



Molecular Determinants of Antibiotic Resistance in the Costa Rican *Pseudomonas aeruginosa* AG1 by a Multi-omics Approach: A Review of 10 Years of Study

Jose Arturo Molina-Mora¹ · Fernando García¹

Received: 26 January 2021 / Revised: 7 May 2021 / Accepted: 11 May 2021 / Published online: 17 June 2021
© International Human Phenome Institutes (Shanghai) 2021

Abstract

Pseudomonas aeruginosa AG1 (PaeAG1) is a Costa Rican strain that was isolated in 2010 in a major Hospital. This strain has resistance to multiple antibiotics such as β -lactams (including carbapenems), aminoglycosides, and fluoroquinolones. PaeAG1 is considered critical (Priority 1) due to its resistance to carbapenems, and it was the first report of a *P. aeruginosa* isolate carrying both VIM-2 and IMP-18 genes encoding for metallo- β -lactamases (MBL) enzymes (both with carbapenemase activity). Owing to these traits, we have studied this model for 10 years using diverse approaches including multi-omics. In this review, we summarize the main points of the different steps that we have studied in PaeAG1: preliminary analyses of this strain at the genomic and phenomic levels revealed that this microorganism has particular features of antibiotic resistance. In the multi-omics approach, the genome assembly was the initial step to identify the genomic determinants of this strain, including virulence factors, antibiotic resistance genes, as well as a complex accessory genome. Second, a comparative genomic approach was implemented to define and update the phylogenetic relationship among complete *P. aeruginosa* genomes, the genomic island content in other strains, and the architecture of the two MBL-carrying integrons. Third, the proteomic profile of PaeAG1 was studied after exposure to antibiotics using 2-dimensional gel electrophoresis (2D-GE). Fourth, to study the central response to multiple perturbations in *P. aeruginosa*, i.e., the core perturbome, a machine learning approach was used. The analysis revealed biological functions and determinants that are shared by different disturbances. Finally, to evaluate the effects of ciprofloxacin (CIP) on PaeAG1, a growth curve comparison, differential expression analysis (RNA-Seq), and network analysis were performed. Using the results of the core perturbome (pathways that also were found in this perturbation with CIP), it was possible to identify the “exclusive” response and determinants of PaeAG1 after exposure to CIP. Altogether, after a decade of study using a multi-omics approach (at genomics, comparative genomics, perturbomics, transcriptomics, proteomics, and phenomics levels), we have provided new insights about the genomic and transcriptomic determinants associated with antibiotic resistance in PaeAG1. These results not only partially explain the high-risk condition of this strain that enables it to conquer nosocomial environments and its multi-resistance profile, but also this information may eventually be used as part of the strategies to fight this pathogen.

Keywords *Pseudomonas aeruginosa* · PaeAG1 · Antibiotic resistance · Multi-omics · High risk

Abbreviations

2D-GE 2-Dimensional gel electrophoresis
CIP Ciprofloxacin

MBL Metallo- β -lactamase
MLST Multilocus sequence typing
MIC Minimum inhibitory concentration
PaeAG1 *Pseudomonas aeruginosa* AG1
ST Sequence type
WHO World Health Organization (WHO)

✉ Jos Artur Molina-Mora
jose.molinamora@ucr.ac.cr

Fernando García
fernando.garcia@ucr.ac.cr

¹ Centro de Investigación en Enfermedades Tropicales (CIET)
& Facultad de Microbiología, Universidad de Costa Rica,
San José, Costa Rica

Context

Antibiotic resistance is a major threat to public health, because it compromises the administration of appropriate antibiotic therapy. *Pseudomonas aeruginosa* is an opportunist pathogen that causes infections among immunocompromised hosts. In 2010, *P. aeruginosa* strain AG1 (PaeAG1) was isolated from a Costa Rican Hospital. This microorganism is a high-risk sequence type 111 clone (ST-111, a molecular genotype associated with epidemics and multidrug resistance; see more details in the “Comparative genomics” section), showing a resistance profile to multiple antibiotics such as β -lactams including carbapenems, aminoglycosides, and fluoroquinolones. PaeAG1 was identified as the first report of a *P. aeruginosa* isolate carrying both VIM-2 and IMP-18 genes encoding for metallo- β -lactamases (MBL) enzymes, both with carbapenemase activity. According to the World Health Organization (WHO), carbapenem-resistant *P. aeruginosa* is considered critical, being classified as Priority 1 group because of its resistance to this last resource antibiotic class. PaeAG1 has particular features at genomic and phenomic levels, many of them related to antibiotic resistance. Owing to these traits, for 10 years, we have been studying the molecular determinants of antibiotic resistance in PaeAG1 using a multi-omics approach.

Multi-omics

Biological systems rely on the DNA–RNA–protein information transfer paradigm that determines the phenotype of an organism (O’Donnell et al. 2020). The comprehensive or global assessment of a set of molecules, which requires interpretation of molecular intricacy and variations at multiple levels, has been referred to as “-omics sciences” (Subramanian et al. 2020). The current high-throughput nature of these techniques, as well as their increased accessibility in terms of time and cost, has triggered the volume of information that can be gathered in individual studies including multiple-omics levels, which together are called “multi-omics” (O’Donnell et al. 2020; Subramanian et al. 2020).

Multi-omics can provide a greater understanding of the flow of information in biological systems, from the original biological set-up or conditions (genetic, environmental, or developmental) to the functional consequences or relevant interactions (Hasin et al. 2017; Civelek and Lusic 2014). This makes it possible to draw more comprehensive conclusions on the biological processes in which these data sets must be integrated and analyzed as a holistic system (O’Donnell et al. 2020). Also, integrated approaches

that combine individual omics data help to bridge the gap from genotype to phenotype, are considered a promising strategy to understand the complexity of biological systems and unravel the mechanisms underlying the biological condition of interest (Subramanian et al. 2020; Civelek and Lusic 2014).

In this context, a comprehensive multi-omics approach has been implemented to study molecular determinants of antibiotic resistance in the PaeAG1 model, including genomics, transcriptomics, and proteomics (Fig. 1).

Antibiotic Resistance

Resistance arises from the extensive and intense use that humans have made of antibiotics over the last eight decades in different settings, including community, hospitals, veterinary medicine, animal production, and plant production. This generates a strong selection pressure in favor of microorganisms with mutations in chromosomal genes or with horizontally acquired genes that confer the ability to grow at therapeutic concentrations (Brauner et al. 2016; Berti and Hirsch 2020).

Antibiotic resistance is a current major threat to public health, because it compromises the administration of appropriate antibiotic therapy. The number of deaths directly attributable to antimicrobial resistance has been estimated to be about 300 million people through 2050 (approximately 10 million people per year), with a total financial loss of up to \$ 100 trillion (O’Neill J 2016). This scenario is indicative of the large scale and seriousness of the problem, mainly because of the lack of therapeutic options to treat infections, increasing patient morbidity and mortality (Woodford et al. 2011; Farajzadeh Sheikh et al. 2019), as well as an increase in the costs of health services. The situation is aggravated by the emergence of strains simultaneously resistant to multiple antibiotics (Firme et al. 2010), the knowledge limitation of interactions with pathogens and mechanisms of the action of antimicrobial agents, and the reduced development of new antibiotics (Brazas et al. 2005).

The minimum inhibitory concentration (MIC) is an in vitro criterion applied in the clinical setting for humans and animals, and indicates the concentration that antibiotics must reach in the tissues of infected individuals to have therapeutic efficacy (Andersson and Hughes 2014). Although it has no practical application outside the clinical setting, the MIC value also reveals the degree of resistance that a bacterial isolate has to a particular antibiotic. A concentration below the MIC is considered sub-inhibitory. Subinhibitory concentrations of antibiotics can generate a selection pressure in favor of antibiotic resistance genes in bacterial populations, causing phenomena of resistance or tolerance to antibiotics, and can also induce diverse biological responses in bacteria, which perceive them as chemical substances

