Genomics and the evolution of antibiotic resistance

Michael R. Gillings,1 Ian T. Paulsen,2 and Sasha G. Tetu2

1Department of Biological Sciences. 2Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, Australia

Address for correspondence: Michael R. Gillings, Department of Biological Sciences, Macquarie University, Sydney, NSW 2019, Australia. michael.gillings@mq.edu.au

Antibiotic resistance arises as a consequence of complex interactions among genes, mobile elements, and their bacterial hosts, coupled with the intense selection pressures imposed by humans in an attempt to control bacterial growth. Understanding the evolution of resistance requires an understanding of interacting cellular and genetic components. Here, we review how DNA analysis has helped reconstruct the origins of the mosaic, multiresistant mobile elements that have spread through pathogens in the last 60 years. This history helps inform the future, such that resistance might be better managed. Whole-genome sequencing has great potential for epidemiological tracking and for understanding the development of resistance via experimental evolution. DNA analysis also offers the opportunity for constructing databases that record genes of interest, the mobile elements that move these genes, and the cells or species that acquire such genes. Linking these DNA elements to their human and animal hosts and to the environments where they occur should help us establish a more robust ecological and evolutionary framework for controlling and managing resistance. Such efforts need to be well coordinated because, like many other issues that face humanity, antibiotic resistance is a global problem that requires global solutions.

Keywords: integron; SXT ICE; evolution; resistome; transposon; database

Introduction

The rapid appearance and spread of antibiotic resistance in bacteria over the last 80 years is one of the best examples of evolution by natural selection.1 Understanding the genes and mobile DNA elements that drove this phenomenon was instrumental in the development of modern molecular biology and DNA analysis.2 It is perhaps then fitting that this revolution has come full circle, and the power of genome sequencing is increasingly being applied to understand the past, survey the present, and predict the future of antibiotic resistance.3 In this review, we examine the potential uses of DNA sequence analysis to understand and manage the evolution of antibiotic-resistant organisms.

Antibiotic resistance is often referred to as emerging4 from a “reservoir,” usually identified as some nonhuman environmental compartment.5 While such metaphors can be useful, care needs to be taken that they continue to be recognized as just metaphors, since their uncritical use can obscure evolutionary understanding.6 For instance, clinicians often suggest that contemporary resistance is imported from elsewhere or is caused by others.7 Further, the use of the reservoir concept implies that resistance genes are in storage, waiting to confer inconvenient phenotypes.

Put simply, the root cause of resistance is strong selection for rare genetic events that co-opt or alter existing genes whose original function may have had nothing to do with the particular antibiotics used in clinical practice. Two kinds of rare event are involved: incremental changes to existing genes by mutation or co-option of existing genes for a new function. The former normally arises within a cell lineage, while the latter often occurs via lateral gene transfer between cells of different lineages.1 Both of these mechanisms can be fruitfully investigated using DNA technology.

During investigations of modern resistance phenomena, we must also remember that the past is a foreign country.8 We live in a world where gene
frequencies, species abundance, ecosystem functions, and biogeochemistry have all been profoundly altered by human activity. Consequently, the data we accumulate using DNA technology have to be considered against the background of recent global changes to all components of the biosphere. Nevertheless, we can attempt to reconstruct some of the key events that were instrumental in the rapid dissemination of antibiotic-resistance phenotypes among bacteria of relevance to human welfare. In performing these reconstructions, we can gain some insight into the fundamental drivers of resistance evolution, with the aim of controlling such processes in the future.

**Antibiotics and resistance genes in the pre-antibiotic era**

Antibiotics are not a natural chemical group, but rather they are a diverse set of molecules with the common property that they inhibit bacterial growth at high concentrations. They are often naturally occurring compounds and are members of a large cohort of small, bioactive molecules called secondary metabolites. Various bioactive molecules can function in cell-to-cell communication, regulating genetic and metabolic activities. Consequently, they can have biological effects at much lower concentrations than those used in antibiotic therapy.

The molecules that we now use as antibiotics may never reach clinical concentrations in natural environments, and we should consider antibiotic resistance against the background of the likely original function of these molecules. Viewing antibiotics as secondary metabolites can help to inform our thinking on resistance. Secondary metabolites are synthesized and/or degraded in biochemical pathways, can be exported or imported by membrane proteins, and have cellular targets or receptors. The genes involved in each of these various processes could potentially be co-opted or manipulated to interact with the antibiotic compound and are thus nascent resistance genes.

