

The evolution of plasmid gene content in Enterobacteriaceae

Dissertation

in fulfillment of the requirements for the degree of
Doktor der Naturwissenschaften (Dr. rer. nat.)
in the Faculty of Mathematics and Natural Sciences
at Kiel University

submitted by
Yiqing Wang

Kiel, 2023

First examiner: Prof. Dr. Tal Dagan

Second examiner: Prof. Dr. Mathieu Groussin

Date of the oral examination: 06.07.2023

Declaration

I hereby declare that the thesis entitled "The evolution of plasmid gene content in Enterobacteriaceae" has been carried out in the Institute of General Microbiology at Kiel University, Kiel, Germany, under the guidance of Prof. Dr. Tal Dagan. The work is original and has not been submitted in part or full by me for any degree at any other university. I further declare that the material obtained from other sources has been duly acknowledged in the thesis. My work has been produced in compliance to the principles of good scientific practice in accordance with the guidelines of the German science foundation. I hereby assure that I have not been revoked any of my academic degrees.

Kiel, 2023

Wang, Yiqing

Table of contents

Abstract.....	1
Zusammenfassung.....	2
Introduction.....	3
Thesis outline	13
Chapter I - Essential gene acquisition destabilizes plasmid inheritance	15
Chapter II - Gene sharing among plasmids and chromosomes reveals barriers for antibiotic resistance gene transfer.....	47
Chapter III - Transposable elements drive the evolution of multiple drug resistance plasmids	75
Concluding remarks	109
Acknowledgements	111

Abstract

Plasmids are extra-chromosomal genetic elements that play a crucial role in the genome evolution of prokaryotes. Plasmid-encoded genes may be beneficial for their host, e.g., genes encoding antibiotic resistance. In this thesis, I investigated the gene content evolution in plasmid genomes, with a focus on essential and beneficial genes. In **chapter one**, I asked if plasmid-encoded genes may be essential for the host. For that, I surveyed the presence of essential genes on plasmids in *Escherichia coli* and found that these are rarely found on plasmids. Through phylogenetic analyses and evolution experiments, we provided a mechanistic explanation of why essential genes are rarely found on plasmids. Our results highlight a barrier to lateral transfer of essential genes from a chromosome to plasmids. In **chapter two**, I compared the distribution of different antibiotic resistance genes (ARG) between plasmids and chromosomes in *Escherichia*, *Salmonella*, and *Klebsiella*. My results reveal that the function of the encoded resistance mechanisms is an important determinant of gene mobility. In **chapter three**, I quantified the cumulative impact of transposable elements on the evolution of ARG content in different types of plasmids. My results reveal the trends and barriers to plasmid-mediated ARG transfer and emphasized the important role of transposable elements in generating antibiotic resistance islands in plasmid genomes. Overall, the studies in this thesis unveil several constraints on the evolution of plasmid gene content, and contributions of transposable elements to plasmid genome evolution.

Zusammenfassung

Plasmide sind extra-chromosomale genetische Elemente, die eine entscheidende Rolle bei der Genomevolution von Prokaryoten spielen. Plasmid-kodierte Gene können für ihren Wirt von Nutzen sein, z. B. Gene, die Antibiotikaresistenzen kodieren. In dieser Arbeit habe ich die Entwicklung des Geninhalts in Plasmidgenomen mit Fokus auf essenzielle und nützliche Gene untersucht. Im **ersten Kapitel** ging ich der Frage nach, ob Plasmid-kodierte Gene für den Wirt essenziell sein können. Dazu untersuchte ich, ob essenzielle Gene auf Plasmid in *E. coli* vorhanden sind und konnte feststellen, dass solche Gene auf Plasmiden selten sind. Mit Hilfe von phylogenetischen Analysen und Evolutionsexperimenten haben wir eine Erklärung, warum essenzielle Gene selten auf Plasmiden zu finden sind. Unsere Ergebnisse weisen auf eine Barriere für den lateralen Transfer essenzieller Gene von einem Chromosom auf Plasmide hin. Im **zweiten Kapitel** habe ich die Verteilung verschiedener Antibiotikaresistenzgene (ARG) zwischen Plasmiden und Chromosomen in *Escherichia*, *Salmonella* und *Klebsiella* verglichen. Meine Ergebnisse zeigen, dass die Funktion der kodierten Resistenzmechanismen ein wichtiger Faktor für die Mobilität von Genen ist. Im **dritten Kapitel** habe ich den kumulativen Einfluss von transponierbaren Elementen auf die Entwicklung des ARG-Inhalts in verschiedenen Plasmidtypen quantifiziert. Meine Ergebnisse zeigen den Trend und die Hindernisse für den plasmidvermittelten ARG-Transfer auf und unterstrichen die wichtige Rolle transponierbarer Elemente bei der Bildung von Antibiotikaresistenzinseln in Plasmidgenomen. Insgesamt weisen die Studien in dieser Arbeit auf mehrere Einschränkungen bei der Evolution von Plasmidgenen und auf den Beitrag von Transposonen bei der Evolution des Plasmidgenoms hin.

Introduction

Plasmids are extra-chromosomal replicons

Plasmids are mostly small, circular DNA molecules that are typically found in bacteria and archaea (Rodríguez-Beltrán et al., 2021). They are considered to be autonomous replicating entities and are referred as a group of secondary replicons. It is estimated that approximately 10% of bacterial genomes contain these extra-chromosomal replicons, in addition to the bacterial chromosome (diCenzo & Finan, 2017). Unlike chromosomes, plasmids encode their own replication origin and partition system, and they replicate independently of the chromosome in the same cell. Plasmids are usually present in a circular form and can have a variable copy number relative to the chromosome (Twigg & Sherratt, 1980).

Secondary replicons are classified based on their size and presence of genes encoding essential functions for the bacterial host. Plasmids and megaplasmids are defined as replicons that lack essential genes, with a cutoff of 350 kb distinguishing them in genome size. Chromids, on the other hand, are defined as secondary replicons that encode essential genes (Harrison et al., 2010). It is worth noting that the term "second chromosome" was frequently misused to refer to chromids. In reality, second chromosomes originate from rare instances when a piece of DNA splits from the bacterial chromosome and becomes a secondary replicon in the bacterial cell (diCenzo & Finan, 2017). As such, a second chromosome is different from any plasmid-origin replicons.

Circular plasmids encode genes and sequences necessary for their propagation, including origin(s) of replication (*ori* sequence), a gene that encodes a replication initiation protein (Rep protein), and genes involved in the control of replication. Plasmids use one of three replication mechanisms: Theta-type, strand displacement, and Rolling-Circle replication (del Solar et al., 1998). Theta replication mechanism has been extensively studied among the plasmids of gram-negative bacteria like *E. coli*. Under the electron microscopy, theta replication mechanism forms intermediates in a Θ -form, which is why it is called theta-type. There are mainly two types of mechanism in the theta plasmid replication, which differ based on the initiation factor being required (class A) or not required (class B) (Lilly & Camps, 2015). Most of the plasmids replicating using the theta-type mechanism require a plasmid-encoded initiation factor (Rep protein) in the duplex DNA melting. ColE1 plasmids are, however an exception, as they rely exclusively on host factors for DNA melting and primer synthesis. The two types mechanisms are therefore fundamentally different.

Plasmids are generally considered unstable and are not expected to carry any essential genes (Tazzyman & Bonhoeffer, 2015). However, chromids have plasmid replication mechanism and encode essential genes which are critical to bacterial survival. Due to the

presence of essential genes, chromids are believed as indispensable genome components. Understanding the origin of essential genes on chromids is crucial for understanding plasmid genome evolution. One hypothesis is that chromids arise through the translocation of essential genes from the chromosome to a plasmid in the same cell. Alternatively, chromids may have originated from transferred plasmids carrying redundant copies of essential genes (Hülter et al., 2017). While the essential gene content on the chromosome of *Escherichia coli* has been extensively studied (Goodall et al., 2018; Rousset et al., 2021), the essential gene content on plasmids remains largely unexplored. In this thesis, I aim to screen the plasmid-encoded essential genes using genome data from *E. coli* strains and perform phylogenetic analysis to test the two possible origins of the essential genes on plasmids.

Plasmid classification

The classic plasmid classification originates from an interesting phenomenon. Under non-selective condition, plasmids sharing the same replication control mechanism or partitioning system are not able to stably propagate together in the same cell (termed plasmid incompatibility) (Novick, 1987). Nevertheless, plasmids with different replication mechanisms are usually compatible. Accordingly, plasmids are classified into different incompatibility groups (Couturier et al., 1988). The different plasmid replication types serve as the sequence basis for this classification. In the genomic era, a practical way to assign plasmid incompatibility groups is by comparing the amino acid sequences of Rep proteins. In Enterobacteriaceae, plasmids are assigned to more than 18 major incompatibility groups (Carattoli & Hasman, 2020), including: IncF (prototype plasmid R27, R1, and F factor); IncI (prototype plasmid R64); IncA/C (prototype plasmid R55); IncN (prototype plasmid R46); IncH (prototype plasmid R478); IncP (prototype plasmid RP4); IncX; and IncY (chimeric plasmids also encode phage signature genes, the so-called Phage-Plasmid (Pfeifer et al., 2021)). These Rep-encoding plasmids have diverse genome sizes and may have a narrow or broad host range. Plasmids that do not encode Rep proteins, such as ColE1-like plasmids, are practically classified through comparing the RNA gene sequences included in their replication control mechanism. ColE1-like plasmids are usually small in genome size and have a narrow host range specific to Gammaproteobacteria (Ares-Arroyo et al., 2021).

Plasmids can propagate not only by the host cell division, but also transfer between bacterial cells. In addition to replicon typing, plasmids are also classified into different mobility types based on their ability to transfer. Conjugative plasmids encode an origin of transfer (*oriT*) sequence, a relaxase, a type IV coupling protein and a set of genes encoding a type IV secretion system. These components enable conjugative plasmids to autonomously transfer by conjugation from one cell to another. Mobilizable plasmids are only transferrable between

hosts with the help of other conjugative plasmids. Mobilizable plasmids do not contain the full set of genes encoding type IV secretion system, but they encode at least an *oriT* sequence and a relaxase. Plasmids encoding only an *oriT* sequence may also be transmissible, utilizing the conjugative machinery from other co-existing conjugative plasmids (Ares-Arroyo et al., 2023). The remaining plasmids that do not encode any essential components for mobilization are presumably non-transmissible plasmids (Smillie et al., 2010).

Unlike plasmid replicon types, plasmid mobility is a dynamic trait that can evolve over time. For instance, a conjugative plasmid could become a mobilizable or a transmissible, even a non-transmissible plasmid (Coluzzi et al., 2022). The mechanism behind this mobility transition could be gene loss or gene function loss. Conversely, a non-transmissible plasmid could also evolve into transmissible plasmid by acquiring genes and sequences encoding transfer machinery. Plasmids of the same replicon type could have different mobility types. Interestingly, plasmid size has been found to be associated with plasmid mobility. The plasmid size distribution has been described as bimodal, with a minimum density value at around 20kb (Smillie et al., 2010). However, conjugative plasmids have a unimodal distribution with a maximum density value at 100 kb. In contrast, mobilizable and non-mobilizable plasmids have multimodal distributions with multiple maximum density values. To study plasmid genome evolution, we need to compare the plasmid gene content between plasmids with different replicon or mobility types. However, it is crucial to first disentangle the associations between the different features in plasmid classification systems. In this thesis, I aim to compare plasmid gene content using phylogenetic trees of plasmid marker genes as a guide for my analysis.

Plasmid gene content

To understand plasmid evolution and compare their gene content, it is important to distinguish between genes in the plasmid genome according to their function. Plasmid gene content can be divided into plasmid backbone genes and cargo genes. Plasmid backbones are defined as a set of conserved genes encoding plasmid essential functions, such as replication and replication control genes, genes encoding functions for plasmid stability and maintenance, e.g., partition systems and multimer resolution systems; genes encoding transfer machinery and genes in the leading region encoding stability modules for transferred plasmid DNA (Garcillán-Barcia et al., 2011). In genomic studies, plasmid backbones are usually defined by comparative genomic analysis (e.g., Gilmour et al., 2004). While many plasmid backbone genes have been extensively studied for the encoded functions, there is still a large number of them without known functions in plasmid backbones. Due to their fundamental roles in plasmid replication and transfer, plasmid backbone genes are stably inherited throughout the plasmid propagation. Importantly, different types of plasmids have unique sets of backbone genes.

Given a specific plasmid type, cargo genes are defined as the remaining genes in plasmid genomes besides backbone genes. Cargo genes often encode functions related to the bacterial host fitness, which facilitate the rapid adaptation to different environments. For example, antibiotic resistance genes are the prototype of plasmid cargo gene. These genes often reside in smaller mobile genetic elements, such as transposable elements (Partridge et al., 2018). The replication and transfer of the mobile genetic element drive the evolution of these cargo genes (Rankin et al., 2011).

Notably, plasmid backbone and cargo genes are usually found in plasmid genomes in separated regions. Backbone genes are usually found in a continuous block in the plasmid genome. The gene order is important for the order of gene replication and expression regulation. Consequently, backbone genes are replicated and regulated in a coordinated manner. Adjacent to the backbone genes lies a distinct region within the plasmid genome dedicated to cargo genes. Due to the frequent insertion and recombination events among the plasmids, even the same type of plasmids may have very diverse cargo gene content. Cargo genes encode often independent functions, e.g., antibiotic inactivation proteins. They act without complex interaction with other proteins in the host network. Compared to backbone genes, different cargo genes on the same plasmids interact fewer with each other. In this thesis, I aim to study the distribution of cargo genes on plasmids and the genetic mechanism driving the evolution of cargo gene repertoire.

Plasmids interact with other mobile genetic elements

Bacterial genomes are the playgrounds of a variety of mobile genetic element (MGE) types (Haudiquet et al., 2022). Different MGEs can interact to facilitate gene transfer, resulting in gene gain and loss events in the genome. As genetic vehicles plasmids often carry MGEs, such as transposable elements, which are the most abundant MGE in bacterial genomes (Khedkar et al., 2022). There are many types and structural variants of transposable elements (TE), which include simple insertion sequences, transposons and composite transposons. Transposable elements, with or without cargo genes, can transfer (i.e., jump) between different locations within a plasmid, between different plasmids, and even between plasmids and chromosome in the same host. They form the gene flow among plasmids and the chromosome in a bacterial genome and facilitate gene exchange between genomes of different bacterial lineages by hitchhiking on transferable plasmids.

A second important MGE type found in plasmid genomes is integrons, which also carry cargo genes and are often found in the neighborhood of other MGEs (Partridge et al., 2018). Compared to transposable elements, integrons are found in fewer copies than TE and have fewer structural variants. Phages are a third example of MGEs that interact with plasmids.

Phage-Plasmids were recently defined as a group of plasmids which is composed of signature genes from both phages and plasmids (Pfeifer et al., 2021). Phage genes integrate into the plasmid genome and use plasmid replication mechanisms to propagate. Phage-Plasmids are phages that also act as plasmids, forming a specific group and have stable inheritance in bacterial population. They are the perfect example for an interaction between MGEs - phages with plasmids – that evolved into stable replicons.

Plasmids also have conflicts with other MGEs in the same host cell (Werren, 2011). Random insertions of MGEs may disrupt the crucial gene or genome structure of plasmids. Incoming DNA of other MGEs is a natural competitor for the resident plasmids. To protect themselves from these potential conflicts, plasmids also employ defense functions encoded by genes and sequences, such as the CRISPR-Cas system. Interestingly, plasmid-borne defense systems are found mainly targeting other plasmids, indicating that conflicts between plasmids are prevalent in bacterial populations (Pinilla-Redondo et al., 2022).

The cooperation and competition between plasmids and MGEs are key drivers of plasmid gene content evolution (Horne et al., 2023), particularly in the lateral transfer of accessory genes. The constant gene flux between MGEs results in an open catalogue of cargo genes in plasmid genomes. As plasmids reside in host cells, they inevitably encounter MGEs, but the extent and nature of these interactions vary between different types of plasmids. Transferable plasmids, are found to have greater opportunities to exchange cargo genes with bacterial genomes due to their mobility, while non-mobilizable plasmids have fewer chances to encounter MGEs. In this thesis, I aim to investigate and compare the interactions with MGEs between different plasmid mobility types.

Antibiotic resistance plasmids in pathogenic strains

Enterobacteriaceae is a family of gram-negative bacteria that includes both commensal bacteria found in the human gut and clinically significant pathogens such as *Escherichia*, *Salmonella* and *Klebsiella* (Rock & Donnenberg, 2014). These pathogens not only cause a variety of symptoms, but they are also frequently resistant to antibiotics (van Duin & Doi, 2016). The emergence of new antibiotic resistance pathogenic strains is often mediated by the acquisition of plasmids carrying antibiotic resistance genes (e.g., Kumarasamy et al., 2010). Mediated by small mobile genetic elements e.g., transposons and IS elements, recombination and integration events occur frequently in the plasmid genome (Harmer et al., 2014). The underlying process also leads to the evolution of diverse resistance gene content among plasmid genomes. This, combined with the variable copy number and transferability between bacterial hosts, plasmids played a crucial role in the development of antibiotic resistance in Enterobacteriaceae (Rozwandowicz et al., 2018).