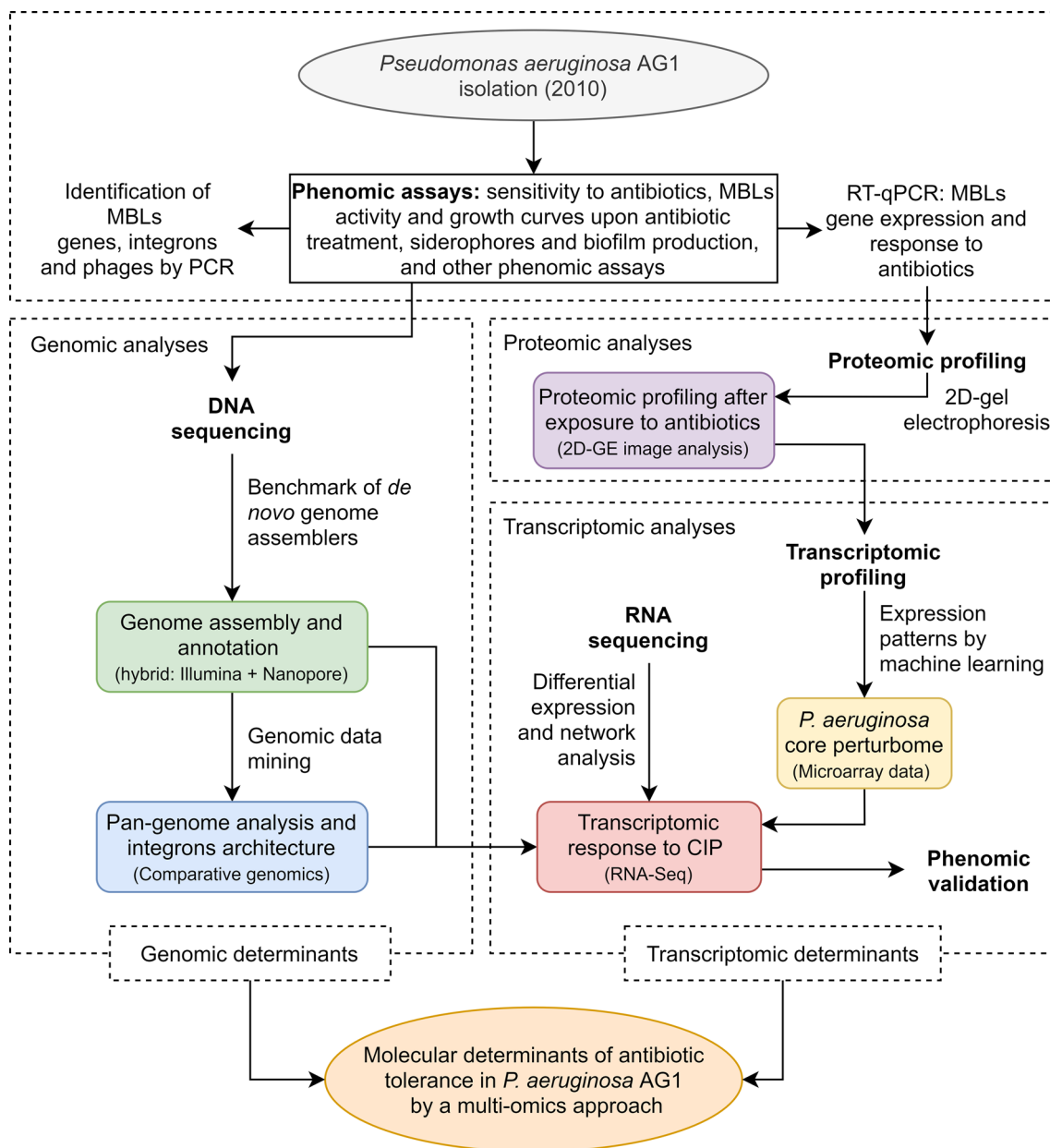


Fig. 1 General workflow to study molecular determinants of antibiotic tolerance in the high-risk *P. aeruginosa* AG1 (PaeAG1) by a multi-omics approach. Different studies have been performed to study PaeAG1. After isolation and determination of MIC of several antibiotics, distinct experimental assays were implemented to identify MBLs, integrons and phage genes (PCR), MBLs expression (RT-qPCR), as well as proteomic profiling to antibiotics (2D-GE). Then, the genome sequence was assembled and annotated, including the

comparison against other genomes to solve the architecture of the integrons. This defined the genomic determinants for PaeAG1. To study the transcriptome response to ciprofloxacin (CIP), a first analysis was done to identify the core perturbome or response to multiple perturbations in expression data. Finally, the transcriptomic determinants in response to CIP were identified using RNA-Seq, including the “exclusive” response not present in the perturbome

in the form of signals that can trigger different cellular responses (Molina Mora et al. 2021; Stockwell and Loper 2005), including alterations in motility patterns or biofilm production, characteristics that can cause an increase in their virulence (McFarland et al. 2015; Stewart et al. 2015). Thus, the use of sub-inhibitory concentration of antibiotics also

contributes to antibiotic resistance as it selects pre-existing resistant organisms and allows the strains to continue growing (McVicker et al. 2014). Since sub-inhibitory antibiotic concentrations are found in many natural environments, as well as in the course of prophylaxis and therapeutic use, bacteria can naturally trigger mechanisms of resistance

(Andersson and Hughes 2014). However, the fundamental mechanisms of bacterial response to antibiotics have not been fully elucidated (Stewart et al. 2015).

In this context, a decade of study of PaeAG1 has not only contributed to gain new insights about the antibiotic resistance in the *Pseudomonas* group, but also new questions have arisen, which motivates to continue studying this model.

***Pseudomonas aeruginosa* Group**

P. aeruginosa is an opportunist and versatile pathogen which is able to survive in a wide variety of environments (Klockgether et al. 2010). With a large genome (6–7.5 Mb), *P. aeruginosa* strains have a large proportion of the genome (> 8%) dedicated to regulatory functions (Cabot et al. 2016) resulting in a consequent diversity of metabolic capabilities and responses to stress. Because of these features, *P. aeruginosa* is responsible for infections among immunocompromised hosts (Lu et al. 2016), nosocomial infections (Fernández et al. 2018), and is a major cause of death among patients with cystic fibrosis (Kumar et al. 2019). Antibiotics such as β -lactams (cell wall synthesis inhibition, including carbapenems as a last-resort treatment, e.g., imipenem), quinolones (inhibition of DNA gyrase *gyrA*, e.g. ciprofloxacin), and aminoglycosides (inhibition of protein synthesis, e.g., tobramycin) antibiotics are used to treat *Pseudomonas* infections.

However, the treatment of *P. aeruginosa* infections is challenging due to its many intrinsic and acquired mechanisms of resistance (Toval et al. 2015), resulting in significant morbidity and mortality. According to the World Health Organization (WHO), resistance to carbapenems in *P. aeruginosa*, *Acinetobacter baumannii*, and Enterobacteriaceae family is considered a critical issue in the context of antibiotic resistance, being classified as Priority 1 group (World Health Organization 2017).

Regarding multi-omics approaches, only a few studies have used this strategy to study *P. aeruginosa*. For example, two closely related *P. aeruginosa* strains (ATCC 33988 and PAO1) grown in n-alkanes or glycerol were analyzed at the transcriptomic and proteomic levels, in which the lack of the activity of a quorum sensing (QS) system was reported as a key element in a better adaptation for growth and survival (Grady et al. 2017). Another study was focused on *P. aeruginosa* causing health-care-associated bloodstream infections to identify a pathogen-derived prognostic biomarker for patients at risk for unfavorable outcomes. The multi-omics approach was implemented using genome sequencing and proteomics as part of a clinical study (Willmann et al. 2018).

Recently, Lood and collaborators studied a jumbo bacteriophage PA5oct in *P. aeruginosa* PAO1 using genomics, transcriptomics, and proteomics. The study elucidated the genome organization and the response it elicits during

infection of its host, as well as the evolutionary relationships with other phages were found using a gene-sharing network analysis (Lood et al. 2020).

Regarding high-risk *P. aeruginosa* clones, associated with epidemics and multidrug resistance, most of them are studied at the genomic level (genotyping or genome sequencing), and, to our knowledge, there are no studies of high-risk *P. aeruginosa* strains using multi-omics approaches. In this sense, our multi-omics approach to study PaeAG1 is the first one for a high-risk clone.

***Pseudomonas aeruginosa* AG1 (PaeAG1)**

Diverse *P. aeruginosa* clones have been identified in several bacterial species that show particular traits associated with resistance to antibiotics and virulence, which stands out for being high-risk clones: their dissemination in hospital settings, their ability to cause infections, as well as the difficulty of their treatment. This phenomenon has been reported not only worldwide (Mulet et al. 2013a; Petitjean et al. 2017), but also in Costa Rica (Toval et al. 2015; Molina-Mora and García 2020; Molina-Mora et al. 2020a).

In Costa Rica, the isolation of carbapenem-resistant *P. aeruginosa* strains is relatively common in some major hospitals, up to 63.1% of prevalence, as previously reported (Toval et al. 2015), much higher than the frequencies observed in other countries (Hong et al. 2015). The Costa Rican strain PaeAG1 was identified as the first report of a *P. aeruginosa* isolate carrying both VIM-2 and IMP-18 genes encoding for MBLs enzymes, both with carbapenemase activity (Toval et al. 2015). Later, another isolate from the United Kingdom with the same enzymes was reported (Turton et al. 2015).

PaeAG1 was grown from a sputum sample of a patient from the Intensive Care Unit in the San Juan de Dios Hospital (San José, Costa Rica) in 2010. This strain has resistance to multiple antibiotics such as β -lactams (including

Table 1 Minimum Inhibitory Concentration (MIC) of several antibiotics against *P. aeruginosa* strains PAO1 (reference) and PaeAG1

Antibiotics	Break-point ($\mu\text{g}/\text{mL}$)	MIC ($\mu\text{g}/\text{mL}$)	
		PAO1	PaeAG1
Third-generation Cephalosporines (β -lactam)	Ceftazidime 32	1	512
Carbapenems (β -lactam)	Meropenem 8	0.250	32
	Imipenem 8	0.250	> 32
Aminoglycosides	Gentamicin 16	1	128
	Tobramycin 16	1	> 192
Quinolones	Ciprofloxacin 4	0.125	32
Polymyxins	Polymyxin B 8	2	1

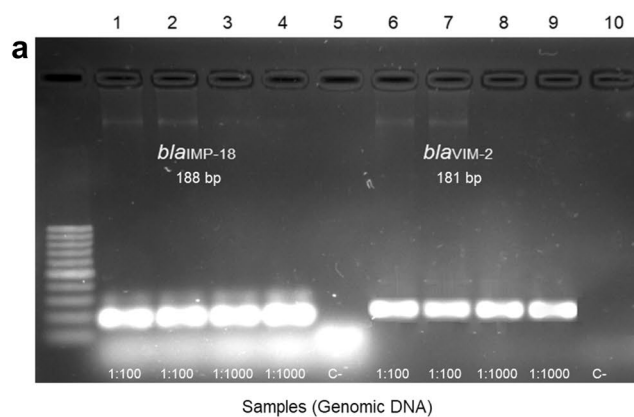
carbapenems), aminoglycosides, and fluoroquinolones, being only sensible to colistin (Table 1).