Given the number of bacterial species in the biosphere and the diversity of secondary metabolites they produce, it is not surprising that bacteria encode an enormous number of genes that are involved in producing or responding to small molecules. Our understanding and knowledge of this gene diversity has expanded dramatically with the widespread use of genome and metagenome sequencing. As outlined above, a proportion of these genes could be co-opted by human pathogens to confer resistance to the compounds we use as antibiotics. This pool of genes that could potentially confer an antibiotic resistance phenotype has been called the resistome. It is important to remember that the resistome is a purely anthropocentric concept, in that it focuses on a single phenotype that is of relevance to humans. The resistome encompasses diverse genes whose original function was probably not resistance in the clinical sense and also includes a wide range of genes for cellular targets that must undergo mutation before they can confer a resistance phenotype.

The resistome concept is useful because it highlights the fact that resistance phenotypes in human and animal pathogens can originate from environmental bacteria, and that resistance genes in clinical isolates are just a small proportion of the potential resistance genes in environmental compartments. Evidence from DNA sequencing shows that the resistome is far larger and much older than the entire corpus of clinically important resistance genes. Gene families capable of conferring resistance date back hundreds of millions of years and can be recovered from permafrost and ancient cave microbiomes. Diverse and ancient gene lineages that confer resistance to aminoglycosides, tetracycline, β-lactams, or glycopeptides can be readily detected using functional assays of metagenomic DNA. These specific properties were probably not the original function of these genes. Further, genes can be recovered from metagenomic DNA that confer resistance to clinically important antibiotics but are unrelated to the resistance genes found in pathogens, meaning they would not be detected by gene homology alone.

**The origins of resistance in human pathogens**

The fixation of resistance determinants in human pathogens and commensals is ultimately a direct result of human attempts to control bacterial growth. The key driver for this process is the strength of the selection pressure that humans apply to bacterial populations. This confers an advantage upon those rare cells in populations that can express a resistance phenotype via gene mutation or via the acquisition of new genes by lateral gene transfer.
Both gene mutation and lateral gene transfer are processes by which bacteria generate novelty and diversity.\textsuperscript{36} The underlying rates of these processes are likely to have remained largely unchanged for millions if not billions of years before the modern era. What has changed, particularly in the last century, is the intensity of selection brought to bear on bacterial populations. Before the human use of antimicrobial agents, relevant mutations or lateral gene transfers within individual cells would have conferred a relative and perhaps only incremental advantage.\textsuperscript{37} After humans began to use high concentrations of antimicrobial compounds, the advantage accruing to a cell with a recent, relevant mutation or lateral transfer event became absolute. That cell would survive, while all other cells would either die or be severely debilitated.

However, there are multiple requirements for such a successful selection event to occur. A mutation must occur in an appropriate gene, modifying its product sufficiently to confer a resistance phenotype. Alternatively, a cell could acquire a gene that confers resistance via the process of lateral gene transfer. This, in turn, requires proximity; the recipient cell might have to share an ecology or environmental compartment with the donor cell. The new gene product must be expressed and successfully integrated into cellular biochemistry in order to confer the resistance phenotype. Most importantly, these events must coincide in both time and space with an appropriate selection pressure.

Selection of a resistant cell occurs at a single place and time and involves an event that occurs randomly inside a single cell. To put this in perspective, it is useful to reflect on the size of the microbiome. The microbial world contains perhaps as many as 10\textsuperscript{12} species\textsuperscript{38} comprising some 1 × 10\textsuperscript{30} prokaryotic cells,\textsuperscript{23} which collectively hold an estimated 1.6 × 10\textsuperscript{37} nucleotides of DNA.\textsuperscript{39} The potential for interactions between these cells and the DNAs that they contain is unimaginably huge. Correspondingly, the frequency of an individual event in a single cell must be a vanishingly small proportion of the total number of events that are occurring at any one time in the microbial biosphere. Consequently, the genetic events that occur in coincidence with an appropriate selective force are extraordinarily rare and are essentially random.

This stochasticity has a number of consequences for investigating the rise of antibiotic resistance. First, the random nature of these events makes it difficult to predict the identities, locations, or nature of the genes and cells that will undergo successful selection for antibiotic resistance. Second, once a particular phenotype is selected and increases in distribution and abundance, it then becomes less likely that a second, independent event will become as successful, even if it confers the same phenotype, simply because it does not have a strong selective or numerical advantage over the first event. Third, each successful selection event results in a genetic bottleneck where a single conserved DNA sequence propagates and is, at least initially, confined to a clonal lineage of a single host species. These general principles can all be illustrated by retracing the genomic history of the class 1 integron, as reconstructed from sequencing data.