Resistance plasmids in Enterobacteriaceae are present in more than a dozen incompatibility (Inc) types, with IncF, IncA/C, IncH, and IncI plasmids being the most prevalent. These plasmids typically have low copy numbers, medium genome sizes, and are often transmissible. Genes on plasmids often encode specific resistance mechanisms, including antibiotic inactivation, antibiotic target protection, and target replacement (Rodríguez-Beltrán et al., 2020). Plasmid-borne antibiotic inactivation genes are particularly noteworthy, as they are often found on plasmids. Inactivation genes encode proteins that destroy antibiotics and protect the host cell. Among the inactivation proteins, beta-lactamases and carbapenems have received the most attention, and their epidemiology have been extensively studied (Logan & Weinstein, 2017). Genes encoding beta-lactamase on plasmids evolve rapidly, hundreds of variants in extended-spectrum beta-lactamase family have been continually discovered (Castanheira et al., 2021). Transmissible resistance plasmids further enable the spread of these fast-evolving resistance genes among bacterial population at different levels. To better understand the genetic mechanisms behind fast-evolving antibiotic resistance and to anticipate upcoming challenges, it is crucial to study rules in plasmid genome evolution.

The increasing amount of bacterial genome data provides us with the opportunity to study plasmid genome evolution on a large scale. Among the available genomes, bacterial strains in the genera *Escherichia*, *Salmonella*, and *Klebsiella* have the most complete assemblies with plasmids. The plasmids in these three genera were found to be similar, i.e., they share common gene content which indicates that they may have same origins (Redondo-Salvo et al., 2020). In this thesis, I used genomic data of strains in Enterobacteriaceae to study plasmid evolution with a focus on the evolution of their antibiotic resistance gene content.

Reference

- Ares-Arroyo, M., Coluzzi, C., & P.C. Rocha, E. (2023). Origins of transfer establish networks of functional dependencies for plasmid transfer by conjugation. *Nucleic Acids Research*, 51(7), 3001–3016. <https://doi.org/10.1093/NAR/GKAC1079>
- Ares-Arroyo, M., Rocha, E. P. C., & Gonzalez-Zorn, B. (2021). Evolution of ColE1-like plasmids across γ -Proteobacteria: From bacteriocin production to antimicrobial resistance. *PLOS Genetics*, 17(11), e1009919. <https://doi.org/10.1371/JOURNAL.PGEN.1009919>
- Carattoli, A., & Hasman, H. (2020). PlasmidFinder and in silico pMLST: Identification and typing of plasmid replicons in Whole-Genome Sequencing (WGS). *Methods in Molecular Biology*, 2075, 285–294. https://doi.org/10.1007/978-1-4939-9877-7_20
- Castanheira, M., Simner, P. J., & Bradford, P. A. (2021). Extended-spectrum β -lactamases: an update on their characteristics, epidemiology and detection. *JAC-Antimicrobial Resistance*, 3(3), dlab092. <https://doi.org/10.1093/JACAMR/DLAB092>

- Coluzzi, C., Garcillán-Barcia, M. P., de la Cruz, F., & Rocha, E. P. C. (2022). Evolution of plasmid mobility: Origin and fate of conjugative and nonconjugative plasmids. *Molecular Biology and Evolution*, 39(6), msac115. <https://doi.org/10.1093/MOLBEV/MSAC115>
- Couturier, M., Bex, F., Bergquist, P. L., & Maas, W. K. (1988). Identification and classification of bacterial plasmids. *Microbiological Reviews*, 52(3), 375–395. <https://doi.org/10.1128/MR.52.3.375-395.1988>
- del Solar, G., Giraldo, R., Ruiz-Echevarría, M. J., Espinosa, M., & Díaz-Orejas, R. (1998). Replication and control of circular bacterial plasmids. *Microbiology and Molecular Biology Reviews*, 62(2), 434–464. <https://doi.org/10.1128/MMBR.62.2.434-464.1998>
- diCenzo, G. C., & Finan, T. M. (2017). The Divided Bacterial Genome: Structure, Function, and Evolution. *Microbiology and Molecular Biology Reviews*, 81(3), e00019-17. <https://doi.org/10.1128/MMBR.00019-17>
- Garcillán-Barcia, M. P., Alvarado, A., & De la Cruz, F. (2011). Identification of bacterial plasmids based on mobility and plasmid population biology. *FEMS Microbiology Reviews*, 35(5), 936–956. <https://doi.org/10.1111/J.1574-6976.2011.00291.X>
- Gilmour, M. W., Thomson, N. R., Sanders, M., Parkhill, J., & Taylor, D. E. (2004). The complete nucleotide sequence of the resistance plasmid R478: defining the backbone components of incompatibility group H conjugative plasmids through comparative genomics. *Plasmid*, 52(3), 182–202. <https://doi.org/10.1016/j.plasmid.2004.06.006>
- Goodall, E. C. A., Robinson, A., Johnston, I. G., Jabbari, S., Turner, K. A., Cunningham, A. F., Lund, P. A., Cole, J. A., & Henderson, I. R. (2018). The essential genome of *Escherichia coli* K-12. *MBio*, 9(1), e02096-17. <https://doi.org/10.1128/MBIO.02096-17>
- Harmer, C. J., Moran, R. A., & Hall, R. M. (2014). Movement of IS26-Associated antibiotic resistance genes occurs via a translocatable unit that includes a single IS26 and preferentially inserts adjacent to another IS26. *MBio*, 5(5), e01801-14. <https://doi.org/10.1128/MBIO.01801-14>
- Harrison, P. W., Lower, R. P. J., Kim, N. K. D., & Young, J. P. W. (2010). Introducing the bacterial ‘chromid’: not a chromosome, not a plasmid. *Trends in Microbiology*, 18(4), 141–148. <https://doi.org/10.1016/J.TIM.2009.12.010>
- Haudiquet, M., De Sousa, J. M., Touchon, M., & Rocha, E. P. C. (2022). Selfish, promiscuous and sometimes useful: how mobile genetic elements drive horizontal gene transfer in microbial populations. *Philosophical Transactions of the Royal Society B*, 377(1861), 20210234. <https://doi.org/10.1098/RSTB.2021.0234>
- Horne, T., Orr, V. T., & Hall, J. P. (2023). How do interactions between mobile genetic elements affect horizontal gene transfer? *Current Opinion in Microbiology*, 73, 102282. <https://doi.org/10.1016/J.MIB.2023.102282>

- Hülter, N., Ilhan, J., Wein, T., Kadibalban, A. S., Hammerschmidt, K., & Dagan, T. (2017). An evolutionary perspective on plasmid lifestyle modes. *Current Opinion in Microbiology*, 38, 74–80. <https://doi.org/10.1016/J.MIB.2017.05.001>
- Khedkar, S., Smyshlyaev, G., Letunic, I., Maistrenko, O. M., Coelho, L. P., Orakov, A., Forslund, S. K., Hildebrand, F., Luetge, M., Schmidt, T. S. B., Barabas, O., & Bork, P. (2022). Landscape of mobile genetic elements and their antibiotic resistance cargo in prokaryotic genomes. *Nucleic Acids Research*, 50(6), 3155–3168. <https://doi.org/10.1093/NAR/GKAC163>
- Kumarasamy, K. K., Toleman, M. A., Walsh, T. R., Bagaria, J., Butt, F., Balakrishnan, R., Chaudhary, U., Doumith, M., Giske, C. G., Irfan, S., Krishnan, P., Kumar, A. V., Maharjan, S., Mushtaq, S., Noorie, T., Paterson, D. L., Pearson, A., Perry, C., Pike, R., ... Woodford, N. (2010). Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: a molecular, biological, and epidemiological study. *The Lancet Infectious Diseases*, 10(9), 597–602. [https://doi.org/10.1016/S1473-3099\(10\)70143-2](https://doi.org/10.1016/S1473-3099(10)70143-2)
- Lilly, J., & Camps, M. (2015). Mechanisms of theta plasmid replication. *Microbiology Spectrum*, 3(1), PLAS-0029-2014. <https://doi.org/10.1128/MICROBIOLSPEC.PLAS-0029-2014>
- Logan, L. K., & Weinstein, R. A. (2017). The epidemiology of carbapenem-resistant Enterobacteriaceae: The impact and evolution of a global menace. *The Journal of Infectious Diseases*, 215(suppl_1), S28–S36. <https://doi.org/10.1093/INFDIS/JIW282>
- Novick, R. P. (1987). Plasmid incompatibility. *Microbiological Reviews*, 51(4), 381–395. <https://doi.org/10.1128/MR.51.4.381-395.1987>
- Partridge, S. R., Kwong, S. M., Firth, N., & Jensen, S. O. (2018). Mobile genetic elements associated with antimicrobial resistance. *Clinical Microbiology Reviews*, 31(4), e00088-17. <https://doi.org/10.1128/CMR.00088-17>
- Pfeifer, E., Moura De Sousa, J. A., Touchon, M., & Rocha, E. P. C. (2021). Bacteria have numerous distinctive groups of phage–plasmids with conserved phage and variable plasmid gene repertoires. *Nucleic Acids Research*, 49(5), 2655–2673. <https://doi.org/10.1093/NAR/GKAB064>
- Pinilla-Redondo, R., Russel, J., Mayo-Muñoz, D., Shah, S. A., Garrett, R. A., Nesme, J., Madsen, J. S., Fineran, P. C., & Sørensen, S. J. (2022). CRISPR-Cas systems are widespread accessory elements across bacterial and archaeal plasmids. *Nucleic Acids Research*, 50(8), 4315–4328. <https://doi.org/10.1093/NAR/GKAB859>
- Rankin, D. J., Rocha, E. P. C., & Brown, S. P. (2011). What traits are carried on mobile genetic elements, and why. In *Heredity*, 106(1), 1–10. <https://doi.org/10.1038/hdy.2010.24>

- Redondo-Salvo, S., Fernández-López, R., Ruiz, R., Vielva, L., de Toro, M., Rocha, E. P. C., Garcillán-Barcia, M. P., & de la Cruz, F. (2020). Pathways for horizontal gene transfer in bacteria revealed by a global map of their plasmids. *Nature Communications*, 11(1), 3602. <https://doi.org/10.1038/S41467-020-17278-2>
- Rock, C., & Donnenberg, M. S. (2014). Human pathogenic Enterobacteriaceae. *Reference Module in Biomedical Sciences*. <https://doi.org/10.1016/B978-0-12-801238-3.00136-7>
- Rodríguez-Beltrán, J., DelaFuente, J., León-Sampedro, R., MacLean, R. C., & San Millán, Á. (2021). Beyond horizontal gene transfer: the role of plasmids in bacterial evolution. *Nature Reviews Microbiology* 2021 19:6, 19(6), 347–359. <https://doi.org/10.1038/s41579-020-00497-1>
- Rodríguez-Beltrán, J., Sørum, V., Toll-Riera, M., de la Vega, C., Peña-Miller, R., & Millán, Á. S. (2020). Genetic dominance governs the evolution and spread of mobile genetic elements in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 117(27), 15755–15762. <https://doi.org/10.1073/PNAS.2001240117>
- Rousset, F., Cabezas-Caballero, J., Piastra-Facon, F., Fernández-Rodríguez, J., Clermont, O., Denamur, E., Rocha, E. P. C., & Bikard, D. (2021). The impact of genetic diversity on gene essentiality within the *Escherichia coli* species. *Nature Microbiology*, 6(3), 301–312. <https://doi.org/10.1038/s41564-020-00839-y>
- Rozwandowicz, M., Brouwer, M. S. M., Fischer, J., Wagenaar, J. A., Gonzalez-Zorn, B., Guerra, B., Mevius, D. J., & Hordijk, J. (2018). Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. *Journal of Antimicrobial Chemotherapy*, 73(5), 1121–1137. <https://doi.org/10.1093/JAC/DKX488>
- Smillie, C., Garcillán-Barcia, M. P., Francia, M. V., Rocha, E. P. C., & de la Cruz, F. (2010). Mobility of Plasmids. *Microbiology and Molecular Biology Reviews*, 74(3), 434–452. <https://doi.org/10.1128/mmbr.00020-10>
- Tazzyman, S. J., & Bonhoeffer, S. (2015). Why there are no essential genes on plasmids. *Molecular Biology and Evolution*, 32(12), 3079–3088. <https://doi.org/10.1093/MOLBEV/MSU293>
- Twigg, A. J., & Sherratt, D. (1980). Trans-complementable copy-number mutants of plasmid ColE1. *Nature* 1980 283:5743, 283(5743), 216–218. <https://doi.org/10.1038/283216a0>
- van Duin, D., & Doi, Y. (2016). The global epidemiology of carbapenemase-producing Enterobacteriaceae. *Virulence*, 8(4), 460–469. <https://doi.org/10.1080/21505594.2016.1222343>
- Werren, J. H. (2011). Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proceedings of the National Academy of Sciences of the United States of America*, 108(SUPPL. 2), 10863–10870. <https://doi.org/10.1073/PNAS.1102343108>

Thesis outline

The objective of this thesis is to study the evolution of plasmid gene content, with a specific focus on essential genes and antibiotic resistance genes. Throughout this study, I aimed to explore several critical questions. How prevalence is essential gene on plasmids in *E. coli*? What is the evolutionary history and the potential function of the plasmid-encoded essential genes? What is the distribution of different antibiotic resistance genes on plasmids compared to chromosomes? What is the driving force of plasmid ARG content evolution and how is the trends and barriers? To address these questions, I conducted three separate projects that comprise the three following chapters of this thesis.

Chapter one focuses on essential genes on plasmids in *Escherichia coli*. In this chapter, I used publicly available *E. coli* genomes and reported essential genes to investigate the prevalence and distribution of essential genes on plasmids. I also constructed the evolutionary history of plasmid-encoded essential genes and their homologs on chromosomes. Our research questions included: can essential genes be found on plasmids in *E. coli*, and what is the origin of these genes? I found that essential genes in *E. coli* are rarely found on plasmids, except for those that serve functions for the plasmids themselves. Furthermore, we discovered that the acquisition of essential genes from the chromosome can lead to plasmid extinction, likely due to the dose effect of the redundant gene copy.

During the investigation of shared gene content between plasmids and chromosomes in Chapter One, I discovered two distinct groups of antibiotic resistance genes. One group was found predominantly on chromosomes, while the other was primarily found on plasmids. In **chapter two**, I conducted a systematic study of the distribution of antibiotic resistance genes on chromosomes and plasmids. I used publicly available complete genomes of strains from the genera *Escherichia*, *Salmonella*, and *Klebsiella*. My analysis revealed that one third of the antibiotic resistance genes were shared between chromosomes and plasmids. Furthermore, I utilized phylogenetic analyses to confirm that the shared genes are resulted from transfer events at various levels. I also combined the classification of chromosome genes into core or accessory groups with the gene transfer data. Together with the resistance mechanism encoded by the genes, we proposed that functional complexity is an important determinant of antibiotic resistance gene transferability.

Chapter three was motivated by our observations in chapters one and two regarding the prevalence of plasmids encoding multiple antibiotic resistance genes and the lateral transfer events. I focused on the evolution of antibiotic resistance gene content on plasmids driven by lateral gene transfer. Utilizing the same genomic data as in chapter two, I examined the co-occurring antibiotic resistance gene pairs on the same plasmids. I found that these frequently co-occurring pairs were conserved across genomes and corresponded mostly to

transposable elements. I then extended my study from two to multiple antibiotic resistance genes and found that these genes were densely encoded in genomic islands, particularly on plasmids. By combining plasmid size and plasmid gene sharing patterns, I was able to address the long-standing question of why plasmid size has a bimodal distribution. My analysis revealed that large and small plasmids have distinct characteristics and evolutionary histories, with transfer of antibiotic resistance genes primarily limited to large plasmids and rarely occurring in small plasmids.

In the end I discuss important results and unresolved questions in my studies to close this thesis.

Chapter I - Essential gene acquisition destabilizes plasmid inheritance

Tanita Wein^{*¶}, Yiqing Wang[¶], Myriam Barz, Fenna T. Stücker, Katrin Hammerschmidt, Tal Dagan

Institute of General Microbiology, Kiel University, Kiel, Germany

[¶] These authors contributed equally to this work.

^{*} Corresponding author

Manuscript published in ***PLoS Genet* 2021**. <https://doi.org/10.1371/journal.pgen.1009656>

Abstract

Extra-chromosomal genetic elements are important drivers of evolutionary transformations and ecological adaptations in prokaryotes with their evolutionary success often depending on their 'utility' to the host. Examples are plasmids encoding antibiotic resistance genes, which are known to proliferate in the presence of antibiotics. Plasmids carrying an essential host function are recognized as permanent residents in their host. Essential plasmids have been reported in several taxa where they often encode essential metabolic functions; nonetheless, their evolution remains poorly understood. Here we show that essential genes are rarely encoded on plasmids; evolving essential plasmids in *Escherichia coli* we further find that acquisition of an essential chromosomal gene by a plasmid can lead to plasmid extinction. A comparative genomics analysis of *Escherichia* isolates reveals few plasmid-encoded essential genes, yet these are often integrated into plasmid-related functions; an example is the GroEL/GroES chaperonin. Experimental evolution of a chaperonin-encoding plasmid shows that the acquisition of an essential gene reduces plasmid fitness regardless of the stability of plasmid inheritance. Our results suggest that essential plasmid emergence leads to a dose effect caused by gene redundancy. The detrimental effect of essential gene acquisition on plasmid inheritance constitutes a barrier for plasmid-mediated lateral gene transfer and supplies a mechanistic understanding for the rarity of essential genes in extra-chromosomal genetic elements.