The first molecular analysis of the genes in PaeAG1 was done to identify MBLs (Fig. 2A) and phages genes by end-point PCR. Sanger sequencing (primer walking method) confirmed that VIM-2 and IMP-18 genes are encoded in class 1 integrons, which were registered as KC907378 and KC907377 in Genbank, as we reported previously (Toval et al. 2015). At the proteomic level (Fig. 2B, control with LB medium), periplasmic proteins were analyzed using 2-dimensional gel electrophoresis (2D-GE, see below for more details). In addition, at the phenomic level, preliminary comparison to the reference strain (*P. aeruginosa* PAO1)

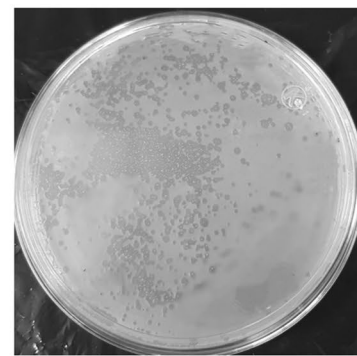
showed that PaeAG1 has particular features after exposure to different antibiotics (Fig. 2B), including pigment production, biofilm formation, phage plaque induction (Fig. 2C), and others in comparison to the reference strain *P. aeruginosa* PAO1 (Toval et al. 2015; Chinchilla 2018).

In this context, PaeAG1 represents an excellent study model for the analysis of the genomic, evolutionary, metabolic, virulence, and, especially, clinical characteristics of the clonal lineage ST-111 in the *P. aeruginosa* species. The deep knowledge of these traits in PaeAG1 establishes a platform for the analysis of alternatives to control its dissemination and the treatment of infections caused by this high-risk clone.

A PCR to detect MBLs



C Phage plaques



B Proteomic profiling

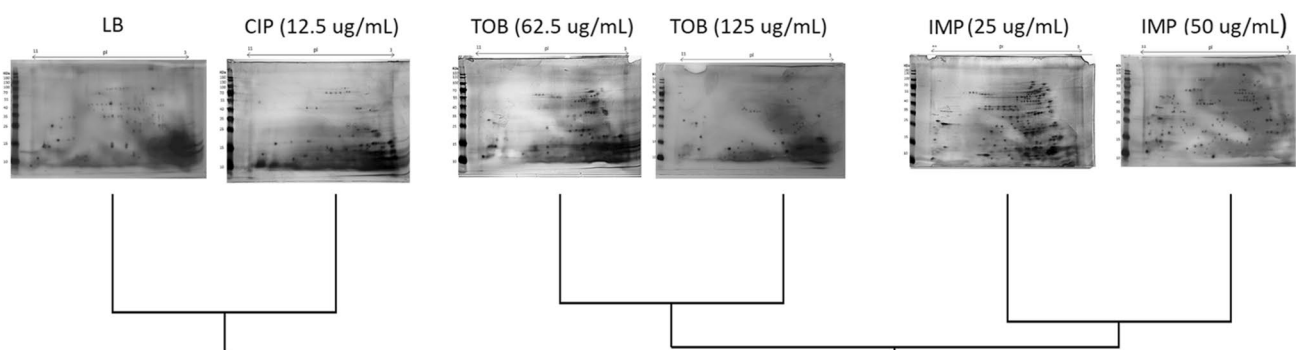


Fig. 2 Different experimental assays to study PaeAG1. **a** End-point PCR assays were standardized to identify genes of MBLs and other elements of the integrons, and phage genes (two genomic DNA concentrations were used). MBLs expression was also assessed using RT-qPCR (see Text). **b** In addition, the proteomic profile of PaeAG1

after exposure to antibiotics was evaluated using 2D-GE. Other assays included the phage plaques assay, which was used to validate the transcriptomic results regarding phage induction (see details below)

The Multi-omics Approach to Study PaeAG1

In view of the preliminary analysis of genomic and phenomic features of PaeAG1, we were interested in studying PaeAG1 in-depth using a multi-omics approach. To address this, the strategy was developed in five main steps (Fig. 1).

Genome Assembly and Annotation of PaeAG1

First, genome sequencing was done using short- (Illumina) and long-read (Oxford Nanopore Technology, ONT) technologies. Although a reference genome is available for the *P. aeruginosa* group (strain PAO1), a de novo strategy to

assemble the PaeAG1 genome was required, since it was initially estimated that PaeAG1 has ~ 1.0 Mb additional of DNA sequence in its genome.

As detailed in Molina-Mora et al. 2020a, a benchmark of non-hybrid (using a single DNA sequencing technology) and hybrid (using both short and long-read data) assemblers was required to select the optimum model. To make this possible, the 3C criterion (i.e., contiguity, completeness, and correctness) was conceptualized as a set of metrics that can be used to benchmark genome assemblies and select the best approach. Recently, we have used this strategy in another work for the genomic surveillance of SARS-CoV-2 in Costa Rica (Molina-Mora et al. 2021a).

The final assembly (GenBank: CP045739, 7,190,208 bp), using a hybrid approach, revealed that PaeAG1 has not only the expected gene content for the *P. aeruginosa* group

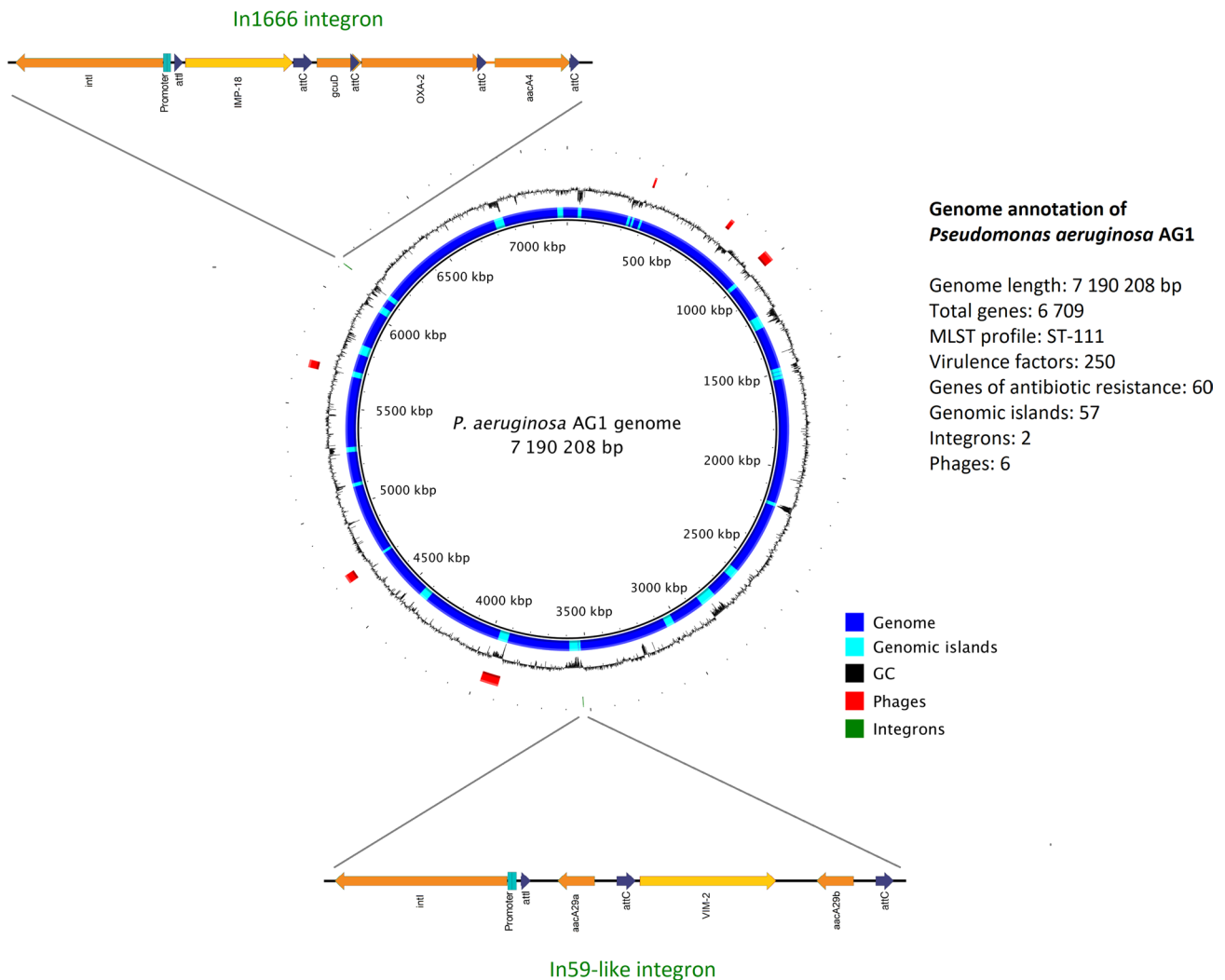


Fig. 3 Genomic determinants of antibiotic resistance in PaeAG1 genome. The genome assembly of PaeAG1 showed not only the expected virulence factors and many antibiotic resistance genes but

also 57 genomic islands harboring phages, integrons, and other antibiotic resistance genes. Using a comparative approach, the architecture of the two integrons was revealed

but also specific elements that are absent in the reference genome (Fig. 3): 57 genomic islands (corresponding to ~1.0 Mb DNA sequence and > 1000 genes) harboring the two complete class 1 integrons, six prophages, mobile genetic elements, and 250 virulence factors. Besides, PaeAG1 has 60 resistance genes, a not functional CRISPR-Cas system (which may explain the high content of genomic islands), and a molecular genotyping profile of a high-risk ST-111 strain.

These particular results are key components of the multi-omics approach with the subsequent analyses. If a mapping to the reference genome had been selected instead of a de novo assembly, the gene content of the extra 1.0 Mb DNA sequence (as part of the accessory genome) could not have been revealed. In this regard, two out of the other studies were focused on elements found in this accessory genome: (i) the two PaeAG1 integrons, and (ii) the role of phages in the response to ciprofloxacin. Importantly, integrons and phages are absent in the reference genome.