**Class 1 integrons and the rise of antibiotic resistance\textsuperscript{40}**

Integrons are a family of genetic elements found on the chromosomes of some 15% of genomesequenced bacterial species.\textsuperscript{41,42} They function to capture exogenous gene cassettes via site-specific recombination into an integron attachment site (att\textsuperscript{I}), an activity catalyzed by the integron-encoded integrase Int\textsuperscript{I}.\textsuperscript{5} The sequence diversity of Int\textsuperscript{I}, its wide phylogenetic distribution, and the diversity of individual arrays of captured gene cassettes demonstrates that integrons are ancient elements that have played important roles in bacterial adaptation and genome evolution. Their roles in adaptation rely on their ability to capture, rearrange, and express members of the vast, and largely uncharacterized, pool of integron gene cassettes.\textsuperscript{43–46}

The initial discovery of integrons\textsuperscript{47} was stimulated by the extraordinary success of one particular integron variant in disseminating antibiotic resistance genes among Gram-negative pathogens. This was called the class 1 integron. Class 1 integrons in clinical contexts have now collectively acquired over 130 resistance genes\textsuperscript{48} and have spread to the majority of Gram-negative pathogens of concern for both medicine and agriculture.\textsuperscript{5} Genomic and metagenomic analyses over the past decade have helped to reconstruct its likely origins and help explain why this particular genetic element has become so successful.

The class 1 integron that is now so widely distributed in pathogens is just one variant of a much
more diverse group of class 1 integrons found in natural environments. These class 1 integrons are a common feature of soil and aquatic bacteria, can occur in 5% of cells, and have some limited ability to transfer horizontally between species. They occur on the chromosomes of a range of non-pathogenic Betaproteobacteria, where they mainly carry gene cassettes of unknown function rather than known resistance genes. One gene family that is identifiable and common in the cassette arrays of environmental class 1 integrons is the versatile qac family of membrane efflux proteins. These characteristics of environmental class 1 integrons suggest that they were in the right place, and had the right properties, to interact with the human commensal microbiota when antimicrobial therapies were first used.

While the history of the clinical class 1 integron will never be known with certainty, we can reconstruct a plausible series of events that led to its appearance (Fig. 1). These events likely occurred as follows. A class 1 integron excised from the chromosome of a betaproteobacterial species and was then captured by a transposon of the Tn402 family, members of which are known from the human commensal flora. The captured integron probably carried a qacE gene cassette, encoding an efflux pump that conferred resistance to quaternary ammonium compounds. These qac cassettes are carried by about half of the environmental class 1 integrons and would have provided a selective advantage for fixation of the integron once it found its way into the human commensal flora. The Tn402 transposon specifically targets the res sites of plasmids, and this activity ensured that the transposon/integron hybrid would rapidly disseminate into a number of plasmid vectors. We can be fairly certain that the vast majority of class 1 integrons now found on plasmids in pathogens arose from this one ancestral event, because they all share a single, highly conserved integron–integrase gene sequence.

Genomic analyses have detected other hybrids between transposons and class 1 integrons that are circulating in the human commensal flora. In addition, the two other integrons of clinical relevance, the class 2 and class 3 integrons, are also thought to have arisen through the capture of chromosomal integrons by transposons. Consequently, it seems likely that interactions between chromosomal integrons and transposable elements are comparatively common. If this is the case, why should this one particular integron–transposon hybrid become so dominant?

It is said that fortune favors the well prepared, and this may be the case for the class 1 integron. This family of integrons is comparatively abundant in bacteria that occupy environments that intersect with the human food chain, and the class 1 integrons exhibit some mobility, making them available for insertion into transposons. Thus, they had
the means and opportunity to colonize the human commensal flora. They also commonly carry \textit{qacE} gene cassettes, giving them a selective advantage, because they could confer resistance to the disinfectants that were introduced before the antibiotic era.\textsuperscript{50,61} Their association with the Tn\textsubscript{402} transposon allowed them to spread rapidly into other mobile elements and thus to diverse bacterial hosts. Finally, the class 1 integron–integrase can capture gene cassettes from the chromosomal cassette arrays of other integron families, giving them access to diverse functions, including resistance genes.\textsuperscript{62,63} While the class 1 integron family exhibited a number of useful preadaptations for a role in disseminating resistance, it nevertheless appears that a great deal of chance was involved, particularly for the individual class 1 sequence variant that has now risen to such abundance in clinical contexts.