Introduction

Plasmids are autonomously replicating genetic elements that are prominent in prokaryotes and have been studied extensively due to their contribution to lateral gene transfer. Plasmid invasion is often accompanied by the acquisition of novel traits that enable bacteria to survive under specific conditions or colonize specific ecological niches. Plasmids carrying beneficial genes, e.g., those that supply resistance against antibiotics (e.g., (Heuer et al., 2011; Blau et al., 2018)) or heavy metals (e.g., (Gullberg et al., 2014; Dziewit et al., 2015)) enable their host to survive under transient selective conditions. A strong selection for the plasmid-encoded trait over many generations was shown to lead to co-adaptation of the plasmid and the host, and the evolution of stable plasmid inheritance (Bouma & Lenski, 1988; Heuer et al., 2007; Sota et al., 2010; Loftie-Eaton et al., 2017; Stalder et al., 2017). Nonetheless, environmental conditions are rarely constant; a decrease in the strength of selection for plasmid-encoded beneficial genes—e.g., due to fluctuating abundance of growth limiting factors—may lead to plasmid loss and extinction. In contrast, plasmids that supply the host with essential functions, i.e., whose benefit to the host is less dependent on temporary environmental conditions, may persist over longer time scales and become an integral component of the lineage genome in the form of chromids (P. W. Harrison et al., 2010). The level of gene essentiality is defined as the extent

to which a gene is required for the reproduction of an organism (Rancati et al., 2018). Borrowing that definition, plasmids that encode for essential genes may be thus considered as vital for the host proliferation. Examples include plasmids encoding genes along the biosynthesis pathways of essential amino acids (e.g., (Gil et al., 2006; diCenzo et al., 2013)) and components of the ribosome machinery (Anda et al., 2015). Nevertheless, it has been previously suggested that plasmids encoding essential genes should be rare in nature as their adaptation to the host is accompanied by reduction of host fitness due to plasmid loss (Tazzyman & Bonhoeffer, 2015). Thus, the evolution of plasmids encoding essential genes may depend on the evolution of stable plasmid-host interactions.

In the absence of selection for plasmid-encoded genes, plasmid persistence largely depends on plasmid stability that comprises plasmid replication and segregation. Indeed, plasmids that have a negligible (i.e., neutral) effect on host fitness may evolve stable inheritance and consequently gain long-term persistence in the population—also in the absence of positive selection (Wein, Hülter, et al., 2019). Stably inherited plasmids may serve as precursors for the evolution of larger plasmids carrying beneficial or essential genes. Notably, the essential plasmids reported in the literature so far are characterized by a stable vertical inheritance. The evolution of such essential plasmids may thus follow two alternative routes (Hülter et al., 2017): *essentiality first*, where the plasmid is initially essential to the host; a plasmid carrying an essential function may persist in the population over long timescales and eventually evolve a stable plasmid reproduction within the host. In the second route, termed *stability first*, a plasmid that initially evolved a stable reproduction cycle may subsequently evolve into an essential plasmid following an essential gene gain. Notwithstanding, as essential genes are typically encoded in the chromosomes (Rousset et al., 2021), the presence of an essential gene on a plasmid may be comparable to the effect of gene duplication. In addition, the expression level of plasmid-encoded genes may be amplified following an increase in plasmid copy number (e.g., (Nicoloff et al., 2019)). Therefore, the evolution of essential genes on plasmids is likely under selection pressure that is typical to gene duplication, such as dose effect (reviewed in (Andersson et al., 2015)). Evidently, most genes ($89 \pm 8\%$) in prokaryotic genomes are found in a single copy (Bratlie et al., 2010).

Here we combine computational genomic analysis and experimental work to investigate the evolution of essential plasmids in the genus *Escherichia*. First, we investigate the frequency of shared genes between plasmids and chromosomes in *Escherichia* as well as the distribution of essential genes in *Escherichia* plasmids. Furthermore, we compare both the *essentiality* and *stability first* evolutionary scenarios using an evolution experiment in *E. coli* where we follow the evolution of stable and unstable plasmids encoding the chaperonin that is indispensable in *E. coli*.

Results

Plasmid gene content is rarely shared with the host chromosome

To study the distribution of chromosomal homologs to plasmid genes we examined isolates of *Escherichia* that are available as complete genome sequences at NCBI (S1 Table). Of the total 599 isolates, 416 isolates harbor between 1 and 10 plasmids. All 88,493 protein-coding genes encoded on plasmids were clustered into 4,780 protein families by sequence similarity. Notably, we did not detect any universal plasmid gene—the largest plasmid protein family includes homologs from 41% (425/1035) plasmids. Additionally, we identified chromosomal homologs to the plasmid protein families using a sequence similarity search. This revealed that 2,682 (56%) of the plasmid protein families had no chromosomal homolog, while the remaining 2,098 (44%) families had a chromosomal homolog in at least one *E. coli* isolate. Previously it has been suggested that gene duplication via gene transfer may be disadvantageous in prokaryotes due to dose effect (e.g., (Jain et al., 1999; Sorek et al., 2007)). Indeed, the distribution of protein family members on chromosomes and plasmids shows that the level of gene sharing between both replicon types is rather low (Fig 1A). A total of 383 (8%) protein families have a chromosomal homolog in a single isolate, and 544 (11%) of the plasmid protein families have a chromosomal homolog in 2–320 isolates (S1 Fig). Thus, in total 927 (19%) of the plasmid protein families have a chromosomal homolog in the same isolate (i.e., the plasmid host), of those, 133 (3%) families correspond to transposable elements (Fig 1A). The most abundant protein family among the transposable elements is that of IS66 that is found in varying copy numbers on 395 examined chromosomes and 302 plasmids.

Essential genes are rarely encoded on plasmids

To study the evolution of essential genes encoded on plasmids, we examined the distribution of genes previously recognized as essential in *E. coli* on chromosomes and plasmids, where the extent of gene essentiality may depend on the environmental and nutritional conditions. The set of essential genes we tested here comprises 504 protein-coding genes (S2 Table); of those, 394 protein-coding genes were identified as essential in *E. coli* K-12 under standard laboratory conditions via gene knockout (Baba et al., 2006) or transposon mutagenesis (Goodall et al., 2018). Additional 110 protein-coding genes were identified as essential in a collection of 18 *E. coli* isolates by CRISPR interference of gene expression ((Rousset et al., 2021); only genes identified as essential in at least one out of the three tested growth media were included; S2 Table). Searching for plasmid-encoded homologs to the essential genes using sequence similarity revealed that only 17 (3%) of the essential genes had a homolog on plasmids (Fig 1A and Table 1). The 464 *Escherichia* plasmids encoding essential genes are typically of medium plasmid size (median: 106Kb), 32% of them are non-mobile and only 4% carry in addition an antibiotic resistance gene (Fig 1B). Importantly, all isolates where we

identified a plasmid homolog of an essential gene also harbored the chromosomal copy of that gene (except one isolate carrying only *metG* on the plasmid).

Fig 1. The distribution of plasmid-encoded essential genes on plasmids and chromosomes. **A**, The distribution of *Escherichia* protein families on plasmids in chromosomes (in grey). Essential genes (red) are found in most of the chromosomes and rarely on plasmids, with *ssb* as an exception. Transposable elements (blue) are frequent on both chromosomes and plasmids. IS66 is the most widely spread gene in *Escherichia* strains. IS66 was split into several protein families in our analysis, which are depicted by multiple data points. Antibiotic resistance (AMR) genes (yellow) are presented for comparison. The AMR genes can be roughly divided into two groups: the first group aligns along the y-axis hence it is more frequently found on chromosomes; those chromosome-encoded AMR genes are typically related to persistence and resilience functions. The second group aligns along the x-axis hence it is more frequently found on plasmids; those plasmid-encoded genes are typically related to antibiotic resistance functions (Wein et al. 2020). **B**, Distribution of plasmid size, mobility group and AMR for plasmids encoding an essential gene. Most of the *ssb*-coding plasmids are conjugative (red) while the majority of *groEL/S* coding plasmids are mobilizable (green). The remaining essential genes are encoded on plasmids that are often non-mobilizable (and non-AMR) (purple).

Escherichia Ssb evolutionary history reveals a deep divergence between the chromosomal and plasmid homologs (Fig 2A). Additionally, the phylogenetic tree suggests several events of lateral gene transfer (LGT) of *ssb* from plasmids to chromosomes as indicated by eight chromosomal *ssb* homologs that branch with plasmid homologs (Fig 2A). To further validate the LGT events, we compared the reconstructed tree likelihood to an alternative constrained topology having a split between chromosomal and plasmid *ssb* homologs. Since the alternative *ssb* topology was rejected ($P = 0.0074$, using AU test), we conclude that, indeed, the *ssb* evolution included multiple LGT events. According to the inferred *ssb* phylogeny, a recent translocation of the plasmid-encoded *ssb* to the chromosome occurred in the seven strains, all of them retained the original chromosomal copy and lacked a plasmid-encoded *ssb*; these seven isolates thus have two *ssb* copies encoded in their chromosome. These *ssb* LGT events correspond to the translocation of an essential plasmid gene to the chromosome.

Further phylogenetic reconstruction of the remaining plasmid homologs to essential genes revealed evidence for a recent transfer from the chromosome to the plasmid in six of the protein families including *ftsH*, *dnaC*, *metG*, *glnS*, *proA* and *proB* (Table 1 and S2–S7 Figs). The evolution of those gene families is evidence for rare duplication of essential genes onto a resident plasmid. In contrast, the phylogenetic trees of nine essential genes including *groEL*, *groES*, *ubiX*, *nrdB*, *nrdA*, *def*, *ilvC*, *lolC* and *kdsD* revealed a clear split between the chromosomal and plasmid homologs (Table 1 and Figs 2C and S8–S14). In two of these gene families—*groEL* and *groES*—we observed gene translocation from plasmid to chromosome (Table 1 and Fig 2C). The divergence between the plasmid and the chromosomal homologs in these gene families suggests that the plasmid homologs have undergone a sub- or neo-functionalization (i.e., similar to the plasmid Ssb).

Previous studies suggested that laterally transferred genes are often non-functionalized due to codon-usage incompatibility or the absence of regulatory elements (Hao & Brian Golding, 2006). To investigate whether plasmid-encoded homologs of essential genes are pseudogenes, we inferred the strength of selection pressure on the plasmid branch relative to chromosome branch in all relevant protein families. Whereas our results revealed the presence of relaxed selection on branches leading to plasmid homologs in *groEL* and *nrdA* (Table 1 and S15 and S16 Figs), we could not detect evidence for non-functionalization of those genes (e.g., frameshift or truncation). Our results thus suggest that the observed relaxation of selection is likely associated with sub- or neo-functionalization of the essential genes encoded on plasmids.

Table 1. Data and phylogenetic analysis of essential genes encoded on plasmids.

Gene	Product	Essential list	No. Plasmids	No. Isolates	Relaxed selection	Tree topology
<i>ssb</i>	single-stranded DNA-binding protein	TKP	420	320	n.a.	Mixed
<i>groL</i>	chaperonin GroEL	TP	11	11	On plasmid branch	One translocation to Chromosome
<i>groS</i>	co-chaperone GroES	TKP	13	13	No evidence	One translocation to Chromosome
<i>dnaB</i>	replicative DNA helicase	TKP	31	31	No evidence	Two translocations to Chromosome
<i>ubiX</i>	aromatic acid decarboxylase	T	1	1	n.a.	Split between plasmids and chromosomes
<i>nrdB</i>	ribonucleotide-diphosphate reductase subunit beta	TKP	20	20	No evidence	Split between plasmids and chromosomes
<i>nrdA</i>	ribonucleoside-diphosphate reductase subunit alpha	TKP	19	19	On plasmid branch	Split between plasmids and chromosomes
<i>ftsH</i>	ATP-dependent metalloprotease FtsH	TKP	2	2	n.a.	Three transfers to plasmid
<i>def</i>	peptide deformylase	TKP	13	13	No evidence	Split between plasmids and chromosomes
<i>dnaC</i>	DNA replication protein DnaC	TKP	1	1	n.a.	Single transfer to plasmid
<i>metG</i>	methionine—tRNA ligase	TKP	19	19	No evidence	Mixed
<i>ilvC</i>	ketol-acid reductoisomerase, NAD(P)-binding	M9	3	3	No evidence	Split between plasmids and chromosomes
<i>glnS</i>	glutamine—tRNA ligase	TKP	5	5	n.a.	Five transfers to plasmid
<i>lolC</i>	lipoprotein-releasing ABC transporter permease subunit	TKP	2	2	n.a.	Split between plasmids and chromosomes
<i>proB</i>	gamma-glutamate kinase	M9	2	2	n.a.	Two transfers to plasmid
<i>proA</i>	gamma-glutamylphosphate reductase	M9	2	2	n.a.	Two transfers to plasmid
<i>kdsD</i>	D-arabinose 5-phosphate isomerase	LB, GMM	1	1	n.a.	Split between plasmids and chromosomes

<https://doi.org/10.1371/journal.pgen.1009656.t001>

Essential list indicates the source of essential genes: TKP is TraDIS-Keio-PEC, TP is TraDIS-PEC, T is TraDIS (Goodall et al. 2018); LB, M9 and GMM means this gene is essential in each of the three medias (Rousset et al. 2020). Relaxed selection shows evidence for relaxed selection on plasmid homologs' branch was found using HyPhy-RELAX (Murrell et al. 2012). Tree topology shows the conclusions from the phylogenetic reconstruction of essential genes on plasmids and chromosomes: *split between plasmids and chromosomes* describe phylogenies that had a deep split between plasmid and chromosomal homologs (i.e., they are diverged); LGT events from chromosome to plasmid are labeled by *transfer to plasmid*; LGT events from plasmid to chromosome are labeled by *transfer to chromosome*; *mixed* is used to label phylogenies that contain deep divergence and LGT events. For gene trees with evidence for gene transfer (*groEL*, *groES*), we further tested the support in the plasmid/chromosome split while excluding the putatively transferred gene. In all tested cases the result showed that the constrained tree could not be rejected in a topology test, thus validating the gene transfer inference. All of the transfer event from plasmid to chromosome are better described as translocation of the plasmid homolog to chromosome.

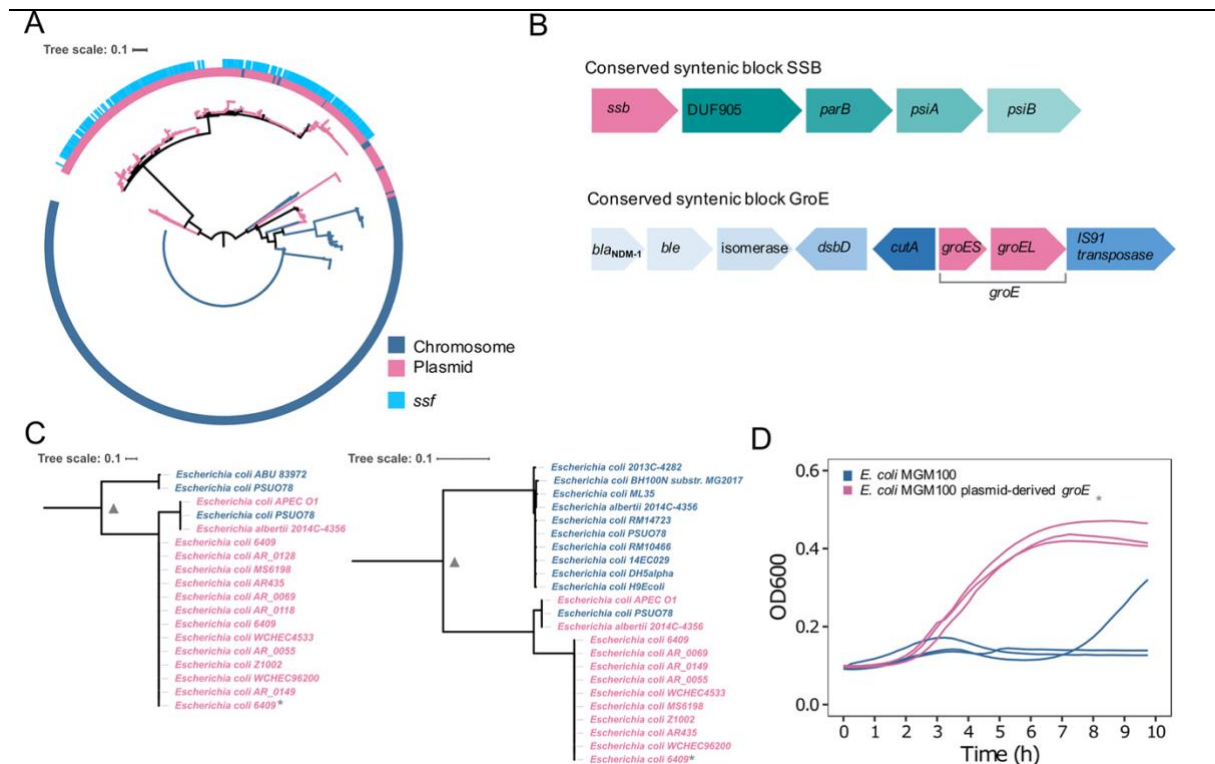


Fig 2. Phylogeny of plasmid-encoded essential genes in *Escherichia*. **A**, Phylogeny of the single-stranded DNA-binding protein Ssb. Ssb homologs encoded on plasmid are shown in pink and encoded on chromosome are shown in blue. The plasmid *ssb* gene termed *ssf* is most abundant across the plasmid Ssb homologs. **B**, The conserved neighborhood (conserved syntentic block, CSB) of *ssb* and the chaperone *groE* encoded on plasmids (S3 and S4 Tables). **C**, Phylogeny of the chaperonin GroES (left) and GroEL (right). The triangle symbol marks the branch split that was constrained in the test for an alternative tree topology. **D**, Growth measurements of *E. coli* MGM100 and MGM100 carrying the plasmid-derived GroE (marked by *). The exceptional growth behavior observed in one *E. coli* MGM100 replicate should be considered an outlier since we found no evidence for other explanations to that result.