Comparative Genomics: Pan-Genome Analysis and Architecture of Integrons

To describe the landscape of the genomic regions associated with the two integrons of PaeAG1, a comparative genomic strategy was performed as a second main step in the multi-omics approach. It was first demonstrated that VIM-2 and IMP-18 are inducible genes under exposure to carbapenems using RT-qPCR (Molina-Mora et al. 2021b). We then described the phylogenetic relationships among all the complete genomes of *P. aeruginosa* strains using a pan-genome

analysis (Fig. 4A). This led to identify not only the core and the accessory genome for this group, but also other strains sharing the PaeAG1 genomic islands. Phylogenetically related strains were also classified as ST-111 clones, but a variant profile of the PaeAG1 genomic island content was found in other strains. ST-111 is a lineage that belongs to the high-risk group in *P. aeruginosa* (Oliver et al. 2015), which is frequently associated with epidemics where multi-drug resistance confounds treatment (Petitjean et al. 2017). Many *P. aeruginosa* high-risk clones carry genomic determinants of antibiotic resistance such as carbapenemases or extended-spectrum β -lactamases (Oliver et al. 2015). In databases, only 10 complete genome sequences of ST-111 strains (including PaeAG1) are available and, as shown in Fig. 4B, they are distributed worldwide.

Since PaeAG1 has special genomic features regarding antibiotic multi-resistance, with the carbapenemase activity by the VIM-2 and IMP-18 genes, the profile of genomic island content in phylogenetically related genomes was used to gain insights into the evolution and landscape of genomic regions around the MBL-carrying integrons of PaeAG1. Thus, specific genomic regions associated with the two integrons were reconstructed and characterized to compare the gene content and architecture in close genomes.

The genomic region associated with the VIM-2-carrying integron (identified as an In59-like element, INTEGRALL-database <http://integrall.bio.ua.pt/>) was completely found in the other two ST-111 strains, being considered as an old-acquaintance integron (Fig. 3, bottom). In the case of the IMP-18-carrying integron, the architecture and the surrounding genomic region had never been reported before.

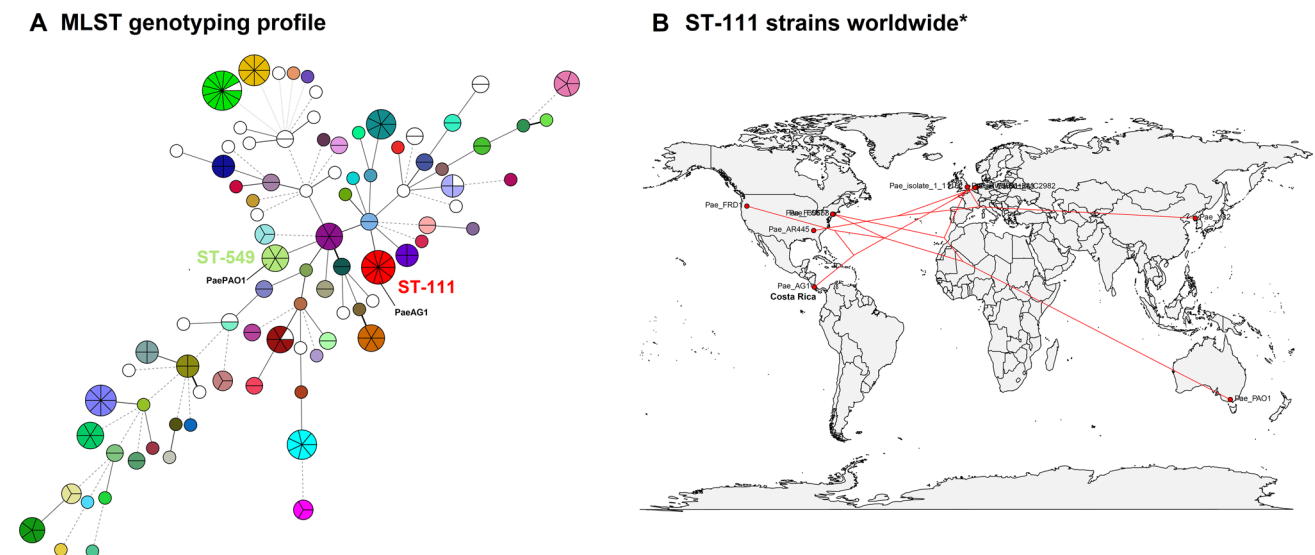


Fig. 4 Comparative analysis of all the *P. aeruginosa* complete genomes. The genome sequences separate strains by MLST genotyping profile, including a well-defined cluster for all the ten ST-111

strains, including PaeAG1 (**A**). The ten ST-111 genomes have been isolated from different places worldwide (**B**). *: The reference strain PaePAO1 (ST-549) was included in (**B**)

The IMP-18-carrying element was registered as a new integron In1666 in INTEGRALL-database (Fig. 3, up).

Jointly, the chromosome assembly and the comparative genomics were able to define the molecular arsenal of PaeAG1 at the genomic level, including multiple genomic determinants of virulence, mobile elements, and antibiotic resistance genes.

Proteomic Profile After Exposure to Antibiotics

On the other hand, in the context of antibiotic resistance, different assays have been performed in PaeAG1 to study its resistance to antibiotics. Antibiotic susceptibility testing is reported in Table 1 (Toval et al. 2015; Chinchilla 2018), and an MBLs' differential expression has been tested not only to carbapenems as demonstrated (Molina-Mora et al. 2021b) but also to other antibiotics (Chinchilla 2018).

At the proteomic level, the protein content in PaeAG1 under exposure to antibiotics was investigated. Two-dimensional gel electrophoresis (2D-GE) analysis was implemented using different imaging and machine learning algorithms (Molina-Mora et al. 2020b). Results revealed that the global proteomic profile after exposure to sub-inhibitory ciprofloxacin (CIP, quinolone) concentration remains close to control (LB medium, without antibiotics), contrasting with the results obtained with tobramycin (an aminoglycoside) and imipenem (a carbapenem) (Fig. 2B). This suggests that the effects of ciprofloxacin on PaeAG1 at the proteomic level are fewer than the changes given by other antibiotics. This is an interesting finding when we compare growth curves (see below and Fig. 6, left panels). Growth curves showed a particular concentration effect for PaeAG1 when exposed to sub-inhibitory CIP concentrations, but not to other tobramycin or imipenem antibiotics at sub-inhibitory concentrations. Thus, to investigate the association between the PaeAG1 growth and sub-inhibitory CIP concentrations, two main transcriptomic analyses were performed: i) the identification of core perturbome in the *P. aeruginosa* group and ii) transcriptomic profiling of PaeAG1 after exposure to CIP.

Central Response to Multiple Perturbations: Core Perturbome

In our work of the study of the molecular response to diverse perturbations (including CIP), term as perturbome, transcriptomic data of *P. aeruginosa* were used (microarray data of the reference strain PAO1 with multiple disturbances, GEO database) (Molina Mora et al. 2021). This makes it possible to generate the landscape of the central regulatory mechanisms of the stress response at the transcriptomic level in this bacterial group. Tolerance to stress conditions is vital for organismal survival, including bacteria

under diverse environmental conditions (such as antibiotics) (DeLong 2012). Thus, to identify the core perturbome of *P. aeruginosa*, a machine learning approach was implemented to recognize gene expression patterns among public transcriptomic data sets, similar to other studies (Ma et al. 2014; Zhao et al. 2016; Cornforth et al. 2018; Glaab et al. 2012). In this regard, only a few studies have used machine learning methods on biological data to describe the effects of multiple perturbations in complex biological systems (Birmingham et al. 2015; Caldera et al. 2019) and so far none in *P. aeruginosa*.

In a subsequent analysis, the specific case of CIP exposure was used to standardize a systems biology pipeline to build large-scale molecular networks (Molina-Mora et al. 2018).

The analysis of the central molecular response to perturbations, by both machine learning and large-scale networks, showed that the stress response is pleiotropic in *P. aeruginosa*, composed of genes belonging to energy metabolism, ribosomal activity, response to stimuli, and DNA metabolism (Fig. 5, circle). Specific effects on gene networks were reflected as changes in gene expression profiles and the complexity of molecular regulation (Molina-Mora et al. 2018).