Once embedded in the human microbiota, the clinical class 1 integron underwent a rapid radiation, facilitated by the population size of humans and their domestic animals, and driven by the diverse, oscillating selection pressures imposed by antimicrobial agents. This rapid diversification can be followed using DNA analyses. The \textit{qacE} gene cassette was partially deleted by insertion of \textit{sull}, a gene conferring resistance to sulfonamide antibiotics, which were introduced in the 1930s.\textsuperscript{64,65} Various independent deletions of the Tn\textsubscript{402} transposition machinery also occurred.\textsuperscript{66,67} The clinical class 1 Tn\textsubscript{402} integron inserted into another transposon containing the \textit{mer} operon (confering resistance to mercury), thus spawning the Tn\textsubscript{21} compound transposon, itself now found on diverse plasmids as a series of complex rearrangements and derivatives.\textsuperscript{68,69} Clinical class 1 integrons have collectively accumulated over 130 gene cassettes that, in total, confer resistance to the majority of antibiotic classes used in medical practice.\textsuperscript{48} They have spread by conjugation and transformation\textsuperscript{70,71} to most Gram-negative pathogens\textsuperscript{72,73} and also into some Gram-positive organisms.\textsuperscript{74}

Thus, what appears to be a single event, originating in a single cell, has spawned an enormously successful and widely distributed group of genetic elements that are now present at high abundance in humans and their domestic animals.\textsuperscript{5} This abundance means they have become a significant component in waste streams, polluting natural environments\textsuperscript{75} and colonizing wild animals.\textsuperscript{76} These events are a direct result of the power of human activities to select for resistance phenotypes. However, an additional element of this selection is often underappreciated: coselection for the ability to participate in complex interactions with diverse elements of the microbiome. In the next section, the emergent properties that arise out of interactions between mobile elements are examined through the lens of genomic data.

**ICEs: a paradigm for combinatorial evolution**

Integrative and conjugative elements (ICEs), also known as conjugative transposons, are mobile DNA elements that can move between cells by conjugation.\textsuperscript{77,78} DNA sequencing analyses suggest that they may be the most abundant type of conjugative elements in prokaryotes.\textsuperscript{79} Unlike plasmids, the other common conjugative elements found in bacteria, ICEs, are normally integrated into the bacterial chromosome. These properties mean that ICEs have two distinct strategies for propagation: via vertical inheritance as part of the bacterial chromosome and via horizontal transfer into new bacterial hosts. Those twin strategies require every ICE to have a core set of genes that encode these abilities.\textsuperscript{80,81}

ICEs encode either a serine or tyrosine recombinase (\textit{Int}) that catalyzes a recombination reaction between a site on the incoming ICE (\textit{attP}) and the host chromosome. The integration site on the chromosome can be highly specific or variable, depending on the ICE. Most ICEs also encode an excisionase (\textit{Xis}) that excises the ICE from the chromosome as a circular molecule. The transfer of the circular ICE into a new cellular host is mediated by genes that control DNA transfer and by the formation of a mating pair between donor and recipient cells. Finally, the processes of integration, excision, and conjugative transfer are regulated by diverse sets of genes that respond to environmental cues (Fig. 2).\textsuperscript{78,80}

The core functions of maintenance, transfer, and regulation are modular, and these modules can be exchanged so that, for instance, similar regulation and conjugation modules can be linked to very different recombinase modules.\textsuperscript{82} Bioinformatic analyses of the core modules have been used to identify new ICEs in a wide variety of genomic contexts and in diverse taxa and have allowed comparisons of diversity within ICE families.\textsuperscript{78} Such comparisons
can now be performed using a dedicated and integrated online ICE database. These comparisons show that the core genes of ICEs compose a flexible scaffold that can, in turn, support a diverse array of cargo genes. The cargo genes often confer phenotypes that benefit the host cell and include phenotypes of importance to human activities, such as resistance to antibiotics and heavy metals, aromatic degradation, and symbiotic nitrogen fixation. The modular construction of ICEs and their ability to acquire diverse cargo genes allows their bacterial hosts to colonize new niches and rapidly adapt to adverse environmental conditions and helps to generate strain variability upon which selection can act. These properties can be illustrated by examining one group of ICEs for which there is considerable genomic data: the SXT/R391 ICE family.

Origins and diversity of the SXT/R391 ICE family

The history of the SXT/R391 family has a number of parallels with the discovery and characterization of the class 1 integron. This family of ICEs first drew notice because some variants conferred antibiotic resistance in *Vibrio cholerae*, the cause of cholera. In fact, the elements were originally named SXT because of their ability to confer resistance to sulfamethoxazole and trimethoprim. Like integrons, they have a modular construction, with a series of internal regions where diverse genes that confer adaptive phenotypes can be readily inserted.

The SXT/R391 ICEs are united by the possession of a highly conserved tyrosine recombinase (Int) that inserts the ICE into the 5′ end of the gene for peptide chain release factor (prfC). SXTs are 52 nearly identical, syntenic core genes. Diversity between different elements in the family is then generated by the insertion of cargo genes into one of five insertion hotspots or three variable regions. The core functions of the SXT ICE, including integration/excision, DNA processing, mating pair formation, and regulation, can be encoded by as few as 25 key genes (Fig. 2).