The plasmid-encoded chaperonin complements the chromosomal function

To further examine the evolution of plasmids encoding essential genes, we studied the implications of essential gene gain on a plasmid by focusing on the chaperonin genes *groEL* and *groES*. The chaperonin is universally encoded in eubacterial chromosomes with only rare exceptions (e.g., *Mollicutes* (Clark & Tillier, 2010)) and plasmids encoding the chaperonin genes have been only reported in *Rhizobium* (Wallington & Lund, 1994). The *groES* and *groEL* genes are typically encoded within the *groE* operon. The GroE chaperonin plays a major role in the bacterial protein-folding pathway and is constitutively expressed in *E. coli* and essential for growth under all conditions (Fayet et al., 1989). An examination of the *groE* genomic neighborhood in *Escherichia* plasmids revealed a conserved neighborhood that includes, in addition to *groEL* and *groES* a *bla*_{NDM-1}, *ble* (bleomycin binding protein Ble-MBL), phosphoribosylanthranilate isomerase, twin-arginine translocation (TAT) pathway signal sequence domain protein and *cutA* (divalent-cation tolerance protein) (Fig 2B and S4 Table). Indeed, a similar integron comprising *groE* has been previously reported in *Escherichia*

(Sekizuka et al., 2011; Zhang et al., 2018). Our results thus indicate that *groE* is likely not a pseudogene. To further test the function of the plasmid-derived *groE*, we cloned the gene sequence into a small model plasmid. The plasmid was then introduced into *E. coli* MGM100 (Fig 2D), which encodes the *groE*-operon under an inducible P_{BAD} promoter, hence it is only viable in medium supplemented with arabinose (i.e., P_{BAD} induction) or when it is complemented with *groE* (McLennan & Masters, 1998). In competition with *E. coli* MG1655, *E. coli* MGM100 has no measurable fitness disadvantage (S18 Fig). To test whether the plasmid-derived *groE* can complement the silenced *groE* in *E. coli* MGM100, we quantified the growth of this strain with and without supplemented arabinose. In addition, we recorded the overnight plasmid loss, as a plasmid that is essential to its host cannot be lost. Our results show that the plasmid can indeed complement the chromosomal GroE function, hence we conclude that the plasmid-encoded *groE* functions as a chaperonin.

Essential gene acquisition destabilizes plasmid inheritance

Our results so far show that essential genes are rarely found on plasmids with the plasmid-encoded GroE in *Escherichia* as a rare exception. Such plasmids that encode an essential gene may evolve via two possible routes—in the *stability first* a preliminary stably inherited plasmid gains an essential gene, while in the *essentiality first* a plasmid that encodes an essential gene gains by that a stable inheritance. To study the evolutionary consequences of the acquisition of an essential chromosomal gene by plasmids, we conducted an evolution experiment with *E. coli* K12 MG1655 carrying small plasmids encoding a chromosomal *groE* copy. To investigate the implications of plasmid stability versus essentiality in the evolution of essential plasmids, we compared the evolutionary dynamics of stable and unstable plasmids in two genetic host backgrounds.

Our unstable model plasmid pCON originated from the pBBR1 backbone that is widely spread and associated with antibiotic resistance (Antoine & Locht, 1992; Wein, Hülter, et al., 2019). The plasmid pCON encodes *nptII* that is constitutively expressed and confers resistance to kanamycin; it has a negligible effect on the host fitness and is characterized by an unstable inheritance in the population (Wein, Hülter, et al., 2019). Previously, we showed that pCON instability is caused by transcription-replication conflicts of the resistance gene transcription and plasmid replication. A stable inheritance of pCON is provided when transcription is silenced or when both machineries are physically separated, i.e., by a plasmid size increase (Wein, Hülter, et al., 2019). Our stable model plasmid pCON-S is a derivative of pCON that is characterized by a stable inheritance due to a short DNA insertion between the origin of replication and the *nptII* gene; similarly to pCON, it has a negligible effect on the host fitness (Wein et al., 2020). We equipped the unstable (pCON) and stable (pCON-S) plasmids with the chromosomal *groE* (including its native promoter) resulting in pGroE and pGroE-S (Fig

3A). The plasmids pGroE and pGroE-S were introduced into two host strains: *E. coli* MG1655 (wt) and *E. coli* MGM100 (Fig 3A). Thus, *groE*-encoding plasmids are essential in *E. coli* MGM100 while they are non-essential in *E. coli* MG1655.

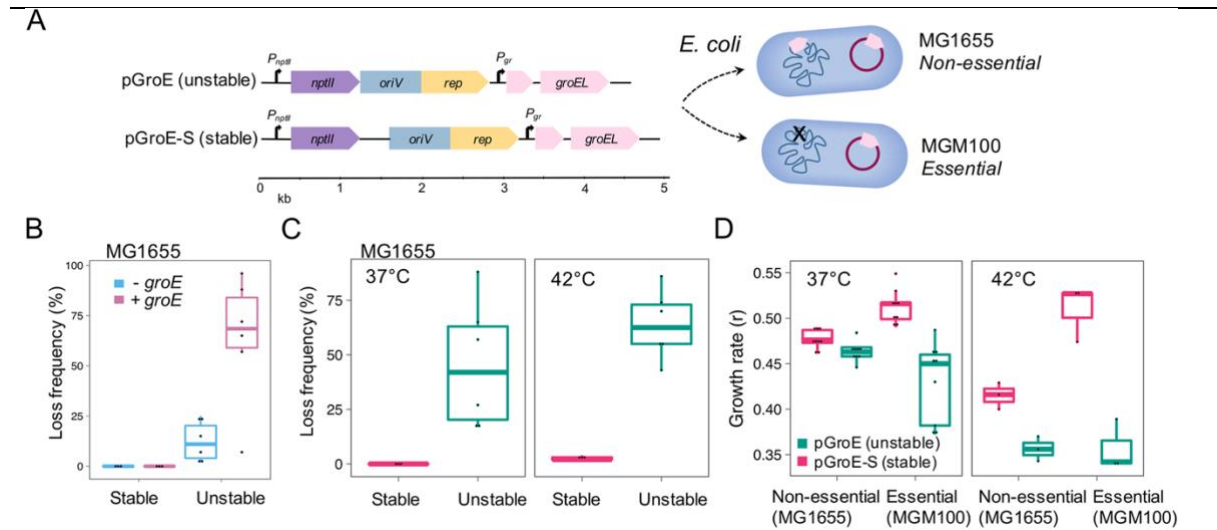


Fig 3. Characterization of plasmids encoding an essential gene. **A**, Genomic map of unstable pGroE and stable pGroE-S plasmids that were introduced into *E. coli* MG1655 (plasmid is non-essential) and *E. coli* MGM100 (plasmid is essential) hosts. **B**, Plasmid loss frequency of plasmids lacking (-) *groE* or encoding (+) *groE* in their genome in the host *E. coli* MG1655 ($H_0: L_+ > L_-$, $P = 0.0247$ using Wilcoxon test $n = 6$). **C**, Plasmid loss frequency of pGroE-S (stable) and pGroE (unstable) after incubation at 37°C or at 42°C ($H_0: L_{42} > L_{37}$, $P = 0.0091$ using Wilcoxon test, $n = 6$). **D**, Growth rates of host strains *E. coli* MG1655 and *E. coli* MGM100 carrying pGroE (unstable) or pGroE-S (stable) after growth at 37°C and 42°C ($H_0: Gr_{42} > Gr_{37}$, $P = 0.051$ using Wilcoxon test, $n = 9$).

To characterize the effect of GroE on plasmid stability, we quantified the loss frequency of pGroE (unstable) and pGroE-S (stable). Our results show that the pGroE loss frequency was significantly higher in comparison to the loss frequency of the plasmid pCON lacking GroE in *E. coli* MG1655 (Fig 3B, $P = 0.0247$ using Wilcoxon test, $n = 6$). Similar to pCON-S, the plasmid pGroE-S was stably maintained in *E. coli* MG1655 in an overnight incubation (Fig 3B). Both plasmids were maintained in the host strain *E. coli* MGM100, where the plasmid is essential for host survival. When the medium was supplemented with arabinose—rendering the plasmid dispensable—we observed a plasmid loss frequency of $21 \pm 3\%$ (SE, $n = 6$) for the unstable pGroE and no loss of the stable pGroE-S ($n = 6$; S19 Fig). These results are in line with the loss frequency of pGroE in *E. coli* MG1655, in which the plasmid is dispensable. Our results show that the introduction of a redundant *groE* copy on an unstable plasmid further destabilizes the plasmid maintenance and may lead to an increase in plasmid loss.

In addition to variation in the extent of plasmid-encoded gene essentiality, the observed plasmid loss may be related to the chaperonin function. The chaperonin GroE is part of the heat shock regulon in *E. coli* hence its expression level is upregulated during growth at high temperature (42°C) (Richmond et al., 1999; Gunasekera et al., 2008). The expression level of plasmid-encoded genes depends not only on their transcriptional regulation but also on the plasmid copy number (PCN). Variation in PCN is thus comparable to gene amplification where fluctuations in PCN lead to variation in the expression level of plasmid-encoded genes, which may enable the host to rapidly adapt to changing environmental conditions (Rodriguez-Beltran et al., 2018; Nicoloff et al., 2019). Thus, under heat stress bacteria that carry plasmid-encoded *groE* are expected to have an advantage over strains with chromosomally encoded *groE* due to a higher GroE abundance in the cell. To evaluate whether heat stress had an effect on plasmid persistence, we measured the plasmid loss frequency after overnight incubation at 42°C. Our results show no measurable plasmid loss for both plasmids when the plasmid is essential (i.e., in *E. coli* MGM100). When the plasmid is not essential (i.e., in *E. coli* MG1655), the pGroE-S remained stable, while the pGroE loss frequency was slightly increased in comparison to pGroE loss at 37°C (Fig 3C). This observation is in line with reduced growth rate of MG1655 at 42°C in comparison to 37°C (Fig 3C, $P = 0.0091$ using Wilcoxon test). Thus, the redundant plasmid-encoded GroE does not confer an advantage in *E. coli* MG1655 at 42°C. Nonetheless, a comparison of the growth rate (r) between *E. coli* MGM100 pGroE-S and MG1655 pGroE-S at 42°C revealed a growth advantage of *E. coli* MGM100 pGroE-S over MG1655 pGroE-S (Fig 3D, $P = 0.051$ using Wilcoxon test) and over the plasmid-free wildtype ($r_{wt} = 0.42 \pm 0.01$, SE, $n = 9$). The growth advantage observed for expression *E. coli* MGM100 pGroE-S likely stems from differences in *groE* expression dynamics when it is expressed solely from the plasmid locus. Additionally, the growth rate of both strains carrying the unstable pGroE was decreased (Fig 3D, $P = 0.0007$ using Wilcoxon test). The reduced growth rate of *E. coli* MGM100 carrying the unstable pGroE is well explained by the occurrence of cell death following plasmid loss (S20 Fig).

We conclude that the presence of a redundant essential gene on an unstable plasmid may destabilize plasmid maintenance in the host regardless of growth conditions. The maintenance of stable plasmids can remain, however, unchanged in the presence of an essential gene. Furthermore, variation of the plasmid-encoded gene expression level due to fluctuations in the plasmid copy number may confer an advantage to the host under specific environmental conditions.

Essential plasmids persist over time regardless of their stability

Constant positive selection for a beneficial gene is expected to result in gene transfer from the plasmid to the chromosome and consequently plasmid loss (Bergstrom et al., 2000; E.

Harrison & Brockhurst, 2012). To test the effect of plasmid essentiality on plasmid evolution, we conducted an evolution experiment comparing both unstable (pGroE) and stable (pGroE-S) plasmids in host backgrounds where the plasmids are essential or non-essential. The experiment was performed with nine replicates of all plasmid and host combinations for approximately 320 generations in a serial batch transfer approach.

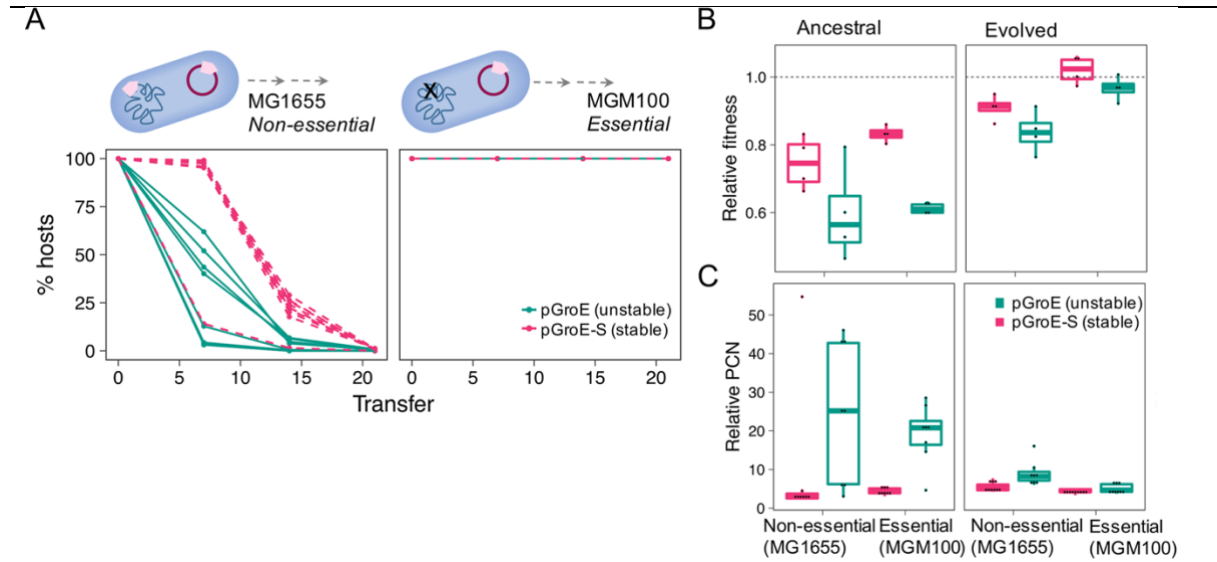


Fig 4. Evolution of plasmids encoding an essential gene. **A**, Evolution experiment of pGroE (unstable) and pGroE-S (stable) in the host strains *E. coli* MG1655 (non-essential, left) and MGM100 (essential, right). **B**, Relative fitness of plasmid-carrying ancestral and evolved populations. Pairwise competition experiments between the plasmid-carrying strains and tagged wild-type (Tm^r). A significant negative fitness effect could be observed for all ancestral populations (w : 0.69, $P = 1.526 \times 10^{-5}$ using Wilcoxon test, $n = 4$ per population) while the fitness of evolved populations increased compared to the ancestral state (w : 0.93, $P = 9.05 \times 10^{-7}$ using Wilcoxon test, $n = 4$ per population). **C**, Plasmid copy number (PCN) of plasmid-carrying ancestral and evolved populations. The median pGroE PCN is significantly different between the ancestral and evolved populations ($P_{MG1655} = 0.03$, $P_{MGM100} = 0.0016$, using Wilcoxon test, $n = 9$ per population), with the PCN being lower in the evolved populations in both hosts.

The results of the evolution experiment reveal that when the plasmids were essential for host survival, they were maintained over time regardless of their stability (Fig 4A). Sequencing a sample of the evolved populations, we did not detect evidence for *groE* transfer to the chromosome or mutations in the inducible $P_{BAD-groE}$ promoter (S5 Table). When the plasmid was non-essential to the host (i.e., in *E. coli* MG1655), the highly unstable pGroE was lost rapidly while the stable pGroE-S decreased in frequency after several transfers (Fig 4A). Thus, non-essential plasmids were lost regardless of their initial stability in the host (i.e., stable

inheritance). To test whether plasmid loss in our experiment led to plasmid extinction, we exposed the evolved populations to antibiotics (kanamycin 25 µg/ml; S21 Fig), which was then followed by serial transfer under non-selective conditions (i.e., without antibiotics). This experiment showed that none of the plasmids went extinct over time. In addition, we observed a similar plasmid loss pattern, where the unstable plasmid was gradually lost from the population and the stable plasmid was maintained at a higher frequency (S21 Fig). Our results thus suggest that the decreased frequency of plasmid carrying hosts is not due to plasmid instability, rather is a result of host fitness differences between plasmid-carrying and plasmid-free cells.

To evaluate the relative fitness of the plasmid-carrying populations, we conducted competition experiments of the ancestral and evolved populations against the plasmid-free wildtype strain. Indeed, we observed a strong negative fitness effect of all *groE*-plasmids on host fitness (Fig 4B; w : 0.697, $P = 1.52 \times 10^{-5}$ using Wilcoxon test, $n = 4$ per population). Nonetheless, the negative impact on fitness decreased over time and the evolved populations revealed a higher relative fitness than their ancestors (Fig 4B, w : 0.93, $P = 9.05 \times 10^{-7}$ using Wilcoxon test, $n = 4$ per population). A reduction in the plasmid effect on the host fitness might have been associated with a decrease in PCN over time. Indeed, this is the case for the unstable plasmids in both host strains where the average PCN decreased about 10-fold (Fig 4C). Previously, we showed that a reduction of PCN for pCON plasmids is often a result of reduction in the plasmid tendency to form multimers (Wein et al., 2019; 2020). Moreover, the variance of pGroE PCN (in MG1655) among replicate populations decreased during the experiment (ancestral CV_{pGroE} : 0.73, evolved CV_{pGroE} : 0.35), and so did the loss frequency (S22 Fig), a result that is akin to the loss of plasmid multimers during the evolution experiment (as in (Wein, Hülter, et al., 2019; Wein et al., 2020)). Since we did not detect any high frequency chromosomal variants in the evolved populations (S5 Table; as previously observed for antibiotic resistance plasmids (Wheatley et al., 2021)), we conclude that the decrease in PCN is likely due to an efficient resolution of plasmid multimers in the evolved population.