Transcriptomic Response in PaeAG1 After Exposure to Ciprofloxacin

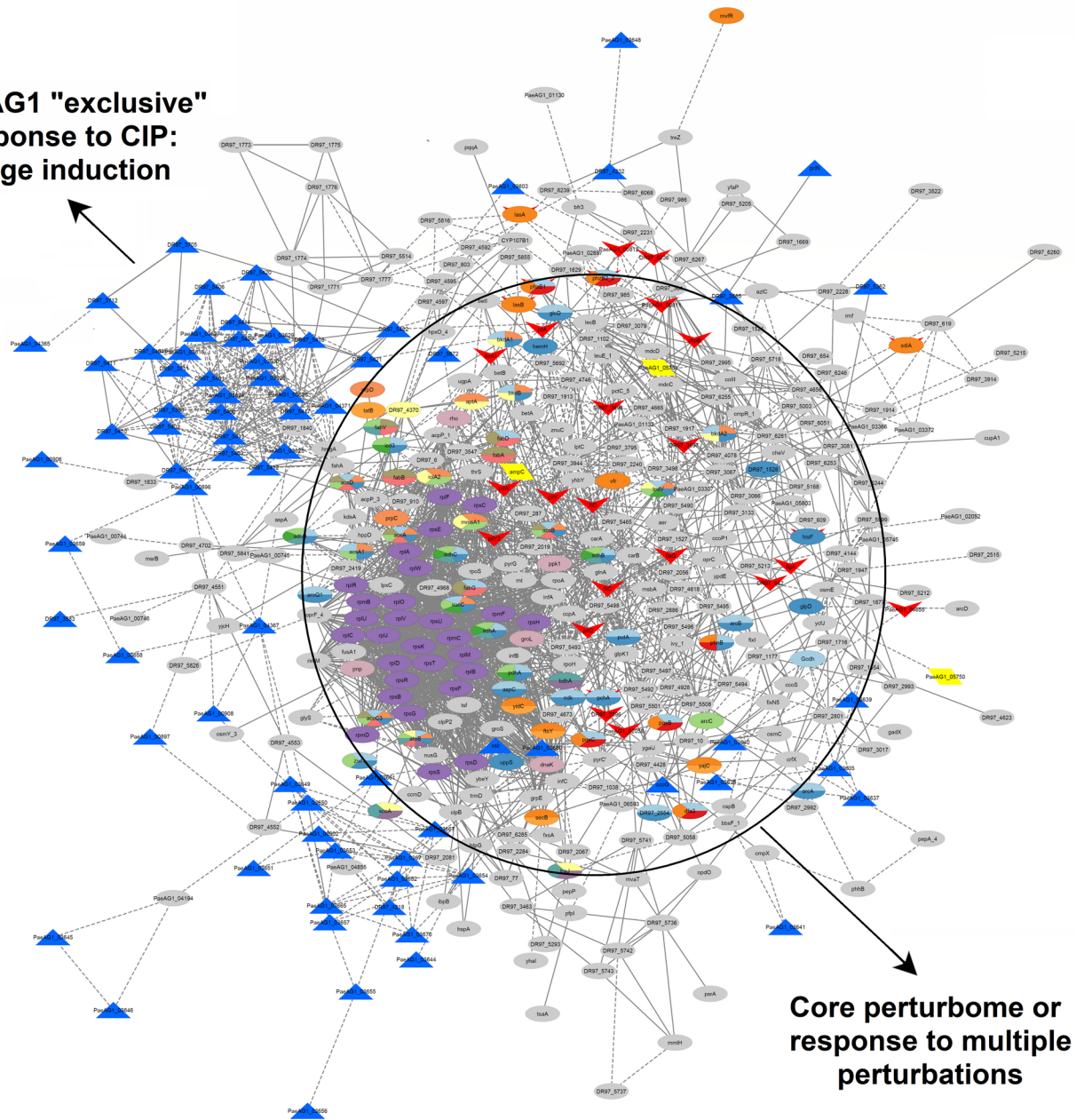
With the identification of the landscape of the core perturbome for *P. aeruginosa*, the study was resumed with the particular response to CIP in PaeAG1, as the final main step of the multi-omics study. Analyses are detailed in a previous study (Molina-Mora et al. 2020c).

The knowledge of the core perturbome was necessary to differentiate the pathways and responses that are shared by other perturbations, but more importantly, to identify the “exclusive” responses to CIP in PaeAG1. As detailed before, growth reduction was evidenced for this strain as sub-inhibitory CIP concentrations were increased. Thus, we identified the transcriptomic determinants associated with the response to CIP in PaeAG1. To address this, we used transcriptomic profiling by RNA sequencing and network analysis by applying a top-down system biology approach.

Transcriptomic determinants included classical elements of the core perturbome for *P. aeruginosa* with down-regulation of pathways related to energy metabolism, ribosomal activity, and DNA metabolism, most of them related to bacterial growth reduction. Also, an exclusive feature, the phage induction, was suggested due to the up-regulation of phage genes creating two well-defined clusters at a network level (Fig. 5).

To validate CIP effects on phage induction, we applied a phage plaque assay (at a phenomic level, Fig. 2C) that

PaeAG1 "exclusive" response to CIP: phage induction



KEGG pathways

- Biosynthesis of antibiotics
- Biosynthesis of secondary metabolites
- Butanoate metabolism
- Carbon metabolism
- Citrate cycle (TCA cycle)

- Fatty acid biosynthesis
- Fatty acid metabolism
- Phenazine biosynthesis
- Propanoate metabolism
- Protein export

- Quorum sensing
- RNA degradation
- Ribosome
- Synthesis and degradation of ketone bodies
- Valine, leucine and isoleucine degradation

Other elements

- ▲ Phages
- ▼ Virulence factors
- ▲ Antibiotic resistance

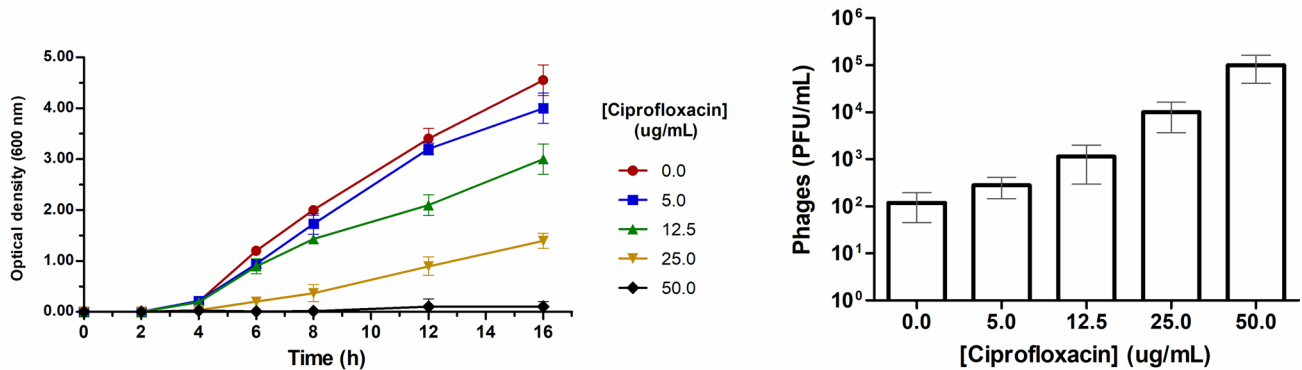
Fig. 5 Transcriptomic network of PaeAG1 after CIP exposure. Different transcriptomic determinants (hub genes, gene clusters, and pathways) are induced, including elements of the central response to multiple perturbations or core perturbome (metabolism, ribosomal

activity, RNA degradation, other), but a “more exclusive” response of phage induction (blue triangles) was identified. The Cytoscape network file is available as the “Supplementary_material-Transcriptomic_network.cys”

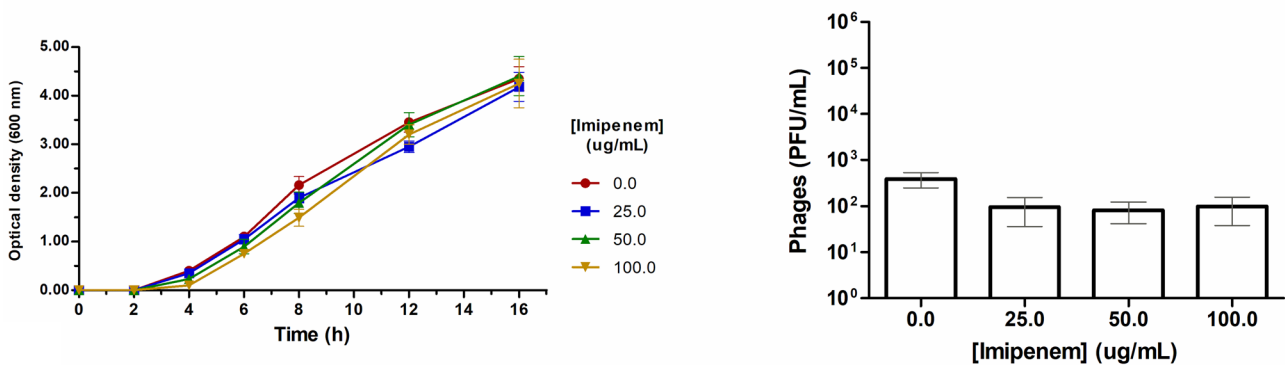
showed an exponential induction as CIP was increased. Since these phages are absent in the reference genome, again, the de novo genome assembly was a critical step

to obtain biological insights for PaeAG1. Although PaeAG1 is resistant to CIP, a sub-inhibitory concentration of this antibiotic can induce a pleiotropic effect at a

A Exposure to ciprofloxacin (MIC 32 mg/mL)



B Exposure to imipenem (MIC >32 mg/mL)



C Exposure to tobramycin (MIC >192 mg/mL)

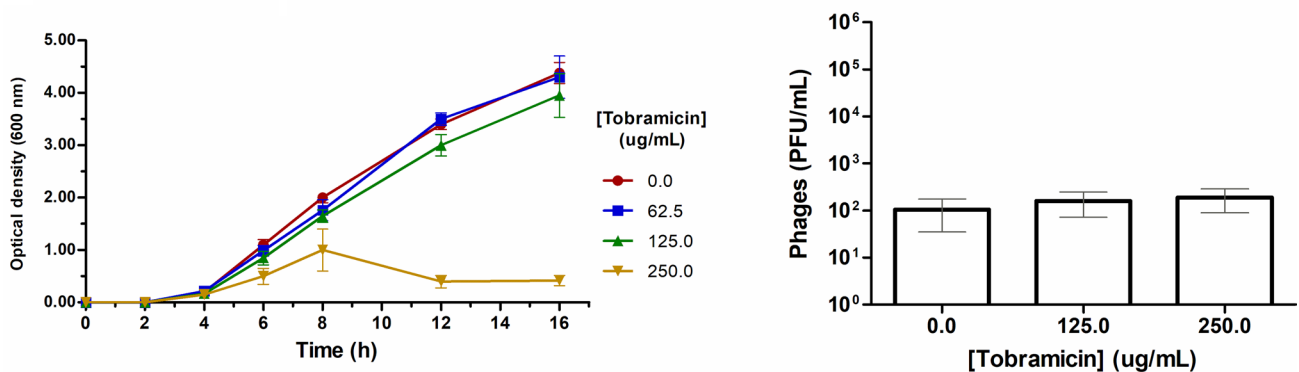


Fig. 6 Growth curves and phage plaque assay of PaeAG1 exposed to antibiotics. For CIP but not for other antibiotics, sub-inhibitory antibiotic concentrations showed decay in the growth as the concentration was increased. The transcriptomic analysis suggested phage induction as a key determinant in the response to CIP (A), which was

demonstrated using the phage plaque assay. This phenomenon was not observed for other antibiotics (imipenem nor tobramycin, B–C respectively). More details in our previous work (Molina-Mora et al. 2020c)

transcriptomic level, including pathways of the core perturbome and phage induction (Fig. 5). In the last case, with the subsequent bacterial cell lysis, the reduction on

the growth curve is explained by CIP in a concentration-dependent manner. This phenomenon is particular to CIP and not found for imipenem or tobramycin (Fig. 6).