Members of the second wave of the seventh cholera pandemic probably first acquired the SXT ICE between 1978 and 1984, with a local origin in the Bay of Bengal. Multiple, independent acquisitions of different SXT/R391 ICEs may have occurred between 1980 and 2000. The resistance determinants found as cargo genes in SXT elements accord with the antibiotics commonly used to treat cholera: tetracycline, furazolidone, and trimethoprim-sulfamethoxazole, strongly suggesting that it was these selective agents that fixed the SXT in *Vibrio* strains belonging to the seventh pandemic. Certainly, 19th century strains of *V. cholerae* did not carry resistance genes.

The original source of SXT in *V. cholerae* was, in all probability, one of the diverse SXT-like ICEs that are widely distributed in Gammaproteobacteria found in aquatic environments. Such elements have been recovered from deep sea *Shewanella*, marine *Vibrio*, *Pseudoalteromonas*, *Photobacterium*,
Enterovibrio, and Marinomonas.\textsuperscript{91,94,101–103} Acquisition of the SXT ICE gave Vibrio a powerful tool for accessing cargo genes from the pangenome of the Gammaproteobacteria.\textsuperscript{98} For instance, the integron carried by the Vibrio SXT ICE is closely related to chromosomal integrons found in the Gammaproteobacteria Shewanella.\textsuperscript{41} The mobility of SXT elements enabled transfer into different genera, including Providencia, Proteus, and other pathogens,\textsuperscript{91,93,101} and these elements have subsequently spread across the globe.\textsuperscript{77} SXT ICEs may also enhance the virulence and survival of Vibrio through their acquisition of genes that control motility and biofilm formation.\textsuperscript{104} Consequently, the contribution of SXT ICEs to pathogen phenotypes and the survival and dissemination of cholera is far from over.

**Combinatorial evolution and emergent properties**

The modular nature of the SXT ICEs and the sharing of cargo genes between the different SXT variants allow them to interact and exchange modules. Recombination between SXT variants can then generate novel elements, effectively promoting their own diversity.\textsuperscript{89,98,105} This emergent property is further enhanced when IS elements, transposons, or integrons are inserted into SXT hotspots as cargo genes, multiplying the potential for complex interactions both between individual SXT ICEs and with mobile elements more generally.\textsuperscript{80,89,106} Thus, the modular construction of SXT ICEs and other mobile elements give them properties that can accelerate the generation of diversity.\textsuperscript{107} Further, the potential for coselection based on the physical linkage of different cargo genes means that diverse conditions, each imposing different selective forces, can all select for the same SXT ICE.\textsuperscript{108}

It has recently become apparent that the distinction between ICEs and plasmids is not as clear cut as once thought, with the observation that at least some SXT/R391 elements can replicate independently.\textsuperscript{78,109} This accords with findings that the IncA/C plasmids and the SXT/R391 ICEs may have a common ancestor,\textsuperscript{89,110} and that ICEs more generally may arise through combinatorial events between phages and conjugative plasmids, underscoring the importance of interactions between mobile elements.\textsuperscript{111} Further, the loss of conjugation or integration functions in an ICE may generate genomic islands, which can be thought of as defective ICEs.\textsuperscript{96,112} Certainly, SXT/R391 or IncA/C functions can activate phylogenetically unrelated genomic islands in trans,\textsuperscript{95} or ICEs can insert into genomic islands to generate novel, hybrid conjugative elements.\textsuperscript{80,112}

**Combinatorial evolution is widespread**

The properties and recent history of the SXT/R391 family are repeated in the histories of an increasing number of ICEs and genomic islands that have been characterized using genomics. ICEs and genomic islands are found in diverse bacterial phyla, exhibit extraordinary diversity, and often encode antibiotic resistance among their cargo genes.\textsuperscript{78,80,96,111,112} It appears that conjugative plasmids can switch to and from an ICE-like lifestyle and that ICEs may be the most abundant conjugative elements in bacteria.\textsuperscript{79} The ability of ICEs to incorporate and mobilize antibiotic resistance determinants is a consistent theme, as is their ability to undergo complex interactions with other mobile elements.