Our results show that plasmid essentiality due to complementation of the chaperonin function leads to plasmid persistence that is accompanied by an increase in overall fitness of the evolved populations. Nonetheless, carrying a redundant chromosomal gene proved here to be disadvantageous for plasmid persistence regardless of plasmid stability. Consequently, we conclude that plasmids encoding an essential gene may have a higher chance to persist in the population over longer time scale if a chromosomal gene copy is either lacking or functionally divergent.

Discussion

The acquisition of an additional copy of single-copy chromosomal genes has been shown to be often deleterious, hence sensitivity to dose effect has been suggested as a barrier to gene acquisition by lateral gene transfer (Sorek et al., 2007). Newly acquired genes are often initially disconnected from the organisms' metabolic network and their integration may take a long time (Lercher & Pál, 2008). Thus, additional copies of core essential genes can disrupt the cell physiology through protein dosage alterations that change metabolic fluxes (e.g., (Baltrus, 2013)). Regulatory networks in prokaryotes have evolved to respond to feedback loops that include monitoring of transcription level and substrate concentrations. Hence, the acquisition of a redundant gene copy may be deleterious due to indirect effect on the organism metabolism and cellular processes (Jain et al., 1999; Cohen et al., 2011). Notwithstanding, plasmid-mediated dose effect is not always deleterious to the host and may depend on the protein sequence characteristics and interactions with other proteins in the cell (Acar Kirit et al., 2020). Our results for the chaperonin-encoding plasmid reveal that functional redundancy of an essential gene following plasmid acquisition can be highly disadvantageous for host fitness. In our experiment, the GroE abundance is associated with the PCN, hence the reduction in fitness effect of the plasmid in the evolved populations is likely explained by the decrease in PCN that is coupled with a decrease in GroE dose within the cell. An elevated expression level of the chaperonin has been shown to be detrimental for *E. coli* fitness under optimal growth conditions (Fares et al., 2002; Sabater-Muñoz et al., 2015). The fitness cost of GroE overexpression has been suggested to stem from the increased energetic investment in protein folding due to the high chaperonin availability (Bogumil & Dagan, 2012). The plasmid-mediated increase in GroE abundance may furthermore hamper the chaperonin assembly in the cell. Plasmids that gain a redundant essential gene are thus at risk of rapid loss and eventually extinction.

Based on our results, we suggest that the evolution of plasmids encoding essential genes is conditioned by either non-functionalization (or loss) of the chromosomal copy or its functional divergence, e.g., via neo-functionalization or sub-functionalization due to modification of the transcriptional regime. This constraint provides an explanation for the rarity of essential genes in *Escherichia* plasmids. Notably, two of the plasmid-encoded essential genes we report here—*ssb* and *groE*—seem to have undergone a sub-functionalization as part of the plasmid life cycle such that their function is no longer essential for the host. In other words, those essential genes were acquired by the plasmid and repurposed into essential plasmid genes that are required for persistence in the host. The *Escherichia* plasmids encoding *ssb* or *groE* are thus carrying homologs to essential genes but should not be considered essential plasmids. Hence, we predict that essential genes are rarely found on plasmids in other prokaryotic taxa and natural environments. The detrimental effect of essential gene

acquisition on plasmid stability constitutes a barrier for plasmid-mediated lateral transfer of essential genes. Nonetheless, we cannot exclude that plasmids encoding an essential gene may become advantageous under specific environmental conditions. Our results show that the translocation of *groE* to a plasmid increased host fitness during growth at high temperature where *groE* expression is naturally upregulated (Richmond et al., 1999; Gunasekera et al., 2008). Indeed, *groE* may be found on plasmids in nature as in the facultative plant symbiont *Sinorhizobium meliloti*, in which two out of five functional *groE* copies are encoded on the plasmid pSymA and one copy is encoded on the plasmid pSymB (Galibert et al., 2001). The chromosomal *groE1* copy was shown to encode the major housekeeping chaperonin function, yet it could be functionally complemented by the pSymA-encoded *groE2* (Bittner et al., 2007). The reason why multiple *groE* copies are retained in *S. meliloti* remains unknown. It is tenable to speculate that the facultative symbiotic lifestyle of *S. meliloti* entailed a sub-functionalization of the *groE* copies. An example for the translocation of an essential gene from the chromosome to a plasmid was reported in bacterial symbionts of lice that inhabit human, chimpanzee and gorilla (*Riesia* sp.). In these organisms, the B5 vitamin biosynthesis pathway is encoded on a small plasmid that is vital for the bacteria-louse symbiosis. In contrast, symbionts of lice that inhabit old world monkeys (*Puchtella* sp.) are lacking the plasmid and the B5 vitamin pathway is chromosomally encoded (Boyd et al., 2017). The phylogenetic relations of the two symbionts indicate that the chromosomal B5 copy is ancient and the gene translocation to the plasmid occurred during the lice adaptation to inhabit new-world monkeys (Boyd et al., 2017). We hypothesize that the fixation of the gene translocation on the plasmid may be related to the symbiotic *Riesia* lifestyle; a retention of the plasmid encoded gene copy may be advantageous under conditions that require a flexible gene expression level, which can be regulated rapidly by transient changes in plasmid copy number (Rodriguez-Beltran et al., 2018).

Considering the two possible routes for the evolution of essential genes on plasmids—our study supplies evidence that the *essentiality first* route is the more likely scenario. To better understand the difference between the two routes we shift our perspective to consider the fitness of the plasmid. Plasmid loss in our experimental system stems from two main reasons: plasmid instability of the unstable pGroE or fitness disadvantage of the stable pGroE-S-hosts; indeed, incomplete segregation and negative fitness effects are known causes of plasmid loss (e.g., (Levin & Stewart, 1980; Lau et al., 2013)). The comparison between the loss dynamics of the unstable and stable plasmids (Fig 4A) indicates that plasmid loss due to intracellular processes—namely plasmid instability—may be more rapid in comparison to plasmid loss due to processes at the population level—where plasmid-hosts are outcompeted by non-hosts. Our results indicate that the acquisition of an essential chromosomal gene by a plasmid leads to a significant reduction in plasmid fitness, even when the plasmid was highly stable prior to the essential gene acquisition (as in the *stability first* route). Considering the constraint on the

presence of chromosomal essential genes on plasmids, most plasmids are unlikely to be essential for their host.

Methods

Computational genomic analysis

All 599 completely sequenced genomes of *Escherichia* genus strains were downloaded from NCBI (version 2018) and analyzed as described in (Wein et al., 2020). Briefly, plasmid-encoded protein coding genes were clustered into homologous protein families. Sequence similarity of plasmid-encoded protein sequences was assessed from the results of reciprocal best BLAST hits (rBBHs) applying a threshold of E-value $\leq 1 \times 10^{-5}$ (using BLAST version 2.6.0+ [55]). Pairs of rBBHs were further compared by global alignment with needle (version 6.6.0.0; EMBOSS package (Camacho et al., 2009)). Sequence pairs with $\geq 50\%$ identical amino acids were clustered into protein families using the Markov clustering algorithm (MCL) (version 12–135) with the default parameters (Enright et al., 2002). To identify chromosome-encoded homologs of plasmid genes, the sequences of all plasmid protein families were blasted against all the chromosome proteins using BLAST with a threshold of E-value $\leq 1 \times 10^{-10}$. The resulting chromosomal hits were further compared by global alignment using needle; chromosome-encoded protein sequences were considered homologs using a threshold of $\geq 40\%$ identical amino acids. AMR protein families were identified using Resistance Gene Identifier (version 5.1.0), with CARD database (version 3.0.7) (Alcock et al., 2020). Plasmid mobility was inferred depending on the presence of the Mob relaxase and T4SS encoding genes in the plasmid genomes (as listed in (Smillie et al., 2010)). Plasmids were classified as mobilizable if they encode a Mob protein and as conjugative if they have at least 15 *tra* genes. The dataset of essential genes in *E. coli* includes 394 protein-coding genes included in the essential genome of *E. coli* K-12 (Goodall et al., 2018). These genes were identified manually in the genome of *E. coli* K-12 substr. MG1655 according to their gene symbol. This dataset was supplemented with additional 110 protein-coding genes identified as essential in a large-scale study that included multiple *E. coli* isolates and three growth media (Rousset et al., 2021), conditions that the gene was essential in all 18 tested isolates in at least one of the tested media (LB, GMM, M9). These genes were identified manually in the genome of *E. coli* str. K-12 substr. MG1655 according to the gene symbol and locus tag. The final set included 504 protein-coding essential genes (listed in S2 Table). The protein sequences of chromosome-encoded essential genes and their plasmid homologs were aligned using MAFFT (version v7.123b) (Katoh & Standley, 2013). Maximum likelihood trees were reconstructed using IQ-Tree (version 1.5.5) (Nguyen et al., 2015) with LG model. The 17 alignments and trees are available as supplementary data. Trees were rooted using the midpoint criterion. The topology of trees comprising a deep plasmid-chromosome split and putative LGT events was further tested with IQ-TREE as

following: a constrained tree topology assuming vertical evolution only (i.e., assuming a deep chromosome-plasmid split) was reconstructed; The likelihood of the resulting constrained topology was compared to the unconstrained topology (i.e., with the LGT event) using the approximately unbiased (AU) topology test. A rejection of the constrained topology thus supports the LGT inference. Codon-aware alignments were produced using PAL2NAL program (Suyama et al., 2006). For trees with a split formed by homologs on different replicons, selective strength on different branches were calculated using HyPhy-RELAX (version 2.5.15) (Murrell et al., 2012). The resulting trees were visualized using iTOL v5 (<https://itol.embl.de/>) (Letunic & Bork, 2019). Conserved syntenic blocks (i.e., gene order) were identified using CSBfinder-S (Svetlitsky et al., 2020) allowing < 4 insertions.

Bacterial strains, plasmids and culture conditions

The strain *Escherichia coli* K-12 MG1655 was used as the model organism in all experiments (DSM No. 18039, German Collection of Microorganisms and Cell Cultures, DSMZ). In addition, we used the *E. coli* strain MGM100 that encodes the *groE* operon under an inducible P_{BAD} promoter [34]. For the purpose of competition experiments, *E. coli* MG1655 equipped with a chromosomal mini-Tn7 insertion (attTn7:miniTn7(*dhfrII*) [16]) conferring resistance to trimethoprim (Tm^r) was used. The strain *E. coli* DH5 α was used during plasmid construction. All *E. coli* strains were routinely grown at 37°C in Luria Bertani (LB- Lennox) medium at 250 rpm shaking. 1% Arabinose was absent from the growth media, unless when stated otherwise. For molecular cloning and documentation purpose, the plasmid carrying strains were grown in LB supplemented with kanamycin (25 μ g per ml).

The plasmids in this study are derived from the plasmids pCON and pCON-S that were constructed previously (Wein, Hülter, et al., 2019; Wein et al., 2020). Their backbone is comprised of a pBBR1 origin (*rep* and *oriV* (Antoine & Loch, 1992)) and they carry the antibiotic resistance gene *nptII* encoding a neomycin phosphotransferase that confers resistance to kanamycin (including the natural Tn5 promoter (Beck et al., 1982)). The plasmids pGroE and pGroE-S were built by PCR amplification of pCON and pCON-S (for primers see S6 Table). The chromosomal operon *groE*, including its native promoter, was amplified from the genome of *E. coli* MG1655 by PCR using the primers *groE_for/rev* (S6 Table). Subsequently, the plasmids were assembled using the NEBuilder® protocol (New England Biolabs). The plasmid-derived GroE (WP_004201172.1, WP_004201176.1) was custom synthesized by Genscript and comprised the sequence of the strain *E. coli* 6409 (NZ_CP020056.1; Fig 2C). The sequence was introduced into pCON-S by PCR amplification (for primers see S6 Table) and NEBuilder® assembly. All plasmids were extracted using the GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) and DNA concentrations were measured using a NanoDrop (Thermo Fisher Scientific).

Fitness experiments

The relative fitness (w) (Lenski, 1991) of plasmid-carrying versus ancestral plasmid-free strain (wt) was estimated by pairwise competition experiments, with four replicates per plasmid type. All competition experiments were initiated with a 1:1 mixture of 1:100 diluted plasmid-carrying and ancestral strain (*E. coli* MG1655 Tm^r (Wein, Hülter, et al., 2019)) from overnight cultures, in a total volume of 1 ml of non-selective LB medium. The relative fitness of plasmid host strains was calculated by evaluating cell counts at time points 0 h and 24 h. The strains were distinguished through plating on non-selective (LB) and selective media (LB supplemented with trimethoprim 150 µg per ml and LB supplemented with kanamycin 25 µg per ml). The chromosomal integration (Tm^r) as well as the plasmid pCON have no measurable impact on the fitness of *E. coli* MG1655 (previously shown in (Wein, Hülter, et al., 2019)).

Evolution experiment

Evolution experiments were conducted with plasmid-carrying strains under non-selective conditions. On the onset of the experiment, the host strains were plated on selective media to ensure plasmid carriage (kanamycin, 25 µg per ml). The experiments were founded by nine replicate colonies of each plasmid type which were inoculated in LB medium at 37°C constant shaking. After overnight growth all cultures were diluted 1:100,000 and transferred into 96-deep-well plates in a total volume of 1 ml. The diluted cultures were incubated at 37°C with constant shaking. The populations were transferred every 24 h by diluting the cultures 1:100,000 and the serial transfer was repeated over a total of 23 transfers for pGroE/pGroE-S populations. The number of generations was routinely measured by evaluating the cell number through plating directly after the dilution and before the next subsequent transfer. We observed a total number of ~320 generations for the pGroE/pGroE-S populations. During the evolution experiment, the frequency of plasmids in the population was estimated from the proportion of hosts, which was determined by replica plating (LEDERBERG & LEDERBERG, 1952) every 7 transfers. Briefly, this was performed by first serially diluting the grown cultures followed by plating of ~500 cells on non-selective LB media. The number of plated cells was increased with decreasing plasmid frequency up to ~1000 cells. The plated populations were incubated for overnight growth. Colonies were counted and replicated using velvet cloth on selective media (LB supplemented with kanamycin 25 µg per ml). Colonies growing on the selective media were counted as plasmid hosts (for a detailed protocol see (Wein, Stücker, et al., 2019)).

To test for plasmid extinction and plasmid stability evolution, the evolved populations were transferred into a 96-deep-well plate containing selective media (LB supplemented with kanamycin, 25 µg per ml) and incubated for overnight growth (12 h). Thereafter, the cultures

were transferred into non-selective conditions for a follow-up evolution experiment. Plasmid host frequency was monitored via replica plating.

Plasmid loss frequency assays

The plasmid loss frequency was estimated from the frequency of plasmid free cells occurring during overnight growth in non-selective media. To determine the loss frequency, cultures were inoculated from single colonies grown on selective media to ensure plasmid carriage. After 12 h growth in 37°C (approximately 8.5 generations), the cultures were serially diluted and plated on non-selective LB media. After overnight incubation the plates were replicated. Following overnight growth, colonies grow on replicated plates were counted as plasmid carrying. The loss frequency was calculated from plasmid-free cells (not resistant) and the total number of colonies tested.

Plasmid copy number determination

Plasmid copy number (PCN) was determined using quantitative real time PCR (qPCR) (as described in (Wein, Hülter, et al., 2019)). Bacterial cells were lysed by 10 min incubation at 98°C directly followed by 10 min at -20°C. The qPCR was conducted with primers targeting the chromosome and the plasmid DNA. The chromosomal primers were complementary to the *idnT* gene of *E. coli* (q_idnT_F/R) and the plasmid primers targeted the *nptII* gene (q_nptII_F/R). The qPCR reactions were conducted in volume of 10 µl containing 1x iTaq Universal SYBR Green Supermix (Bio-Rad Laboratories), 100 nM of each primer (final concentration) and 1 µl sample. All qPCR reactions including positive and non-template controls were performed in technical replicates on a Real-Time PCR Detection System (Bio-Rad) using: 95°C for 3 min, and 40 cycles of 10 s at 95°C and 1 min at 59°C cycling conditions for all reactions. The ratio between the number of plasmid amplicons and chromosome amplicons is defined as the plasmid copy number (comparative C_T ($\Delta\Delta C_T$) method) and was calculated while considering the amplification efficiencies of both primer pairs.

Population growth dynamics

The growth dynamics of *E. coli* populations carrying different plasmids were measured by determining the optical density at 600 nm (OD₆₀₀) using a photo-spectrometer (Thermo Fisher Scientific) during growth at 37°C or 42°C for 24 h. Thereafter, the R package *growthcurver* (Sprouffske & Wagner, 2016), was used to fit a logistic growth model to the growth dynamics and estimate the growth parameters including the growth rate (*r*).

Bacterial viability assay

Bacterial viability was evaluated by staining *E. coli* cells with Propidium iodide (PI, Sigma). PI enters only compromised bacterial membranes and is therefore an indicator of membrane integrity (i.e., does not enter alive cells) and only stains dead cells. *E. coli* populations were inoculated in liquid media and the cultures were sampled after two hours (log cells) and ten hours (stationary cells) of growth (3 or 4 replicates per strain). The cells were washed in PBS (20 μ l) and incubated with PI at a final concentration of 30 μ M for 10 min in the dark at room temperature. Thereafter, 10 μ l of the mixture were transferred to an agar-coated slide and the cells were visualized with an epifluorescence microscope (Zeiss Axio Imager 2, Plan-Apochromat 63 \times /1.40 Oil DIC M27 objective). For counting purposes at least 10 images were taken per sample at random locations. The cell number of stained versus non-stained cells were evaluated using the software ImageJ.