Phage induction by CIP in PaeAG1 can be used as a complementary strategy to fight *Pseudomonas* infections. The fact that PaeAG1 phages are resident elements of the genome and not exogenous elements as in other studies (Kamal and Dennis 2015; Fothergill et al. 2011) represents an advantage to eventual further implementations. Future studies are required to evaluate the modulation of the CIP response using genetic engineering (knock-out, knock-down, and the like), other—omics approaches (proteomics, ChIP-Seq, etc.), and in vivo models.

Using our results and previous work, we defined that the response to CIP in PaeAG1 is complex (Fig. 7). First, the mechanism of action of CIP induces a DNA damage response, in which repressor genes of the SOS pathways or phage induction are modulated. Besides, the physiological consequences of the CIP by a general stress response are orchestrated in part by RpoS, with specific consequences on metabolism, protein synthesis, and regulation of virulence factors. Finally, these determinants influence other cell processes which can have different effects on the cell metabolism and the profile of resistance to antibiotics.

Integration and Conclusions

Antibiotic resistance is a major threat to public health because of its continuous emergence, worldwide spread, and increasing prevalence (Hong et al. 2015). Unlike highly host-adapted pathogens and symbionts undergoing genome reduction, as a versatile environmental organism, *P. aeruginosa* continually expands its genomic repertory (Mathee et al. 2008). With a high-risk ST-111 profile, PaeAG1 is a critical organism given its resistance to multiple antibiotics, including carbapenems (World Health Organization 2017).

Using a multi-omics approach, it was able to study molecular determinants of antibiotic resistance in PaeAG1. Genome assembly using a benchmark strategy led to building a high-quality sequence. A de novo approach allowed assembling around 1.0 Mb of sequence that is absent in the reference genome. These exclusive regions are composed of 57 genomic islands harboring two MBL-carrying integrons, phages, and many other genes. Comparison to all available complete sequences showed that the genome could be grouped by MLST profile, including a clear ST-111 cluster containing PaeAG1. In addition, a landscape of genomic regions surrounding integrons was

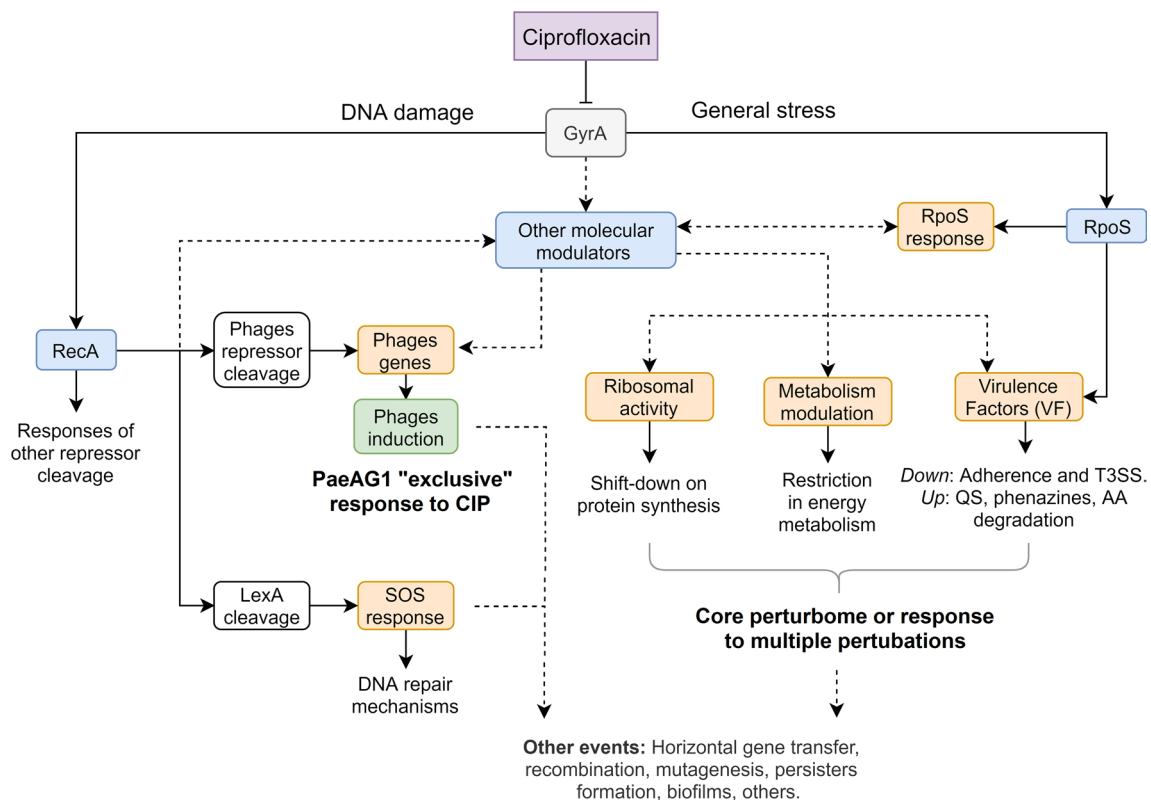


Fig. 7 Transcriptomic determinants of PaeAG1 in response to CIP. After exposure to CIP, different response to DNA damage and general stress are induced. This includes different pathways of the central response to multiple perturbations or core perturbome, as well as

particular pathways in response to CIP, including the SOS response (well known to be induced by CIP in *P. aeruginosa* group and other bacteria) and the “exclusive” phage induction in PaeAG1. Image was modified from our previous work (Molina-Mora et al. 2020c)

described in which an IMP-18-carrying integron was characterized for the first time. Multi-resistance profile, antibiotic resistance genes, the MLST profile, clusters of the pan-genome analysis, and the architecture of integrons the evolutionary history of the genome of PaeAG1.

To study the central response to perturbations in the *P. aeruginosa* group, the core perturbome, and to identify gene expression patterns, we used a machine learning approach. Pathways of energy metabolism, ribosomal activity, DNA metabolism, and others were enriched. Similar findings of enriched pathways were obtained for the specific case of PaeAG1 exposed to CIP, but particular genes (absent in the reference strain, such as phage genes) were also identified. Phage induction upon CIP treatment, suggested by phage genes up-regulation, was validated at a phenomic level. Particular key genes, gene clusters, and pathways were recognized as transcriptomic determinants of antibiotic resistance in PaeAG1.

Together, these genomic and transcriptomic elements are molecular determinants of antibiotic resistance in PaeAG1. This is particularly relevant for critical clones with the ability to conquer nosocomial environments and to develop a multi-resistance profile. As has been suggested, the biological markers of high-risk clones could be useful for the future design of specific treatments and infection control strategies (Mulet et al. 2013b). Thus, to study the implications of these genomic and transcriptomic determinants in PaeAG1, more detailed analyses are needed and we are planning to continue with other strategies, such as different levels of molecular regulation, other expression analyses (including proteomic level), other stress conditions to define the perturbome, genetic and phenotypic variability, validation of the effect and power of hub genes, modeling molecular circuits, explorations of the relationship between the presence of specific virulence traits and severity, in vivo models, and phage induction as a potential therapy to overcome resistance.

Finally, as shown here, the study of the molecular determinants in PaeAG1 was possible thanks to the integration of sequencing data, phenotypes, and bioinformatics pipelines. Because of the data complexity and results depending on algorithms, benchmarking strategies were required to analyze the data and to select the best protocols according to different criteria. Although we studied a bacterial genome (small in comparison to eukaryotic models), high-performance computational infrastructure was necessary mainly for comparative genomic and transcriptomic analyses. Besides, isolation and antibiotic resistance profiling, genome and RNA sequencing, as well as proteomic and other phenomic assays have been implemented for the last 10 years to study this bacterial model, implying a high cost. All these considerations remind us that these types of projects demand high-performance computational infrastructure,

best bioinformatics practices, and investment in scientific research in general.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s43657-021-00016-z>.

Acknowledgements We thank all members of Centro de Investigación en Enfermedades Tropicales (Universidad de Costa, Costa Rica) and all collaborators (researchers, students, and national and international institutions) for their support and work during 10 years studying this bacterial model.

Author Contributions JMM and FG participated in the conception, design, and data integration for this review. JMM drafted the manuscript. Both authors were involved in its revision and the final approbation.

Funding This work was funded by projects B8114, B8152, B7124, B5113, B4658, B4657, B4504, B4094, B4059, and B4051, Vicerrectoría de Investigación, Universidad de Costa Rica (period 2010-2020).