For example, the Tn916 mobile genetic elements are a family of ICEs that are widely distributed among diverse bacterial phyla, primarily the Firmicutes, but are also found in the Actinobacteria, Proteobacteria, and others. Tn916 appears to be undergoing a radiation into a range of commensals and pathogens, possibly driven by their extraordinary ability to acquire cargo genes that encode antibiotic resistance.\textsuperscript{113} In a similar manner, transposon Tn6029 has interacted with integrons, IS elements, and diverse antibiotic-resistance genes to generate a series of complex assembly events in *Escherichia coli* 0104:H4.\textsuperscript{114}

Genome sequencing continues to expand the discovery of modular ICEs and genomic islands that contain mobile elements and antibiotic resistance determinants and to document their spread into new hosts such as the opportunistic pathogens *Pseudomonas aeruginosa*\textsuperscript{115} and *Acinetobacter baumannii*.\textsuperscript{116} These mosaic assemblies have almost certainly been fixed in new species by antibiotic selection. However, each of these cases simply represents a continuation of one means by which bacteria have generated diversity over evolutionary timescales. For instance, in *Streptococcus agalactiae*, over 60% of the regions of divergence between strains are ICEs, and these elements are the major source of ongoing diversity in this species.\textsuperscript{117,118}
Recombination events between ICEs in different *Streptococcus* species can yield new ICEs with additional properties and host ranges and consequently also disseminate antibiotic resistance under the appropriate selection pressures.  

**Antibiotic resistance and databases**

DNA sequencing technologies have helped us to reconstruct the likely origins of many of the resistance genes currently circulating and evolving in human pathogens. While there are genes found in the human commensal flora that can confer resistance to antimicrobial agents, these are generally not the genes of concern in clinical contexts. In contrast, soil bacteria carry genes with 100% identity to genes that confer resistance to major antibiotic classes in a range of human pathogens. This strongly suggests that many resistance genes in pathogens were recently acquired by lateral gene transfer from environmental organisms. If so, each of these acquisitions was a one-off event that occurred in a single cell, at a single point in time, and at a single location. The essentially random nature of these events makes it very difficult to predict the identity or origin of the next resistance gene that might be acquired by lateral gene transfer.

We can, however, document the most likely locations for such events and the ecological drivers that promote them. The establishment and maintenance of databases and bioinformatics tools is essential for this effort, particularly as the accumulation of whole-genome and next-generation sequencing data rapidly expands our knowledge of resistance determinants. There are already a number of databases dedicated to antibiotic resistance determinants (Table 1), each of which has its own advantages and disadvantages.

A consolidated and unified approach for cataloging antibiotic-resistance phenomena would help manage the resistance crisis. The ideal database would tabulate resistance genes, the genetic platforms that capture such genes, and the mobile elements that move these genes between locations and cells. Databases would also benefit from the inclusion of genes that confer resistance to metals and biocides and genes that confer virulence. Ideally, these genes need, where possible, to be assigned to a bacterial host (or hosts). Finally, the genes need to be stratified according to whether they were derived from a known clinical pathogen or the microbiota of humans, animals, or plants, or if they are from environments such as soil, sediment, or water.

Information derived from such databases needs to be carefully interpreted. Currently, there is an understandable bias toward genes from clinical contexts and from the human microbiota. It needs to be borne in mind that these records reflect the post-antibiotic world and are not necessarily a good representation of the activities or locations of these genes before the 1950s. We know that there are diverse and ancient gene families that can confer resistance to clinical antibiotics, but the vast majority of these genes will probably never become clinical problems. When such genes do become clinical problems, often just one gene variant has been co-opted from a family of diverse genes, most of which do not confer resistance. For instance, the aminoglycoside acetyltransferases range from providing little, if any, resistance to full-blown resistance phenotypes. Consequently, without functional studies, attributions of resistance based purely on sequence homology are problematic. Attributes based on sequence homology also identify wild-type genes that must undergo point mutation to confer resistance. These are normally housekeeping genes and are not resistance determinants in the usual sense, unless they have been transferred onto a mobile element. Similarly, genes encoding efflux pumps are ubiquitous, allowing cells to eliminate diverse compounds, and serendipitously this includes some antibiotics.

Newly discovered resistance mechanisms need to be incorporated into databases, but many of these mechanisms may be hard to capture using gene-centric approaches. Examples include intrinsic resistance, tolerance, and persistence. Mechanisms such as lag-time extension and biofilm formation may be polygenic and environmentally determined. Despite these difficulties, well-constructed databases help us to understand the history of resistance phenomena and could be used to manage and predict resistance in the future.

**Genomics, databases, and applications**

There are at least three general ways that genomics and sequence databases focused on antibiotic resistance might help inform the future of resistance management. The first is in identifying and understanding the interplay between the various
Table 1. List of general resources for DNA sequence analysis of antibiotic resistance genes and associated genetic elements