Sequencing analysis

Population sequencing of the ancestral and evolved populations (one or two per treatment) was used to detect genetic variants occurring either on the plasmid or the host chromosomes. Prior to the sequencing, cultures were grown in antibiotics to ensure plasmid presence (see S20 Fig). Total DNA was isolated from 1 ml culture using the Wizard Genomic DNA Purification Kit (Promega). Concentration and quality of the extracted plasmid and genomic DNA was assessed using the NanoDrop (Thermo Fisher Scientific) and Qubit (Invitrogen by Life Technologies). The sample libraries for Illumina sequencing were prepared using the Nextera Flex library kit (Illumina, Inc) and sequencing was performed with paired-end reads on the MiSeq platform (Illumina, Inc).

Sequencing reads were trimmed to remove Illumina specific adaptors and low quality bases using the program Trimmomatic v.0.35 (Bolger et al., 2014) (parameters: ILLUMINACLIP:NexteraPE-PE.fa:2:30:10 CROP:250 HEADCROP:5 LEADING:20 TRAILING:20 SLIDINGWINDOW:4:20 MINLEN:36). As the reference we joined the *E. coli* MG1655 (GenBank acc. no. NC_000913.3) genome with the plasmid sequences (created in SnapGene software v.2.4 (GLS Biotech)). The sequencing reads were mapped to the reference genomes using BWA-MEM v.0.7.5a-r405 (Li & Durbin, 2009). Mapping statistics were retrieved using BAMStats v.1.25 (<https://sourceforge.net/projects/bamstats/files/>). Subsequent indexing and local realignment of sequencing reads were performed using PICARD tools, SAMtools v.0.1.19 (Li et al., 2009) and GATK v.3.6 (McKenna et al., 2010) retaining only paired mapped reads with a minimum mapping quality of 20. SNPs were called using LoFreq v.2.1.2 (Wilm et al., 2012).

Statistical analysis

All statistical tests and data analysis were performed in R version 3.5.1.

Data accessibility

The data are provided in electronic supplementary material. Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.5677445.v1>

Author Contributions

Conceptualization: Tanita Wein, Katrin Hammerschmidt, Tal Dagan.

Data curation: Yiqing Wang.

Formal analysis: Tanita Wein, Yiqing Wang, Tal Dagan.

Funding acquisition: Tanita Wein, Tal Dagan.

Investigation: Tanita Wein, Yiqing Wang, Myriam Barz, Fenna T. Stücker, Tal Dagan.

Supervision: Tanita Wein, Tal Dagan.

Validation: Tanita Wein, Yiqing Wang, Myriam Barz, Fenna T. Stücker.

Visualization: Tanita Wein, Yiqing Wang, Tal Dagan.

Writing – original draft: Tanita Wein, Yiqing Wang, Tal Dagan.

Writing – review & editing: Myriam Barz, Fenna T. Stücker, Katrin Hammerschmidt.

Acknowledgments

We thank Erik Brinks for his assistance with the genome sequencing. We thank Ana Garoña and Nils Hülter for critical comments on the manuscript.

References

- Acar Kirit, H., Lagator, M., & Bollback, J. P. (2020). Experimental determination of evolutionary barriers to horizontal gene transfer. *BMC Microbiology*, 20(1), 326–13. <https://doi.org/10.1186/S12866-020-01983-5>
- Alcock, B. P., Raphenya, A. R., Lau, T. T. Y., Tsang, K. K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen, A. L. V., Cheng, A. A., Liu, S., Min, S. Y., Miroshnichenko, A., Tran, H. K., Werfalli, R. E., Nasir, J. A., Oloni, M., Speicher, D. J., Florescu, A., Singh, B., ... McArthur, A. G. (2020). CARD 2020: Antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Research*, 48(D1), D517–D525. <https://doi.org/10.1093/NAR/GKZ935>
- Anda, M., Ohtsubo, Y., Okubo, T., Sugawara, M., Nagata, Y., Tsuda, M., Minamisawa, K., & Mitsui, H. (2015). Bacterial clade with the ribosomal RNA operon on a small plasmid rather than the chromosome. *Proceedings of the National Academy of Sciences of the*

- United States of America*, 112(46), 14343–14347.
<https://doi.org/10.1073/PNAS.1514326112>
- Andersson, D. I., Jerlström-Hultqvist, J., & Näsvall, J. (2015). Evolution of new functions de novo and from preexisting genes. *Cold Spring Harbor Perspectives in Medicine*, 5(7), a017996–19. <https://doi.org/10.1101/CSHPERSPECT.A017996>
- Antoine, R., & Loch, C. (1992). Isolation and molecular characterization of a novel broad-host-range plasmid from *Bordetella bronchiseptica* with sequence similarities to plasmids from Gram-positive organisms. *Molecular Microbiology*, 6(13), 1785–1799. <https://doi.org/10.1111/J.1365-2958.1992.TB01351.X>
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., & Mori, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: The Keio collection. *Molecular Systems Biology*, 2, 2460–11. <https://doi.org/10.1038/MSB4100050>
- Baltrus, D. A. (2013). Exploring the costs of horizontal gene transfer. *Trends in Ecology and Evolution*, 28(8), 489–495. <https://doi.org/10.1016/J.TREE.2013.04.002>
- Beck, E., Ludwig, G., Auerswald, E. A., Reiss, B., & Schaller, H. (1982). Nucleotide sequence and exact localization of the neomycin phosphotransferase gene from transposon Tn5. *Gene*, 19(3), 327–336. [https://doi.org/10.1016/0378-1119\(82\)90023-3](https://doi.org/10.1016/0378-1119(82)90023-3)
- Bergstrom, C. T., Lipsitch, M., & Levin, B. R. (2000). Natural selection, infectious transfer and the existence conditions for bacterial plasmids. *Genetics*, 155(4), 1505–1519. <https://doi.org/10.1093/GENETICS/155.4.1505>
- Bittner, A. N., Foltz, A., & Oke, V. (2007). Only one of five *groEL* genes is required for viability and successful symbiosis in *Sinorhizobium meliloti*. *Journal of Bacteriology*, 189(5), 1884–1889. <https://doi.org/10.1128/JB.01542-06>
- Blau, K., Bettermann, A., Jechalke, S., Fornfeldt, E., Vanrobaeys, Y., Stalder, T., Top, E. M., & Smalla, K. (2018). The transferable resistome of produce. *MBio*, 9(6), 1–15. <https://doi.org/10.1128/MBIO.01300-18>
- Bogumil, D., & Dagan, T. (2012). Cumulative impact of chaperone-mediated folding on genome evolution. *Biochemistry*, 51(50), 9941–9953. <https://doi.org/10.1021/BI3013643>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bouma, J. E., & Lenski, R. E. (1988). Evolution of a bacteria/plasmid association. *Nature*, 335(6188), 351–352. <https://doi.org/10.1038/335351A0>
- Boyd, B. M., Allen, J. M., Nguyen, N. P., Vachaspati, P., Quicksall, Z. S., Warnow, T., Mugisha, L., Johnson, K. P., & Reed, D. L. (2017). Primates, lice and bacteria:

- Speciation and genome evolution in the symbionts of hominid lice. *Molecular Biology and Evolution*, 34(7), 1743–1757. <https://doi.org/10.1093/MOLBEV/MSX117>
- Bratlie, M. S., Johansen, J., Sherman, B. T., Huang, D. W., Lempicki, R. A., & Drabløs, F. (2010). Gene duplications in prokaryotes can be associated with environmental adaptation. *BMC Genomics*, 11(1), 588. <https://doi.org/10.1186/1471-2164-11-588>
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, 10, 421-9. <https://doi.org/10.1186/1471-2105-10-421>
- Chase, J. W., Merrill, B. M., & Williams, K. R. (1983). F sex factor encodes a single-stranded DNA binding protein (SSB) with extensive sequence homology to *Escherichia coli* SSB. *Proceedings of the National Academy of Sciences of the United States of America*, 80(181), 5480–5484. <https://doi.org/10.1073/PNAS.80.18.5480>
- Clark, G. W., & Tillier, E. R. M. (2010). Loss and gain of GroEL in the Mollicutes. *Biochemistry and Cell Biology*, 88(2), 185–194. <https://doi.org/10.1139/O09-157>
- Cohen, O., Gophna, U., & Pupko, T. (2011). The complexity hypothesis revisited: Connectivity rather than function constitutes a barrier to horizontal gene transfer. *Molecular Biology and Evolution*, 28(4), 1481–1489. <https://doi.org/10.1093/MOLBEV/MSQ333>
- diCenzo, G., Milunovic, B., Cheng, J., & Finan, T. M. (2013). The tRNA^{arg} gene and *engA* are essential genes on the 1.7-Mb pSymB megaplasmid of *Sinorhizobium meliloti* and were translocated together from the chromosome in an ancestral strain. *Journal of Bacteriology*, 195(2), 202–212. <https://doi.org/10.1128/JB.01758-12>
- Dziewit, L., Pyzik, A., Szuplewska, M., Matlakowska, R., Mielnicki, S., Wibberg, D., Schlüter, A., Pühler, A., & Bartosik, D. (2015). Diversity and role of plasmids in adaptation of bacteria inhabiting the Lubin copper mine in Poland, an environment rich in heavy metals. *Frontiers in Microbiology*, 6, 152. <https://doi.org/10.3389/FMICB.2015.00152>
- Enright, A. J., Van Dongen, S., & Ouzounis, C. A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research*, 30(7), 1575–1584. <https://doi.org/10.1093/NAR/30.7.1575>
- Fares, M. A., Ruiz-González, M. X., Moya, A., Elena, S. F., & Barrio, E. (2002). Endosymbiotic bacteria: GroEL buffers against deleterious mutations. *Nature*, 417(6887), 398. <https://doi.org/10.1038/417398A>
- Fayet, O., Ziegelhoffer, T., & Georgopoulos, C. (1989). The *groES* and *groEL* heat shock gene products of *Escherichia coli* are essential for bacterial growth at all temperatures. *Journal of Bacteriology*, 171(3), 1379–1385. <https://doi.org/10.1128/JB.171.3.1379-1385.1989>

- Galibert, F., Finan, T. M., Long, S. R., Pühler, A., Abola, P., Ampe, F., Barloy-Hubler, F., Barnett, M. J., Becker, A., Boistard, P., Bothe, G., Boutry, M., Bowser, L., Buhrmester, J., Cadieu, E., Capela, D., Chain, P., Cowie, A., Davis, R. W., ... Batut, J. (2001). The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science*, 293(5530), 668–672. <https://doi.org/10.1126/SCIENCE.1060966>
- Garcillán-Barcia, M. P., Alvarado, A., & De la Cruz, F. (2011). Identification of bacterial plasmids based on mobility and plasmid population biology. *FEMS Microbiology Reviews*, 35(5), 936–956. <https://doi.org/10.1111/J.1574-6976.2011.00291.X>
- Gil, R., Sabater-Muñoz, B., Perez-Brocal, V., Silva, F. J., & Latorre, A. (2006). Plasmids in the aphid endosymbiont *Buchnera aphidicola* with the smallest genomes. A puzzling evolutionary story. *Gene*, 370(1–2), 17–25. <https://doi.org/10.1016/J.GENE.2005.10.043>
- Goodall, E. C. A., Robinson, A., Johnston, I. G., Jabbari, S., Turner, K. A., Cunningham, A. F., Lund, P. A., Cole, J. A., & Henderson, I. R. (2018). The essential genome of *Escherichia coli* K-12. *MBio*, 9(1), e02096-17. <https://doi.org/10.1128/MBIO.02096-17>
- Gullberg, E., Albrecht, L. M., Karlsson, C., Sandegren, L., & Andersson, D. I. (2014). Selection of a multidrug resistance plasmid by sublethal levels of antibiotics and heavy metals. *MBio*, 5(5), e01918-14. <https://doi.org/10.1128/MBIO.01918-14>
- Gunasekera, T. S., Csonka, L. N., & Paliy, O. (2008). Genome-wide transcriptional responses of *Escherichia coli* K-12 to continuous osmotic and heat stresses. *Journal of Bacteriology*, 190(10), 3712–3720. <https://doi.org/10.1128/JB.01990-07>
- Hao, W., & Brian Golding, G. (2006). The fate of laterally transferred genes: Life in the fast lane to adaptation or death. *Genome Research*, 16(5), 636–643. <https://doi.org/10.1101/GR.4746406>
- Harrison, E., & Brockhurst, M. A. (2012). Plasmid-mediated horizontal gene transfer is a coevolutionary process. *Trends in Microbiology*, 20(6), 262–267. <https://doi.org/10.1016/J.TIM.2012.04.003>
- Harrison, P. W., Lower, R. P. J., Kim, N. K. D., & Young, J. P. W. (2010). Introducing the bacterial ‘chromid’: Not a chromosome, not a plasmid. *Trends in Microbiology*, 18(4), 141–148. <https://doi.org/10.1016/J.TIM.2009.12.010>
- Heuer, H., Fox, R. E., & Top, E. M. (2007). Frequent conjugative transfer accelerates adaptation of a broad-host-range plasmid to an unfavorable *Pseudomonas putida* host. *FEMS Microbiology Ecology*, 59(3), 738–748. <https://doi.org/10.1111/J.1574-6941.2006.00223.X>
- Heuer, H., Schmitt, H., & Smalla, K. (2011). Antibiotic resistance gene spread due to manure application on agricultural fields. *Current Opinion in Microbiology*, 14(3), 236–243. <https://doi.org/10.1016/J.MIB.2011.04.009>

- Hülter, N., Ilhan, J., Wein, T., Kadibalban, A. S., Hammerschmidt, K., & Dagan, T. (2017). An evolutionary perspective on plasmid lifestyle modes. *Current Opinion in Microbiology*, 38, 74–80. <https://doi.org/10.1016/J.MIB.2017.05.001>
- Jain, R., Rivera, M. C., & Lake, J. A. (1999). Horizontal gene transfer among genomes: The complexity hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, 96(7), 3801–3806. <https://doi.org/10.1073/PNAS.96.7.3801>
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772–780. <https://doi.org/10.1093/MOLBEV/MST010>
- Lau, B. T. C., Malkus, P., & Paulsson, J. (2013). New quantitative methods for measuring plasmid loss rates reveal unexpected stability. *Plasmid*, 70(3), 353–361. <https://doi.org/10.1016/J.PLASMID.2013.07.007>
- LEDERBERG, J., & LEDERBERG, E. M. (1952). Replica plating and indirect selection of bacterial mutants. *Journal of Bacteriology*, 63(3), 399–406. <https://doi.org/10.1128/JB.63.3.399-406.1952>
- Lenski, R. E. (1991). Quantifying fitness and gene stability in microorganisms. *Biotechnology (Reading, Mass.)*, 15, 173–192. <https://doi.org/10.1016/B978-0-409-90199-3.50015-2>
- Lercher, M. J., & Pál, C. (2008). Integration of horizontally transferred genes into regulatory interaction networks takes many million years. *Molecular Biology and Evolution*, 25(3), 559–567. <https://doi.org/10.1093/MOLBEV/MSM283>
- Letunic, I., & Bork, P. (2019). Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Research*, 47(W1), W256–W259. <https://doi.org/10.1093/NAR/GKZ239>
- Levin, B. R., & Stewart, F. M. (1980). The population biology of bacterial plasmids: A priori conditions for the existence of mobilizable nonconjugative factors. *Genetics*, 94(2), 425–443. <https://doi.org/10.1093/GENETICS/94.2.425>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/ map format and SAMtools. *Bioinformatics (Oxford, England)*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Loftie-Eaton, W., Bashford, K., Quinn, H., Dong, K., Millstein, J., Hunter, S., Thomason, M. K., Merrih, H., Ponciano, J. M., & Top, E. M. (2017). Compensatory mutations improve general permissiveness to antibiotic resistance plasmids. *Nature Ecology and Evolution*, 1(9), 1354–1363. <https://doi.org/10.1038/S41559-017-0243-2>

- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303. <https://doi.org/10.1101/gr.107524.110>
- McLennan, N., & Masters, M. (1998). GroE is vital for cell-wall synthesis. *Nature*, 392(6672), 139. <https://doi.org/10.1038/32317>
- Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., & Kosakovsky Pond, S. L. (2012). Detecting individual sites subject to episodic diversifying selection. *PLoS Genetics*, 8(7), e1002764. <https://doi.org/10.1371/JOURNAL.PGEN.1002764>
- Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32(1), 268–274. <https://doi.org/10.1093/MOLBEV/MSU300>
- Nicoloff, H., Hjort, K., Levin, B. R., & Andersson, D. I. (2019). The high prevalence of antibiotic heteroresistance in pathogenic bacteria is mainly caused by gene amplification. *Nature Microbiology*, 4(3), 504–514. <https://doi.org/10.1038/S41564-018-0342-0>
- Porter, R. D., & Black, S. (1991). The single-stranded-DNA-binding protein encoded by the *Escherichia coli* F factor can complement a deletion of the chromosomal *ssb* gene. *Journal of Bacteriology*, 173(8), 2720–2723. <https://doi.org/10.1128/JB.173.8.2720-2723.1991>
- Rancati, G., Moffat, J., Typas, A., & Pavelka, N. (2018). Emerging and evolving concepts in gene essentiality. *Nature Reviews Genetics*, 19(1), 34–49. <https://doi.org/10.1038/NRG.2017.74>
- Richmond, C. S., Glasner, J. D., Mau, R., Jin, H., & Blattner, F. R. (1999). Genome-wide expression profiling in *Escherichia coli* K-12. *Nucleic Acids Research*, 27(19), 3821–3835. <https://doi.org/10.1093/NAR/27.19.3821>
- Rodriguez-Beltran, J., Hernandez-Beltran, J. C. R., Delafuente, J., Escudero, J. A., Fuentes-Hernandez, A., MacLean, R. C., Peña-Miller, R., & San Millan, A. (2018). Multicopy plasmids allow bacteria to escape from fitness trade-offs during evolutionary innovation. *Nature Ecology and Evolution*, 2(5), 873–881. <https://doi.org/10.1038/S41559-018-0529-Z>
- Rousset, F., Cabezas-Caballero, J., Piastra-Facon, F., Fernández-Rodríguez, J., Clermont, O., Denamur, E., Rocha, E. P. C., & Bikard, D. (2021). The impact of genetic diversity on gene essentiality within the *Escherichia coli* species. *Nature Microbiology*, 6(3), 301–312. <https://doi.org/10.1038/s41564-020-00839-y>
- Sabater-Muñoz, B., Prats-Escriche, M., Montagud-Martínez, R., López-Cerdán, A., Toft, C., Aguilar-Rodríguez, J., Wagner, A., & Fares, M. A. (2015). Fitness trade-offs determine

