genes. The database contained at least one sequence for each previously reported AR gene, including genes involved in susceptibility and/or resistance in *P. aeruginosa* [1]. Each genome was thus run through the database via BLASTn v.2.2.28+ for AR genes with at least 80% identity and 80% coverage. Overlapping alignments belonging to the same AR gene were clustered into a single hit if the overlap covered >10% of the aligned reference sequences. Only the best hit, defined as the alignment with the highest percentage identity times the reference sequence coverage, was kept to infer the haplotype state of the gene. Note that for a given isolate and for a given AR gene, several distinct annotations can be reported when hits appeared at distinct genomic locations (e.g. duplicated genes). However, precise plasmid copy number definition was impossible using this approach.

2.3. Resistome genotyping

The 672 isolates were genotyped for allele counts both at the 'locus' and 'variant' levels. By allele counts at the locus level, the number of copies of a given AR gene (e.g. bla_{TEM}, *mexX*, etc.) was measured irrespective of the number of alleles of the gene. Regarding the allele counts at the variant level, all of the detected sequences of a given AR gene were aligned using MAFFT v.6.861 with high-precision mode [15] in order to identify both single nucleotide polymorphisms (SNPs) including tri- and quadri-allelic sites and indels. Finally, for a given isolate, the number of occurrences of each SNP and indel alleles was counted and this was used as the corresponding allele count at the variant level (see Supplementary Table S2).

2.4. Integron analyses

Integrase genes were detected by aligning (tBLASTn) the assembled genome sequences with the protein sequences of three integrases (IntI1, IntI2 and IntI3) as derived from a targeted UniProt search (sequences are provided in Supplementary File S1). Contrary to IntI1, for which we were able to retrieve 36 sequences, only 3 sequences were found for IntI3 and 1 sequence for IntI2. A 70% identity and 70% protein coverage cut-off was applied for the alignments. Then, recognisable and annotated AR genes upstream of the integrase start codon were searched for and the physical distance between the start codon and the 5′ end of the resistance element was reported. All AR elements within a 10 kb window were included.

2.5. Resistome-wide association study (RWAS)

First, the strength of association between established genotypes linked to increases or decreases in the MIC (generally beyond the resistance breakpoint) was assessed. Second, novel candidate genotypes associated with additional variability in MIC values were looked for, taking the effect of known genotypes into account. Fig. 1 provides a literature-supported overview of the causal genes or variants thereof known to increase/decrease the MIC for each antibiotic in the association study.

2.6. Population structure

Results may be inflated by the presence of cryptic correlations between population structure and the MIC status of the strains. This could be due to population-wide linkage disequilibrium (LD) between causal mutations and genetic structure, or to a sampling bias leading to over-representation of related strains. Confounding effects related to population structure were restrained by computing the principal components (PCs) from the core–SNP genotype matrix and using them as covariates in the statistical association framework. To avoid extensive correlations that could mask the effect of a core gene on resistance, genotypes associated with polymorphisms in, for example, *gyrA*, *gyrB*, *parC*, *parE*, *folP* etc., were removed [16].

2.7. Associative modelling

Ordinal regression was used to develop the core statistical RWAS framework [17]. This provides the benefit of adequately accounting for the ordinal nature of the MIC values and facilitating inclusion of covariates. For each antibiotic, the PC was defined to include as covariates for population structure control, using a forward selection procedure [18]. This allows to build Z, the $n \times k$ matrix of PCs, where n is the number of isolates and k is the number of retained PCs. For each antibiotic, an optimal set of unequivocal genotypes was selected using a backward elimination procedure based on the Akaike information criterion (AIC) [18,19]. An optimal although conservative subset of genuine genotypes associated with MIC variability

Fig. 1. Overview of the phenotypic, genomic and resistome data of 672 *Pseudomonas aeruginosa* isolates included in the present analysis. The phylogenetic tree is inferred from core gene content and depicts the three major groups of *P. aeruginosa* in shades of purple. Depicted below the tree in dark green bars are the isolates that were found to be resistant to one or more of the nine antibiotics analysed in this study. Below, strains with draft genomes larger or smaller than the median are depicted as black or grey bars, respectively. The left panel provides the names of genes and non-synonymous mutations constituting the resistome for the nine antibiotics. Different antibiotic resistance mechanisms are colour coded (bar at the right of the determinant list). The right panel reviews resistance gene content on a per isolate basis, with grey shading denoting the presence of a given resistance gene or allele. The percentage of strains harbouring given resistance genes or alleles is shown on the far right, allowing for easy discrimination of core and accessory elements. The resistome structure illustrates that antibiotic inactivation genes are more likely to belong to the accessory resistome than efflux genes. Note that resistance genes embedded in integrons are colour coded as dark red bars. Additional columns with global information (e.g. plasmid content, percentage GC, etc.) can be added if needed.