Data Availability The annotated final assembly of the PaeAG1 chromosome was deposited in GenBank under the accession number CP045739. Short-read and long-read raw data were uploaded to the NCBI Sequence Read Archive (SRA) with the accessions numbers SRX7088413 and SRX7088414, respectively. The RNA-seq raw data and processed files of transcripts quantification are available at the NCBI Gene Expression Omnibus (GEO) database under accession number GSE139866. Annotation tables, processed files, and other documents are available at <https://github.com/josemolina6/>.

Declarations

Conflict of Interest No competing interests to declare.

References

- Andersson DI, Hughes D (2014) Microbiological effects of sublethal levels of antibiotics. *Nat Rev Microbiol* 12:465–478. <https://doi.org/10.1038/nrmicro3270>
- Bermingham ML, Pong-Wong R, Spiliopoulou A, Hayward C, Rudan I, Campbell H, Wright AF, Wilson JF, Agakov F, Navarro P, Haley CS (2015) Application of high-dimensional feature selection: evaluation for genomic prediction in man. *Sci Rep* 5:1–12. <https://doi.org/10.1038/srep10312>
- Berti AD, Hirsch EB (2020) Tolerance to antibiotics affects response. *Science* (80-) 367:141–142
- Brauner A, Fridman O, Gefen O, Balaban NQ (2016) Distinguishing between resistance, tolerance and persistence to antibiotic treatment. *Nat Rev Microbiol* 14:320–330
- Brazas MD, Brazas MD, Hancock REW, Hancock REW (2005) Ciprofloxacin induction of a susceptibility determinant in *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother* 49:3222–3227. <https://doi.org/10.1128/AAC.49.8.3222>
- Cabot G, Zamorano L, Moyà B, Juan C, Navas A, Blázquez J, Oliver A (2016) Evolution of *Pseudomonas aeruginosa* antimicrobial resistance and fitness under low and high mutation rates. *Antimicrob Agents Chemother* 60:1767–1778. <https://doi.org/10.1128/AAC.02676-15.Address>
- Caldera M, Müller F, Kaltenbrunner I, Licciardello MP, Lardeau CH, Kubicek S, Menche J (2019) Mapping the perturbome network

- of cellular perturbations. *Nat Commun.* <https://doi.org/10.1038/s41467-019-13058-9>
- Chinchilla D (2018) Patrones de expresión de los genes de las metalo- β -lactamasas blaIMP-18 y blaVIM-2 e IMP-18 en la cepa *Pseudomonas aeruginosa* AG1 resistente a carbapenems. Tesis del Posgrado en Microbiología con énfasis en Bacteriología. Universidad de Costa Rica, San José
- Civelek M, Lusic AJ (2014) Systems genetics approaches to understand complex traits. *Nat Rev Genet* 15:34–48. <https://doi.org/10.1038/nrg3575>
- Cornforth DM, Dees JL, Ibberson CB, Huse HK, Mathiesen IH, Kirketerp-Møller K, Wolcott RD, Rumbaugh KP, Bjarnsholt T, Whiteley M (2018) *Pseudomonas aeruginosa* transcriptome during human infection. *Proc Natl Acad Sci U S A.* <https://doi.org/10.1073/pnas.1717525115>
- DeLong EF (2012) *Prokaryotes: prokaryotic physiology and biochemistry.* Springer, Berlin
- Farajzadeh Sheikh A, Shahin M, Shokoohzadeh L, Halaji M, Shahcheraghi F, Ghanbari F (2019) Molecular epidemiology of colistin-resistant *Pseudomonas aeruginosa* producing NDM-1 from hospitalized patients in Iran. *Iran J Basic Med Sci* 22:38–42. <https://doi.org/10.22038/ijbms.2018.29264.7096>
- Fernández M, Corral-Lugo A, Krell T (2018) The plant compound rosmarinic acid induces a broad quorum sensing response in *Pseudomonas aeruginosa* PAO1. *Environ Microbiol* 20:4230–4244. <https://doi.org/10.1111/1462-2920.14301>
- Firme M, Kular H, Lee C, Song D (2010) RpoS contributes to variations in the survival pattern of *Pseudomonas aeruginosa* in response to ciprofloxacin. *J Exp Microbiol Immunol* 14:21–27
- Fothergill JL, Mowat E, Walshaw MJ, Ledson MJ, James CE, Winstanley C (2011) Effect of antibiotic treatment on bacteriophage production by a cystic fibrosis epidemic strain of *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother* 55:426–428. <https://doi.org/10.1128/AAC.01257-10>
- Glaab E, Bacardit J, Garibaldi JM, Krasnogor N (2012) Using rule-based machine learning for candidate disease gene prioritization and sample classification of cancer gene expression data. *PLoS ONE.* <https://doi.org/10.1371/journal.pone.0039932>
- Grady SL, Malfatti SA, Gunasekera TS, Dalley BK, Lyman MG, Striebich RC, Mayhew MB, Zhou CL, Ruiz ON, Dugan LC (2017) A comprehensive multi-omics approach uncovers adaptations for growth and survival of *Pseudomonas aeruginosa* on n-alkanes. *BMC Genomics* 18:334. <https://doi.org/10.1186/s12864-017-3708-4>
- Hasin Y, Seldin M, Lusic A (2017) Multi-omics approaches to disease. *Genome Biol* 18:83. <https://doi.org/10.1186/s13059-017-1215-1>
- Hong DJ, Bae IK, Jang IH, Jeong SH, Kang HK, Lee K (2015) Epidemiology and characteristics of metallo- β -lactamase-producing *Pseudomonas aeruginosa*. *Infect Chemother* 47:81–97. <https://doi.org/10.3947/ic.2015.47.2.81>
- Kamal F, Dennis JJ (2015) Burkholderia cepacia complex phage-antibiotic synergy (PAS): antibiotics stimulate lytic phage activity. *Appl Environ Microbiol* 81:1132–1138. <https://doi.org/10.1128/AEM.02850-14>
- Klockgether J, Munder A, Neugebauer J, Davenport CF, Stanke F, Larbig KD, Heeb S, Schöck U, Pohl TM, Wiehlmann L, Tümmler B (2010) Genome diversity of *Pseudomonas aeruginosa* PAO1 laboratory strains. *J Bacteriol* 192:1113–1121. <https://doi.org/10.1128/JB.01515-09>
- Kumar SS, Penesyan A, Elbourne LDH, Gillings MR, Paulsen IT (2019) Catabolism of nucleic acids by a cystic fibrosis *Pseudomonas aeruginosa* isolate: an adaptive pathway to cystic fibrosis sputum environment. *Front Microbiol* 10:1–14. <https://doi.org/10.3389/fmicb.2019.01199>
- Lood C, Danis-Włodarczyk K, Blasdel BG, Bin JH, Vandenheuvel D, Briers Y, Noben J, Noort V, Drulis-Kawa Z, Lavigne R (2020) Integrative omics analysis of <sc> </sc> *Pseudomonas aeruginosa* </sc> virus PA5oct highlights the molecular complexity of jumbo phages. *Environ Microbiol* 22:2165–2181. <https://doi.org/10.1111/1462-2920.14979>
- Lu P, Wang Y, Zhang Y, Hu Y, Thompson KM, Chen S (2016) RpoS-dependent sRNA RgsA regulates Fis and AcpP in *Pseudomonas aeruginosa*. *Mol Microbiol* 102:244–259. <https://doi.org/10.1111/mmi.13458>
- Ma C, Xin M, Feldmann KA, Wang X (2014) Machine learning-based differential network analysis: a study of stress-responsive transcriptomes in arabidopsis. *Plant Cell* 26:520–537. <https://doi.org/10.1105/tpc.113.121913>
- Mathee K, Narasimhan G, Valdes C, Qiu X, Matewish JM, Koehrsen M, Rokas A, Yandava CN, Engels R, Zeng E, Olavarietta R, Doud M, Smith RS, Montgomery P, White JR, Godfrey PA, Kodira C, Birren B, Galagan JE, Lory S (2008) Dynamics of *Pseudomonas aeruginosa* genome evolution. *Proc Natl Acad Sci* 105:3100–3105. <https://doi.org/10.1073/PNAS.0711982105>
- McFarland KA, Dolben EL, LeRoux M, Kambara TK, Ramsey KM, Kirkpatrick RL, Mougous JD, Hogan DA, Dove SL (2015) A self-lysis pathway that enhances the virulence of a pathogenic bacterium. *Proc Natl Acad Sci U S A* 112:8433–8438. <https://doi.org/10.1073/pnas.1506299112>
- McVicker G, Prajsnar TK, Williams A, Wagner NL, Boots M, Renshaw SA, Foster SJ (2014) Clonal expansion during staphylococcus aureus infection dynamics reveals the effect of antibiotic intervention. *PLoS Pathog.* <https://doi.org/10.1371/journal.ppat.1003959>
- Molina-Mora JA, Campos-Sanchez R, Garcia F (2018) Gene expression dynamics induced by ciprofloxacin and loss of lexA function in *Pseudomonas aeruginosa* pao1 using data mining and network analysis. In: 2018 IEEE international work conference on bioinspired intelligence (IWOBI). IEEE, pp 1–7
- Molina-Mora J-A, Campos-Sánchez R, Rodríguez C, Shi L, García F (2020a) High quality 3C de novo assembly and annotation of a multidrug resistant ST-111 *Pseudomonas aeruginosa* genome: benchmark of hybrid and non-hybrid assemblers. *Sci Rep* 10:1392. <https://doi.org/10.1038/s41598-020-58319-6>
- Molina-Mora JA, Chinchilla-Montero D, Castro-Peña C, García F (2020b) Two-dimensional gel electrophoresis (2D-GE) image analysis based on cell profiler. *Medicine (baltimore)* 99:e23373. <https://doi.org/10.1097/MD.00000000000023373>
- Molina-Mora JA, Chinchilla D, Chavarría M, Ulloa A, Campos-Sanchez R, Mora-Rodríguez RA, Shi L, García F (2020c) Transcriptional determinants of the response of ST-111 *Pseudomonas aeruginosa* AG1 to ciprofloxacin identified by a top-down systems biology approach. *Sci Rep* 10:1–23. <https://doi.org/10.1038/s41598-020-70581-2>
- Molina-Mora JA, Cordero-Laurent E, Godínez A, Calderón-Osorno M, Brenes H, Soto-Garita C, Pérez-Corrales C, Drexler JF, Moreira-Soto A, Corrales-Aguilar E, Duarte-Martínez F (2021a) SARS-CoV-2 genomic surveillance in Costa Rica: evidence of a divergent population and an increased detection of a spike T1117I mutation. *Infect Genet Evol* 92:104872. <https://doi.org/10.1016/j.meegid.2021.104872>
- Molina-Mora JA, Chinchilla-Montero D, García-Batán R, García F (2021b) Genomic context of the two integrons of ST-111 *Pseudomonas aeruginosa* AG1: a VIM-2-carrying old-acquaintance and a novel IMP-18-carrying integron. *Infect Genet Evol* 89:104740. <https://doi.org/10.1016/j.meegid.2021.104740>
- Molina-Mora JA, Garcia F (2020) The 3C criterion: contiguity, completeness and correctness to assess de novo genome assemblies. *BMC Bioinform Algorithms Appl* 21:5. <https://doi.org/10.1186/s12859-020-03838-2>
- Molina Mora JA, Montero-Manso P, García-Batán R, Campos-Sánchez R, Fernández JV, García F (2021) A first perturbome of *Pseudomonas aeruginosa*: identification of core genes related to