- the role of the molecular chaperonin GroEL in buffering mutations. *Molecular Biology and Evolution*, 32(10), 2681–2693. <https://doi.org/10.1093/MOLBEV/MSV144>
- Sekizuka, T., Matsui, M., Yamane, K., Takeuchi, F., Ohnishi, M., Hishinuma, A., Arakawa, Y., & Kuroda, M. (2011). Complete sequencing of the *bla*_{NDM-1}-positive IncA/C plasmid from *Escherichia coli* ST38 isolate suggests a possible origin from plant pathogens. *PLoS ONE*, 6(9), e25334. <https://doi.org/10.1371/JOURNAL.PONE.0025334>
- Smillie, C., Garcillán-Barcia, M. P., Francia, M. V., Rocha, E. P. C., & de la Cruz, F. (2010). Mobility of Plasmids. *Microbiology and Molecular Biology Reviews*, 74(3), 434–452. <https://doi.org/10.1128/membr.00020-10>
- Sorek, R., Zhu, Y., Creevey, C. J., Francino, M. P., Bork, P., & Rubin, E. M. (2007). Genome-wide experimental determination of barriers to horizontal gene transfer. *Science*, 318(5855), 1449–1452. <https://doi.org/10.1126/SCIENCE.1147112>
- Sota, M., Yano, H., M Hughes, J., Daughdrill, G. W., Abdo, Z., Forney, L. J., & Top, E. M. (2010). Shifts in the host range of a promiscuous plasmid through parallel evolution of its replication initiation protein. *ISME Journal*, 4(12), 1568–1580. <https://doi.org/10.1038/ISMEJ.2010.72>
- Sprouffske, K., & Wagner, A. (2016). Growthcurver: An R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinformatics*, 17(1), 172. <https://doi.org/10.1186/s12859-016-1016-7>
- Stalder, T., Rogers, L. M., Renfrow, C., Yano, H., Smith, Z., & Top, E. M. (2017). Emerging patterns of plasmid-host coevolution that stabilize antibiotic resistance. *Scientific Reports*, 7(1), 4853. <https://doi.org/10.1038/S41598-017-04662-0>
- Suyama, M., Torrents, D., & Bork, P. (2006). PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, 34(Web Server issue), W609-12. <https://doi.org/10.1093/NAR/GKL315>
- Svetlitsky, D., Dagan, T., & Ziv-Ukelson, M. (2020). Discovery of multi-operon colinear syntenic blocks in microbial genomes. *Bioinformatics*, 36, I21–I29. <https://doi.org/10.1093/BIOINFORMATICS/BTAA503>
- Tazzyman, S. J., & Bonhoeffer, S. (2015). Why there are no essential genes on plasmids. *Molecular Biology and Evolution*, 32(12), 3079–3088. <https://doi.org/10.1093/MOLBEV/MSU293>
- Wallington, E. J., & Lund, P. A. (1994). *Rhizobium leguminosarum* contains multiple chaperonin (cpn60) genes. *Microbiology*, 140(1), 113–122. <https://doi.org/10.1099/13500872-140-1-113>
- Wein, T., Hülter, N. F., Mizrahi, I., & Dagan, T. (2019). Emergence of plasmid stability under non-selective conditions maintains antibiotic resistance. *Nature Communications*, 10(1), 2595. <https://doi.org/10.1038/S41467-019-10600-7>

- Wein, T., Stücker, F. T., Hülter, N. F., & Dagan, T. (2019). Quantification of Plasmid-Mediated Antibiotic Resistance in an Experimental Evolution Approach. *Journal of Visualized Experiments*, 154. <https://doi.org/10.3791/60749>
- Wein, T., Wang, Y., Hülter, N. F., Hammerschmidt, K., & Dagan, T. (2020). Antibiotics interfere with the evolution of plasmid stability. *Current Biology*, 30(19), 3841-3847.e4. <https://doi.org/10.1016/J.CUB.2020.07.019>
- Wheatley, R., Diaz Caballero, J., Kapel, N., de Winter, F. H. R., Jangir, P., Quinn, A., del Barrio-Tofiño, E., López-Causapé, C., Hedge, J., Torrens, G., Van der Schalk, T., Xavier, B. B., Fernández-Cuenca, F., Arenzana, A., Recanatini, C., Timbermont, L., Sifakis, F., Ruzin, A., Ali, O., ... MacLean, C. (2021). Rapid evolution and host immunity drive the rise and fall of carbapenem resistance during an acute *Pseudomonas aeruginosa* infection. *Nature Communications*, 12(1), 2460. <https://doi.org/10.1038/S41467-021-22814-9>
- Wilm, A., Aw, P. P. K., Bertrand, D., Yeo, G. H. T., Ong, S. H., Wong, C. H., Khor, C. C., Petric, R., Hibberd, M. L., & Nagarajan, N. (2012). LoFreq: A sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Research*, 40(22), 11189–11201. <https://doi.org/10.1093/nar/gks918>
- Zhang, X., Feng, Y., Zhou, W., McNally, A., & Zong, Z. (2018). Cryptic transmission of ST405 *Escherichia coli* carrying bla_{NDM-4} in hospital. *Scientific Reports*, 8(1), 390. <https://doi.org/10.1038/S41598-017-18910-W>

Supporting information

S1 Fig. An empirical cumulative distribution function (CDF) describing the number of isolates per plasmid protein family in which a chromosomal homolog to a plasmid protein family was found. An inset plot (in red dashed line) shows an enlargement of the top left corner. The distribution shows that 81% plasmid protein families have no chromosomal homolog in the same isolate. The remaining 19% plasmid protein families have a chromosomal homolog in at least one isolate. <https://doi.org/10.1371/journal.pgen.1009656.s001>

S2 Fig. Phylogeny of ATP-dependent metalloprotease FtsH homologs in *Escherichia* strains. Isolate names are colored according to gene location with red for plasmids and blue for chromosomes. <https://doi.org/10.1371/journal.pgen.1009656.s002>

S3 Fig. Phylogeny of DNA replication protein DnaC homologs in *Escherichia* strains.

Isolate names are colored according to gene location with red for plasmids and blue for chromosomes. <https://doi.org/10.1371/journal.pgen.1009656.s003>

S4 Fig. Phylogeny of methionine—tRNA ligase MetG homologs in *Escherichia* strains.

Isolate names are colored according to gene location with red for plasmids and blue for chromosomes. MetG in *Escherichia coli* isolate 2014C-3307 is found only on a plasmid (i.e., it has no chromosomal homolog). <https://doi.org/10.1371/journal.pgen.1009656.s004>

S5 Fig. Phylogeny of glutamine—tRNA ligase GlnS homologs in *Escherichia* strains.

Isolate names are colored according to gene location with red for plasmids and blue for chromosomes. <https://doi.org/10.1371/journal.pgen.1009656.s005>

S6 Fig. Phylogeny of gamma-glutamylphosphate reductase ProA homologs in

***Escherichia* strains.** Isolate names are colored according to gene location with red for plasmids and blue for chromosomes. <https://doi.org/10.1371/journal.pgen.1009656.s006>

S7 Fig. Phylogeny of gamma-glutamate kinase ProB homologs in *Escherichia* strains.

Isolate names are colored according to gene location with red for plasmids and blue for chromosomes. <https://doi.org/10.1371/journal.pgen.1009656.s007>

S8 Fig. Phylogeny of aromatic acid decarboxylase UbiX homologs in *Escherichia*

strains. Isolate names are colored according to gene location with red for plasmids and blue for chromosomes. <https://doi.org/10.1371/journal.pgen.1009656.s008>

S9 Fig. Phylogeny of ribonucleotide-diphosphate reductase subunit beta NrdB

homologs in *Escherichia* strains. Isolate names are colored according to gene location with red for plasmids and blue for chromosomes.

<https://doi.org/10.1371/journal.pgen.1009656.s009>

S10 Fig. Phylogeny of ribonucleoside-diphosphate reductase subunit alpha NrdA

homologs in *Escherichia* strains. Isolate names are colored according to gene location with red for plasmids and blue for chromosomes.

<https://doi.org/10.1371/journal.pgen.1009656.s010>

S11 Fig. Phylogeny of peptide deformylase Def homologs in *Escherichia* strains.

Isolate names are colored according to gene location with red for plasmids and blue for chromosomes. <https://doi.org/10.1371/journal.pgen.1009656.s011>

S12 Fig. Phylogeny of ketol-acid reductoisomerase LivC homologs in *Escherichia*

strains. Isolate names are colored according to gene location with red for plasmids and blue for chromosomes. <https://doi.org/10.1371/journal.pgen.1009656.s012>

S13 Fig. Phylogeny of lipoprotein-releasing ABC transporter permease subunit LolC homologs in *Escherichia* strains.

Isolate names are colored according to gene location with red for plasmids and blue for chromosomes.

<https://doi.org/10.1371/journal.pgen.1009656.s013>

S14 Fig. Phylogeny of D-arabinose 5-phosphate isomerase KdsD homologs in

***Escherichia* strains.** Isolate names are colored according to gene location with red for plasmids and blue for chromosomes. <https://doi.org/10.1371/journal.pgen.1009656.s014>

S15 Fig. Branch specific parameters of relaxation of selection inferred for GroEL

phylogeny using HyPhy-RELAX. Branches are colored according to the selection intensity parameter k , ranging between brown for $k < 1$ (relaxed selection) and green for $k > 1$

(intensified selection). Note that branch length information is excluded in this figure. Isolate

names are colored according to gene location with red for plasmids and blue for

chromosomes. <https://doi.org/10.1371/journal.pgen.1009656.s015>

S16 Fig. Branch specific relaxation parameters are inferred for NrdA phylogeny using

HyPhy-RELAX. Branches are colored according to the selection intensity parameter k ,

ranging between brown for $k < 1$ (relaxed selection) and green for $k > 1$ (intensified selection).

Note that branch length information is excluded in this figure. Isolate names are colored

according to gene location with red for plasmids and blue for chromosomes.

<https://doi.org/10.1371/journal.pgen.1009656.s016>

S17 Fig. Phylogeny of replicative DNA helicase DnaB homologs in *Escherichia* strains.

Isolate names are colored according to gene location with red for plasmids and blue for

chromosomes. <https://doi.org/10.1371/journal.pgen.1009656.s017>

S18 Fig. Relative Fitness of *E. coli* MGM100 in comparison to *E. coli* MG1655 ($n = 4$).

<https://doi.org/10.1371/journal.pgen.1009656.s018>

S19 Fig. Plasmid loss of pGroE and pGroE-S in *E. coli* MGM100 after overnight growth in the presence of 1% arabinose (n = 6).

<https://doi.org/10.1371/journal.pgen.1009656.s019>

S20 Fig. Quantification of the number of dead cells along the growth phase of *E. coli* MG1655 and MGM100 populations. The number of dead cells was evaluated in the log phase and in the stationary growth phase (n = 4).

<https://doi.org/10.1371/journal.pgen.1009656.s020>

S21 Fig. Plasmid frequency following evolution experiment. Plasmid carrying populations were exposed to a transfer in antibiotics (kanamycin 25 µg/ml).

<https://doi.org/10.1371/journal.pgen.1009656.s021>

S22 Fig. Plasmid loss frequency of evolved plasmids pGroE-S and pGroE in the host strain *E. coli* MG1655. The plasmid loss was measured after overnight incubation (n = 6).

<https://doi.org/10.1371/journal.pgen.1009656.s022>

Chapter II - Gene sharing among plasmids and chromosomes reveals barriers for antibiotic resistance gene transfer

Yiqing Wang¹, Aditi Batra², Hinrich Schulenburg² and Tal Dagan^{1*}

¹Institute of General Microbiology, and ²Zoological institute, Kiel University, Kiel, Germany

* Corresponding author

Manuscript published in ***PHILOS T R SOC B*** 2021. <https://doi.org/10.1098/rstb.2020.0467>

Abstract

The emergence of antibiotic resistant bacteria is a major threat to modern medicine. Rapid adaptation to antibiotics is often mediated by the acquisition of plasmids carrying antibiotic resistance (ABR) genes. Nonetheless, the determinants of plasmid-mediated ABR gene transfer remain debated. Here, we show that the propensity of ABR gene transfer via plasmids is higher for accessory chromosomal ABR genes in comparison with core chromosomal ABR genes, regardless of the resistance mechanism. Analysing the pattern of ABR gene occurrence in the genomes of 2635 Enterobacteriaceae isolates, we find that 33% of the 416 ABR genes are shared between chromosomes and plasmids. Phylogenetic reconstruction of ABR genes occurring on both plasmids and chromosomes supports their evolution by lateral gene transfer. Furthermore, accessory ABR genes (encoded in less than 10% of the chromosomes) occur more abundantly in plasmids in comparison with core ABR genes (encoded in greater than or equal to 90% of the chromosomes). The pattern of ABR gene occurrence in plasmids and chromosomes is similar to that in the total *Escherichia* genome. Our results thus indicate that the previously recognized barriers for gene acquisition by lateral gene transfer apply also to ABR genes. We propose that the functional complexity of the underlying ABR mechanism is an important determinant of ABR gene transferability.

Introduction

Bacteria harbour vast potential to rapidly adapt to different selective environments, including environments defined by antibiotics. Notably, antibiotic resistance (ABR) predates the clinical use of antibiotics: ABR genes have been reported from permafrost (Dcosta et al., 2011) and ancient human remains (Warinner et al., 2014). Today, the ability of bacteria to adapt to antibiotics forms the core of the ABR crisis (Laxminarayan et al., 2013). Drug therapy is aimed at killing bacteria by targeting key components and processes that are crucial for their reproduction; the repertoire of genetic variants giving rise to ABR therefore includes many core genes (Baquero & Levin, 2021). For example, bacterial efflux pumps, which function in resistance to diverse antibiotics, have multiple roles in bacterial physiology in addition to removal of antibiotics from the cell, including the extrusion of heavy metals and cell signalling molecules (Blanco et al., 2016). Indeed, genes encoding efflux pumps and their regulation are typically part of bacterial core genomes, i.e. they are nearly universally present in bacterial genomes (Gallagher et al., 2017). By contrast, other resistance mechanisms are encoded by genes in the accessory genome and can be found also on mobile genetic elements (e.g. genes involved in target replacement or target protection) (Rodríguez-Beltrán et al., 2020). While the emergence of ABR under long-term drug therapy is common via *de novo* mutations, lateral transfer of ABR genes by mobile genetic elements has the potential to disseminate resistance

to various antibiotics within microbial communities in clinical settings, agriculture and also the environment (Heuer et al., 2011; Andersson et al., 2021).

The acquisition of ABR genes can be mediated by three main transfer mechanisms: transformation, transduction and conjugation. Bacterial organisms that are naturally competent may acquire ABR genes via natural transformation of mobile genetic elements carrying ABR genes, including transposons, insertion sequence (IS) elements and integrons (e.g. (Domingues et al., 2012)). Transduction can cause the transfer of ABR genes encoded in the genome of lysogenic phages (e.g. (Mazaheri Nezhad Fard et al., 2011; Popa et al., 2017)) or the transfer of ABR gene-containing plasmids by phages (e.g. (Varga et al., 2012)). Notwithstanding, plasmid-mediated ABR gene transfer occurs most commonly via conjugation, which facilitates the transfer of conjugative and mobilizable plasmids (e.g. (Blau et al., 2018)). The dissemination of ABR genes by plasmids is often associated with the evolution of plasmid-encoded integrons that may comprise multiple ABR genes (Che et al., 2021). The type of vehicle that mediates the ABR gene transfer may have implications for the phylogenetic range of the transfer event. Phage-mediated gene transfer is often restricted to closely related donors and recipients owing to phage host specificity (Popa et al., 2017). Similarly, the apparent taxonomic pattern in networks of plasmid sequence similarity (Redondo-Salvo et al., 2020) suggests that plasmid host range constrains the range of plasmid-mediated gene transfer to closely related donor and recipients (with the exception of broad host range plasmids, e.g. (Jørgensen et al., 2014)).

The prevalence of gene transfer may furthermore depend on genetic and functional characteristics of the transferred gene. Genes that encode proteins whose function depends on interaction with multiple proteins, e.g. within the same complex or cellular process, are rarely transferred, likely because their acquisition leads to interference with the component sociometry (Jain et al., 1999a; Cohen et al., 2011a). An increase in gene copy number following lateral gene transfer may often lead to disadvantageous dose effect, which has been shown to pose a barrier to lateral gene transfer (Sorek et al., 2007; Acar Kirit et al., 2020). The acquisition of novel laterally transferred genes may have deleterious effects on the recipient owing to diverse reasons ranging from the consequences of foreign DNA integration into the genome to proper expression and function of the protein product in the cell (e.g. (Socha et al., 2019); reviewed in (Baltrus, 2013)). A potential disruption of chromosomal architecture may be minimized if genes are transferred by and expressed from extra-chromosomal genetic elements like plasmids (i.e. assuming non-integrative plasmids). Nonetheless, ABR gene expression from a plasmid locus may be accompanied by fitness cost to the host owing to various other reasons. For example, ABR mechanisms that rely on drug inactivation may require a lower energetic investment (i.e. in ATP) of the cell in comparison with efflux pumps

[24]. Such trade-offs in the adaptation of ABR plasmids in newly colonized hosts are expected to have consequences for the evolution of plasmid ABR gene content.