<table>
<thead>
<tr>
<th>Resource</th>
<th>Website</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Antibiotic resistance</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARDB—antibiotic resistance gene database</td>
<td><a href="http://ardb.cbcb.umd.edu/">http://ardb.cbcb.umd.edu/</a></td>
<td>185</td>
</tr>
<tr>
<td>CARD—comprehensive antibiotic resistance</td>
<td><a href="https://card.mcmaster.ca/">https://card.mcmaster.ca/</a></td>
<td>186</td>
</tr>
<tr>
<td>database</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAC—repository of antibiotic resistance</td>
<td><a href="http://rac.aihi.mq.edu.au/rac/">http://rac.aihi.mq.edu.au/rac/</a></td>
<td>187</td>
</tr>
<tr>
<td>cassettes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ResFams—antibiotic resistance function</td>
<td><a href="http://www.dantaslab.org/resfams/">http://www.dantaslab.org/resfams/</a></td>
<td>128</td>
</tr>
<tr>
<td>annotation</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Other resistance and virulence determinants</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MvirDB—virulence factors and toxins</td>
<td><a href="http://mvirdb.llnl.gov/">http://mvirdb.llnl.gov/</a></td>
<td>190</td>
</tr>
<tr>
<td>PATRIC—pathosystems resource integration</td>
<td><a href="https://www.patricbrc.org/portal/portal/patric/Home#">https://www.patricbrc.org/portal/portal/patric/Home#</a></td>
<td>191</td>
</tr>
<tr>
<td>center</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BacMet—biocide and metal resistance genes</td>
<td><a href="http://bacmet.biomedicine.gu.se/">http://bacmet.biomedicine.gu.se/</a></td>
<td>192</td>
</tr>
<tr>
<td>Efflux pumps—review</td>
<td><a href="http://cmr.asm.org/content/28/2/337.short">http://cmr.asm.org/content/28/2/337.short</a></td>
<td>134</td>
</tr>
<tr>
<td><strong>Vectors and mobile elements</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTEGRALL—the integron database</td>
<td><a href="http://integrall.bio.ua.pt/">http://integrall.bio.ua.pt/</a></td>
<td>193</td>
</tr>
<tr>
<td>Integen Finder</td>
<td><a href="https://github.com/gem-pasteur/Integen_Finder">https://github.com/gem-pasteur/Integen_Finder</a></td>
<td>194</td>
</tr>
<tr>
<td>ACLAME—phages, plasmids, and</td>
<td><a href="http://aclame.ulb.ac.be/">http://aclame.ulb.ac.be/</a></td>
<td>195</td>
</tr>
<tr>
<td>transposons</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IS Finder—insertion element finder</td>
<td><a href="https://www-is.biotoul.fr/">https://www-is.biotoul.fr/</a></td>
<td>146</td>
</tr>
<tr>
<td>ICEBerg—integrative conjugative element</td>
<td><a href="http://db-mml.sjtu.edu.cn/ICEberg/">http://db-mml.sjtu.edu.cn/ICEberg/</a></td>
<td>83</td>
</tr>
<tr>
<td>analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IslandViewer—genomic island prediction</td>
<td><a href="http://www.pathogenomics.sfu.ca/islandviewer/">http://www.pathogenomics.sfu.ca/islandviewer/</a></td>
<td>196</td>
</tr>
<tr>
<td>PAIDB—pathogenicity island database</td>
<td><a href="http://www.paidb.re.kr/about_paidb.php">http://www.paidb.re.kr/about_paidb.php</a></td>
<td>197</td>
</tr>
<tr>
<td>PlasmidFinder—In silico plasmid detection</td>
<td><a href="http://cge.cbs.dtu.dk/services/PlasmidFinder/">http://cge.cbs.dtu.dk/services/PlasmidFinder/</a></td>
<td>198</td>
</tr>
</tbody>
</table>

Note: This tabulation is representative rather than exhaustive. Detailed considerations of the relative advantages of some of these resources have recently been published.125,126

agents involved in resistance: the resistome, the mobilome, the bacterial hosts, and their environmental compartments.22 The second is in rapid identification and tracking of both genes and species of interest.141 The third is in examining the development of resistance in real time, using experimental evolution.142

Accumulating sequence data make it more and more apparent that the human use of antibiotics does not just exert effects on the abundance and distribution of resistance genes, but also on the abundance of mobile DNA elements more generally. For instance, insertion sequences (IS) are the simplest type of transposable element in bacterial genomes, and these have undergone a number of recent expansions143 driven by their linkage to resistance genes. Indeed, some ISs specifically target the recombination sites of integrons.144 More generally, genome sequencing has expanded our knowledge of IS, such that there are now over 4000 of these elements known,145 and the range of known strategies for their propagation is growing.146

In particular, the association of the ISCR1 element with class 1 integrons has been instrumental in the widespread dissemination of resistance in Gram-negative organisms and in the assembly of resistance islands.106,111 This IS, and other elements such as IS26, helps to assemble new transposable elements and fuse these with antibiotic resistance genes.147,148 Such compound elements help to move resistance determinants between cells and species, increasing the abundance and distribution of the resistance genes themselves and the mobile elements that facilitate their spread. The increasing
modularity and mosaicism of mobile resistance elements offers opportunities for tracking and identifying bacterial populations that contain such modules and for better understanding the mechanisms that promote their dissemination.\textsuperscript{81,107}