- multiple perturbations by a machine learning approach. Biosystems. <https://doi.org/10.1101/2020.05.05.078477>
- Mulet X, Cabot G, Ocampo-Sosa AA, Domínguez MA, Zamorano L, Juan C, Tubau F, Rodríguez C, Moyà B, Peña C, Martínez-Martínez L, Oliver A (2013a) Biological markers of *Pseudomonas aeruginosa* epidemic high-risk clones. *Antimicrob Agents Chemother* 57:5527–5535. <https://doi.org/10.1128/AAC.01481-13>
- Mulet X, Cabot G, Ocampo-Sosa AA, Domínguez MA, Zamorano L, Juan C, Tubau F, Rodríguez C, Moyà B, Peña C, Martínez-Martínez L, Oliver A, Spanish Network for Research in Infectious Diseases (REIPI) (2013b) Biological markers of pseudomonas aeruginosa epidemic high-risk clones. *Antimicrob Agents Chemother* 57:5527–5535. <https://doi.org/10.1128/AAC.01481-13>
- O'Donnell ST, Ross RP, Stanton C (2020) The progress of multi-omics technologies: determining function in lactic acid bacteria using a systems level approach. *Front Microbiol* 10:3084. <https://doi.org/10.3389/fmicb.2019.03084>
- O'Neill J (2016) Review on antimicrobial resistance - tackling drug-resistant infections globally: final report and recommendations. Review on antimicrobial resistance
- Oliver A, Mulet X, López-Causapé C, Juan C (2015) The increasing threat of *Pseudomonas aeruginosa* high-risk clones. *Drug Resist Update* 21–22:41–59. <https://doi.org/10.1016/j.drug.2015.08.002>
- Petitjean M, Martak D, Silvant A, Bertrand X, Valot B, Hocquet D (2017) Genomic characterization of a local epidemic *Pseudomonas aeruginosa* reveals specific features of the widespread clone ST395. *Microb Genomics* 3:e000129. <https://doi.org/10.1099/mgen.0.000129>
- Stewart PS, Franklin MJ, Williamson KS, Folsom JP, Boegli L, James GA (2015) Contribution of stress responses to antibiotic tolerance in *Pseudomonas aeruginosa* biofilms. *Antimicrob Agents Chemother* 59:3838–3847. <https://doi.org/10.1128/AAC.00433-15>
- Stockwell VO, Loper JE (2005) The sigma factor RpoS is required for stress tolerance and environmental fitness of *Pseudomonas fluorescens* Pf-5. *Microbiology* 151:3001–3009. <https://doi.org/10.1099/mic.0.28077-0>
- Subramanian I, Verma S, Kumar S, Jere A, Anamika K (2020) Multi-omics data integration, interpretation, and its application. *Bioinform Biol Insights* 14:1177932219899051. <https://doi.org/10.1177/1177932219899051>
- Toval F, Guzmán-Martínez A, Madriz V, Somogyi T, Rodríguez C, García F (2015) Predominance of carbapenem-resistant *Pseudomonas aeruginosa* isolates carrying blaIMP and blaVIM metallo- β -lactamases in a major hospital in Costa Rica. *J Med Microbiol* 64:37–43. <https://doi.org/10.1099/jmm.0.081802-0>
- Turton JF, Wright L, Underwood A, Witney AA, Chan YT, Al-Shahib A, Arnold C, Doumith M, Patel B, Planche TD, Green J, Holliman R, Woodford N (2015) High-resolution analysis by whole-genome sequencing of an international lineage (Sequence Type 111) of pseudomonas aeruginosa associated with metallo-carbapenemases in the United Kingdom. *J Clin Microbiol* 53:2622–2631. <https://doi.org/10.1128/JCM.00505-15>
- Willmann M, Götting S, Bezdán D, Maček B, Velic A, Marschal M, Vogel W, Flesch I, Markert U, Schmidt A, Kübler P, Haug M, Javed M, Jentzsch B, Oberhettinger P, Schütz M, Bohn E, Sonabend M, Klein K, Autenrieth IB, Ossowski S, Schwarz S, Peter S (2018) Multi-omics approach identifies novel pathogen-derived prognostic biomarkers in patients with *Pseudomonas aeruginosa* bloodstream infection. *bioRxiv* 309898
- Woodford N, Turton JF, Livermore DM (2011) Multiresistant Gram-negative bacteria: the role of high-risk clones in the dissemination of antibiotic resistance. *FEMS Microbiol Rev* 35:736–755. <https://doi.org/10.1111/j.1574-6976.2011.00268.x>
- World Health Organization (2017) Guidelines for the prevention and control of carbapenem-resistant Enterobacteriaceae, Acinetobacter baumannii and *Pseudomonas aeruginosa* in health care facilities. Geneva
- Zhao W, Chen JJ, Perkins R, Wang Y, Liu Z, Hong H, Tong W, Zou W, Metzker M, Didelot X, Bowden R, Wilson D, Peto T, Crook D, Koser C, Holden M, Ellington M, Cartwright E, Brown N, Ogilvy-Stuart A, Hsu L, Chewapreecha C, Croucher N, Harris S, Lienau E, Strain E, Wang C, Zheng J, Ottesen A, Keys C, Hammack T, Musser S, Brown E, Allard M, Petrosino J, Highlander S, Luna R, Gibbs R, Versalovic J, Radford A, Chapman D, Dixon L, Chantrey J, Darby A, Hall N, Koboldt D, Steinberg K, Larson D, Wilson R, Mardis E, Zhang J, Chiodini R, Badr A, Zhang G, Hofmann T, Blei D, Ng A, Jordan M, Griffiths T, Steyvers M, Blei D, Jordan M, Datta R, Joshi D, Li J, Wang J, Hofmann T, Papadimitriou C, Tamaki H, Raghavan P, Vempala S, Jordan M, Shivashankar S, Srivathsan S, Ravindran B, Tendulkar A, Bisgin H, Liu Z, Kelly R, Fang H, Xu X, Tong W, Zhang R, Cheng Z, Guan J, Zhou S, Macnab R, Hamming R, Grimont P, Weill F, Allard M, Luo Y, Strain E, Li C, Keys C, Son I, Stones R, Musser S, Brown E, Zhao W, Chen J, Foley S, Wang Y, Zhao S, Basinger J, Zou W, Zhang Z, Schwartz S, Wagner L, Miller W, Allard M, Luo Y, Strain E, Pettengill J, Timme R, Wang C, Li C, Keys C, Zheng J, Stones R, Edgar R, Higgins D, Sharp P, Mardia K, Kent J, Bibby J, Salton G, Wong A, Yang C, Zhao W, Zou W, Chen J, Hubert L, Arabie P, Vapnik V, Breiman L, Blei D, Kauffmann F, Edwards P, Liebana E, Guns D, Garcia-Migura L, Woodward M, Clifton-Hadley F, Davies R, Gaul S, Wedel S, Erdman M, Harris D, Harris I, Ferris K, Hoffman L, Zou W, Lin W, Foley S, Chen C, Nayak R, Chen J, Zou W, Lin W, Hise K, Chen H, Keys C, Chen J, Guard J, Sanchez-Ingunza R, Morales C, Stewart T, Liljebjelke K, Kessel J, Ingram K, Jones D, Jackson C, Fedorka-Cray P, Achtman M, Wain J, Weill F, Nair S, Zhou Z, Sangal V, Krauland M, Hale J, Harbottle H, Uesbeck A, Ranieri M, Shi C, Switt AM, Bakker H, Wiedmann M, Naccache S, Federman S, Veeraraghavan N, Zaharia M, Lee D, Samayoa E, Bouquet J, Greninger A, Luk K, Enge B, Pettengill J, Timme R, Barrangou R, Toro M, Allard M, Strain E (2016) A novel procedure on next generation sequencing data analysis using text mining algorithm. *BMC Bioinform* 17:213. <https://doi.org/10.1186/s12859-016-1075-9>