As the costs of DNA sequencing fall and the speed of analyses improves, it is becoming practical to use whole-genome sequencing to study antibiotic resistance. Genome sequencing can be used to manage individual infections, to improve diagnostic turnaround, and to track outbreaks.\textsuperscript{141,149} It is particularly useful for rapidly determining resistance profiles in slow-growing organisms such as \textit{Mycobacterium}, especially where the molecular basis of resistance is known.\textsuperscript{150,151} Even in fast-growing organisms, genome sequencing can be useful in predicting antimicrobial resistance and has the additional advantages of providing strain identity and other phenotypic information.\textsuperscript{152} At a population level, sequencing is useful for identifying unrecognized transmission routes and real-time tracking of outbreaks,\textsuperscript{153,154} including recently discovered resistance genes.\textsuperscript{155–158}

Genomics and metagenomics offer opportunities for understanding the conduits of gene transfer between bacteria, animal hosts, and environmental compartments.\textsuperscript{5} Investigating resistomes across habitats helps to detect highly mobile resistance determinants and identifies hotspots for lateral transfer of resistance genes, such as wastewater treatment plants or chicken coops.\textsuperscript{159} Such hotspots are important for controlling the spread of resistance, because resistance genes can cluster by environment,\textsuperscript{160} and controlling transfers between different environments is an obvious first choice in controlling resistance more generally. Treating antibiotic resistance genes and their host cells as ecological assemblages within an environment allows the application of well-tested ecological principles to the problem.\textsuperscript{161} Environmental compartments, such as the human gut\textsuperscript{162–164} or the oral cavity,\textsuperscript{165} might then be profitably viewed from the perspectives of dispersal, colonization, invasion, and island biogeography.\textsuperscript{37}

More generally, metagenomics will be useful in understanding the dynamics of lateral transfer of antibiotic resistance in managed ecosystems.\textsuperscript{166,167} It will be particularly important in tracking resistance in farm animals,\textsuperscript{168,169} understanding the linkages between resistance elements,\textsuperscript{170} and understanding the modes by which resistance genes are shared between humans and their domestic animals.\textsuperscript{171} New hotspots for generation of resistance elements will be identified, such as the potential for aquaculture facilities to be incubators of resistance.\textsuperscript{172,173}

The power and precision of genomic technologies will be useful for analyzing \textit{in vitro} and experimental evolution studies on antibiotic resistance.\textsuperscript{142} In particular, there is growing concern that low-level pollution with antibiotics might influence the rate at which resistance evolves.\textsuperscript{174,175} Resistance can be selected at subinhibitory antibiotic concentrations,\textsuperscript{176,177} and the coupling of experimental evolution with genomics is beginning to help us understand the complex dynamics of resistance in populations that are exposed to antimicrobial agents.\textsuperscript{178,179} Exposure to biocides, for instance, selects for mutations in efflux pumps, and these mutations also confer resistance to multiple antibiotics.\textsuperscript{180} Antibiotic exposure more generally is thought to increase mutation rates across the whole genome.\textsuperscript{181} The amenability of most pathogenic microorganisms to laboratory growth, their rapid generation times, the ability to run multiple replicates in small spaces, and the speed and accuracy of genomic technologies all mean that experimental evolution to examine antibiotic resistance is potentially a fruitful and tractable area of research.

**Conclusions**

Our understanding of the mechanisms of antibiotic resistance and the means by which bacteria have become resistant has been largely informed by the revolution in genomics and DNA sequencing. We still have much to learn, particularly about the relative importance of various environmental compartments that contribute to resistance phenomena.\textsuperscript{122,164} Tackling the problem of antibiotic resistance involves monitoring all environments\textsuperscript{182,183} and acquiring a better understanding of the drivers of resistance phenomena.\textsuperscript{127} Finally, microorganisms do not recognize geopolitical boundaries, and a resistance gene arising in one location can rapidly spread around the world. Like many other problems that face humanity, resistance needs to be tackled at a global scale.\textsuperscript{184}

**Conflicts of interest**

The authors declare no conflicts of interest.
References

Genomics and antibiotic resistance


73. van Essen-Zandbergen, A., H. Smith, K. Veldman, et al. 2007. Occurrence and characteristics of class 1, 2 and


82. Roberts, A.P., P.A. Johanesen, D. Lyras, et al. 2001. Comparison of Tn5397 from Clostridium difficile, Tn916 from Enterococcus faecalis and the CW459tet (M) element from Clostridium perfringens shows that they have similar conjugation regions but different insertion and excision modules. Microbiology 147: 1243–1251.