Plasmid gene content is furthermore expected to be subject to purifying selection owing to genetic conflicts between plasmid and chromosomally encoded loci that affect the same phenotype (Werren, 2011). Indeed, a recent study suggests that the evolution of plasmid ABR gene content depends on the level of genetic conflicts with chromosomally encoded genes, which is strongest for ABR genes encoding efflux pumps and weakest for ABR genes encoding antibiotic target inactivation mechanisms (Rodríguez-Beltrán et al., 2020). Considering the association between ABR gene classification into core or accessory genome and the ABR mechanism, we hypothesize that the prevalence of ABR genes on plasmids depends on the ABR gene occurrence in bacterial chromosomes. Here, we examine ABR gene occurrence in bacterial plasmids and chromosomes using a large-scale analysis of three Enterobacteriaceae genera: *Escherichia*, *Salmonella* and *Klebsiella*, for which a large number of genomes are publicly available. We further assess whether ABR gene occurrence on plasmids and chromosomes is associated with the resistance mechanism and inferred ABR gene transfer events using phylogenetics and the temporal pattern of gene occurrence in plasmids and chromosomes.

Results

Plasmid ABR gene content is associated with gene prevalence on chromosomes

The survey for genes that putatively confer ABR revealed 2810 (38%) plasmids that encode an ABR gene homologue (termed here ABR plasmids). The ABR plasmids typically encode multiple ABR genes, with the median number of ABR genes per plasmid ranging between four in *Escherichia* and six in *Salmonella*. The proportion of ABR plasmids was similar among the three genera: 34% (1131) in *Escherichia*, 42% (1256) in *Klebsiella* and 39% (423) in *Salmonella*. The genomes of all isolates in the three genera had at least one homologue to an ABR gene encoded in the chromosome.

The ABR gene homologues in *Escherichia* plasmids and chromosomes fall within three classes: (1) ABR genes exclusively found on plasmids, (2) ABR genes found exclusively on chromosomes, and (3) ABR genes found on both plasmids and chromosomes (figure 1 a). The pattern of gene occurrence on plasmids and chromosomes further reveals two main groups: ABR genes encoded on plasmids and shared with a small proportion of the chromosomes (less than or equal to 10%) - these genes are considered to be part of the accessory genome - and ABR genes that are encoded on plasmids and are shared with greater than or equal to 90% of the chromosomes - these genes are considered to be part of the core genome (see yellow bars in figure 1a). A similar pattern of ABR gene occurrence on plasmids and chromosomes is observed in the genomes of *Klebsiella* and *Salmonella* isolates (electronic

supplementary material, figures S1 and S2). To compare the pattern of ABR gene occurrence with the total *Escherichia* genes, we repeated the analysis using a dataset of protein families in *Escherichia* isolates. The results show that the ABR gene occurrence in plasmids and chromosomes bears similarity to the distribution of the total gene families in *Escherichia* genomes (figure 1b): in both datasets homologues of core chromosomal genes found in greater than or equal to 90% of the chromosomes are less abundant in plasmids in comparison with homologues of accessory chromosomal genes found in fewer than or equal to 10% of the chromosomes. The similar pattern of shared genes among plasmids and chromosomes observed for the ABR genes and the total genome indicates that the association between ABR gene prevalence in plasmids and chromosomes is general within *Escherichia* genomes rather than a unique property of ABR genes.

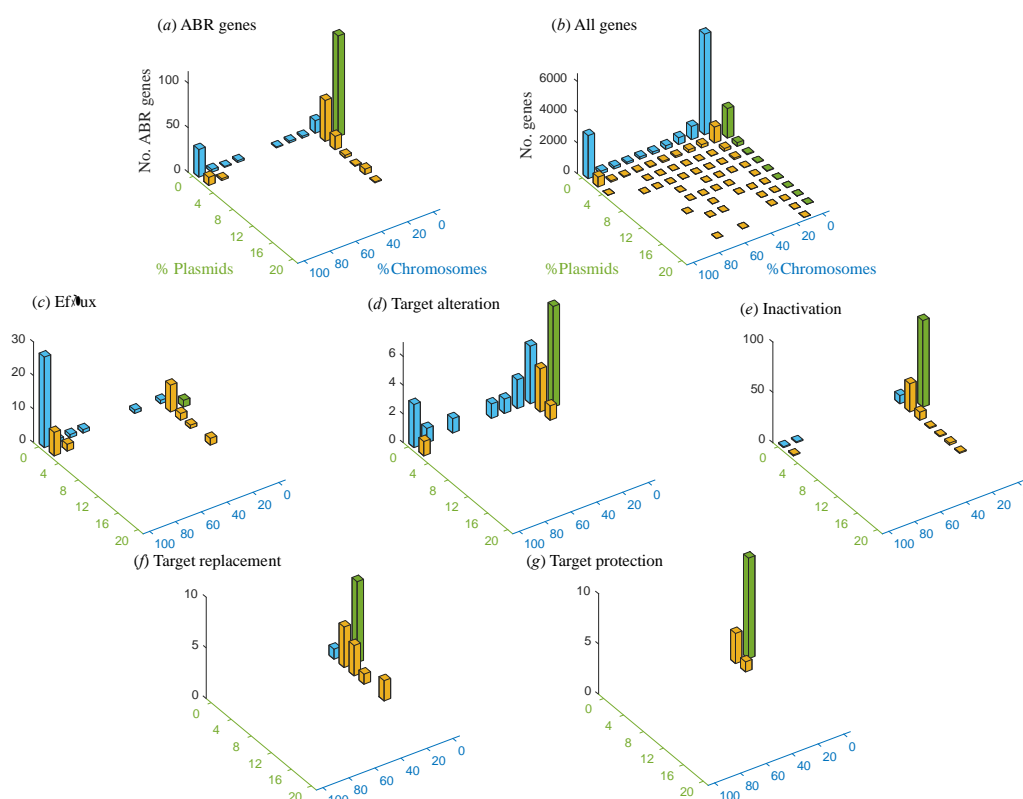


Figure 1. A comparison of gene prevalence in plasmids and chromosomes in *Escherichia* isolates. Plots in the figure present three-dimensional histograms of gene occurrence in plasmids (green axis) and chromosomes (blue axis). The height of bars in the histogram corresponds to the number of genes that are encoded in the same frequency group on chromosomes and plasmids. The bar coordinates correspond to the proportion of replicons (plasmids or chromosomes) where the genes occur. Note that no gene occurs in more than 20% of the plasmids; genes occurring in 100% of the chromosomes correspond to core gene families. Genes occurring only on chromosomes are shown with blue bars. Genes occurring only on plasmids are shown with green bars. Yellow bars correspond to genes that occur in both plasmids and chromosomes. (a) ABR genes in the examined isolates (data supplied in electronic supplementary material, table S3). (b) Gene occurrence of all *Escherichia* gene families. (c–g) ABR genes according to their classification into ABR mechanism (as in CARD [28]).

Previous studies suggested that the occurrence of ABR genes in plasmids may be related to the underlying mechanism of ABR (Turner et al., 2014; Rodríguez-Beltrán et al., 2020). Indeed, further examination of the ABR gene occurrence on plasmids and chromosomes reveals differences across the ABR mechanism classes. ABR genes that are involved in the expression of efflux pumps are typically core chromosomal genes that are rarely found on plasmids (figure 1c). By contrast, ABR genes classified as antibiotic target alteration include both chromosome- and plasmid-exclusive genes, with rare gene sharing between plasmids and chromosomes (figure 1d). ABR genes classified as antibiotic inactivation comprise mostly plasmid-specific genes, with some genes occurring in chromosomes as accessory genes and rarely as core genes (figure 1e). Genes classified as target protection and target replacement are encoded on a small number of plasmids and chromosomes (figure 1f, g).

To further examine the mode of ABR gene evolution, that is, if it is mainly by vertical inheritance or lateral gene transfer, we explored biases in the prevalence of ABR gene homologues towards plasmids and chromosomes (termed hereafter ‘replicon type’) for each of the ABR genes in our dataset. The results of the statistical analysis confirm that most of the ABR genes coding for efflux pumps are significantly more frequent on chromosomes (table 1; electronic supplementary material, table S3). Nonetheless, some exceptional efflux-coding genes are enriched on plasmids, including *tetA/B*, *floR* and *qacEΔ1*, whose homologues are mainly encoded in plasmids in all three genera (electronic supplementary material, table S3). The distribution of ABR genes that encode antibiotic target alteration proteins is balanced between chromosomes and plasmids in *Escherichia*, while in *Klebsiella* and *Salmonella* it is biased towards plasmids. Examples of chromosomal antibiotic target alteration genes include the gene encoding the translation elongation factor EF-Tu, which is found as an exclusive core gene in all genera, and *bacA*, an undecaprenyl pyrophosphate phosphatase (El Ghachi et al., 2004), which is a core chromosomal gene in *Escherichia* and *Salmonella*. Examples of genes encoding antibiotic target alteration proteins that are overrepresented on plasmids include *ermB* and *mcr-1* (electronic supplementary material, table S3).

Table 1. Comparison of ABR frequency in plasmids and chromosomes.

Genus	antibiotic efflux		antibiotic inactivation		antibiotic target alteration		antibiotic target protection		antibiotic target replacement	
	Chromosome	Plasmid	Chromosome	Plasmid	Chromosome	Plasmid	Chromosome	Plasmid	Chromosome	Plasmid
<i>Escherichia</i>	40(74%)	14(26%)	3(2%)	126(98%)	10(50%)	10(50%)	-	13(100%)	-	17(100%)
<i>Klebsiella</i>	14(50%)	14(50%)	3(3%)	100(97%)	8(40%)	12(60%)	-	11(100%)	-	13(100%)
<i>Salmonella</i>	20(61%)	13(39%)	4(5%)	74(95%)	5(38%)	8(62%)	-	9(100%)	-	13(100%)

The number of ABR genes having a bias towards plasmids or chromosomes according to the resistance mechanism classification (using Fisher's exact test in contingency analysis, $\alpha = 0.05$ and correction for multiple comparisons with false discovery rate (FDR) (Benjamini & Hochberg, 1995)). No bias to the chromosome was found in ABR genes for antibiotic target protection and replacement (marked with ‘—’).

ABR genes in the inactivation class are typically encoded on plasmids, with rare genes showing a bias towards the chromosome (table 1). Those exceptional genes include *ampH*, which is a core chromosomal gene in all three genera. This gene encodes a penicillin binding protein that has a function in the determination of cell shape in *Escherichia coli* and can confer resistance to β -lactam antibiotics (González-Leiza et al., 2011). Another example is *ampC1*, which codes for a β -lactamase and is nearly universally encoded in *Escherichia* and *Salmonella* but absent in *Klebsiella* genomes. Additional examples are *aac(6')-Iy* and *aac(6')-Iaa*, both encoding aminoglycoside acetyltransferases, which are exclusively found in *Salmonella* chromosomes. The genes *fosA5* and *fosA6*, mediating resistance against fosfomycin, have been reported in *Escherichia* (Guo et al., 2016), yet in our extensive dataset they are observed exclusively as accessory chromosomal genes in *Klebsiella* (electronic supplementary material, table S3).

Purifying selection on plasmid gene content due to genetic conflicts is expected to differ among taxa having different compositions of their core genome. Nonetheless, the bias of ABR genes towards specific replicon type (i.e. plasmids or chromosomes) is generally similar among the three tested genera in our dataset, likely since they are closely related and their core ABR genome largely overlaps. Nonetheless, 23 ABR genes stand out as exceptions (table 2). Those are mostly plasmid genes with variable presence on the chromosomes, i.e. they can be considered accessory genes. One extreme case is the antibiotic efflux gene *oqxA*, which is a core chromosomal gene in *Klebsiella* and enriched on plasmids in the other two genera. Among the exceptional genes encoding antibiotic inactivation proteins, the gene *bla_{CARB-1}* is found as an accessory chromosomal gene in *Salmonella*, with no bias towards specific replicon type, but only on plasmids in *Escherichia* and *Klebsiella* (electronic supplementary material, table S2).

Table 2. ABR genes whose relative frequency in plasmids (Pls) and chromosomes (Chr) differs among the three tested genera.

ARO	gene	resistance mechanism	<i>Escherichia</i>		<i>Salmonella</i>		<i>Klebsiella</i>		p	p.adj
			Chr	Pls	Chr	Pls	Chr	Pls		
3003922	<i>oqxA</i>	antibiotic efflux	2	45	2	36	728	1	1.02e-105	1.00e-103
3003923	<i>oqxB</i>	antibiotic efflux	2	31	2	19	162	0	5.51e-45	2.70e-43
3000165	<i>tetA</i>	antibiotic efflux	62	283	20	165	13	232	1.02e-05	7.71e-05
3000166	<i>tetB</i>	antibiotic efflux	69	100	39	58	3	6	9.47e-01	1.00e-01
3005010	<i>qacED1</i>	antibiotic efflux	55	291	68	137	116	371	1.73e-05	1.21e-04
3000873	<i>bla</i> _{TEM-1}	antibiotic inactivation	65	333	53	133	9	404	6.65e-22	2.17e-20
3004290	<i>ampC</i>	antibiotic inactivation	1119	1	2	0	0	0	-	-
3003689	<i>mcr-1</i>	antibiotic target alteration	13	95	0	26	0	8	1.10e-01	2.83e-01
3002790	<i>qnrS1</i>	antibiotic target protection	8	105	1	41	2	132	6.61e-02	2.06e-01
3000412	<i>sul2</i>	antibiotic target replacement	90	287	55	187	21	272	2.00e-09	2.79e-08
3002705	<i>floR</i>	antibiotic efflux	33	143	33	121	2	63	9.17e-04	4.99e-03
3000168	<i>tetD</i>	antibiotic efflux	6	1	1	4	2	50	3.69e-06	3.28e-05
3002660	<i>aph(6)-Id</i>	antibiotic inactivation	82	270	60	173	17	249	1.70e-10	4.16e-09
3002639	<i>aph(3'')-Ib</i>	antibiotic inactivation	82	263	59	167	18	245	4.26e-10	8.34e-09
3001059	<i>bla</i> _{SHV-1}	antibiotic inactivation	0	2	0	1	144	7	2.01e-04	1.23e-03
3002601	<i>aadA</i>	antibiotic inactivation	43	130	10	76	3	133	8.62e-09	1.06e-07
3001877	<i>bla</i> _{CTX-M-14}	antibiotic inactivation	23	43	2	13	7	51	7.64e-03	3.57e-02
3002013	<i>bla</i> _{CMY-2}	antibiotic inactivation	22	42	6	70	0	7	1.81e-04	1.18e-03
3002539	<i>aac(3)-IV</i>	antibiotic inactivation	5	37	4	45	12	21	4.48e-03	2.20e-02
3002603	<i>aadA3</i>	antibiotic inactivation	5	29	27	29	97	44	1.70e-08	1.85e-07
3002683	<i>catI</i>	antibiotic inactivation	35	27	26	35	9	81	5.23e-10	8.54e-09
3001926	<i>bla</i> _{CTX-M-65}	antibiotic inactivation	4	13	1	20	0	76	5.21e-04	3.00e-03
3004656	<i>catII</i>	antibiotic inactivation	0	7	4	19	1	88	8.84e-03	3.94e-02
3002862	<i>dfrA7</i>	antibiotic target replacement	13	2	21	23	0	1	9.83e-03	4.19e-02
3002240	<i>bla</i> _{CARB-1}	antibiotic inactivation	0	4	25	0	0	2	1.36e-06	1.33e-05
3001397	<i>bla</i> _{OXA-2}	antibiotic inactivation	3	0	0	1	0	10	2.75e-03	1.42e-02
3000410	<i>sul1</i>	antibiotic target replacement	53	286	66	140	76	371	8.05e-06	6.58e-05
3002863	<i>dfrA7</i>	antibiotic target replacement	13	2	21	23	0	1	9.83e-03	4.19e-02

The table shows ABR gene occurrence on plasmids and chromosomes in the three genera. The ABR gene frequency on both replicons was compared using Fisher's exact test (the p-value is reported in addition to adjusted p-value for multiple comparisons using false discovery rate (FDR)). Chromosome- and plasmid-exclusive genes and genes absent from any of the genera were excluded from this analysis (a total of 98 ABR genes were tested). Genes discussed in the Results 3(c)–3(e) are highlighted in italics.

ABR gene transfer between plasmids and chromosomes

Our results reveal 133 (33%) ABR genes that are found on both replicons in at least one genus, thus indicating that they have been laterally transferred. Of those, 75 ABR genes were observed on both plasmid and chromosome within the same isolate genome in at least one strain. The majority of genes that we identified at least once as duplicated are enriched on plasmids (65 out of 75 genes); of those, 45 genes encode proteins that function in antibiotic inactivation (electronic supplementary material, table S4). Only five of the 75 duplicated genes are enriched for a chromosome location; these correspond to four genes encoding efflux pumps and one encoding translation elongation factor (target alteration). The presence of duplicated ABR genes on both replicons in the same isolate is relatively rare. We note here that deleterious effects in the context of gene gain following lateral gene transfer (Baltrus, 2013) are expected not only for the recipient, but rather also for the donor, following gene transfer

from chromosomes to plasmids. Taken together, our results suggest that genes that are duplicated on the plasmid and chromosome in the same isolate often correspond to genes introduced into the genus as plasmid genes.

To further study the evolutionary history of genes shared between plasmids and chromosomes we reconstructed the phylogenetic trees of the ABR genes and examined their topology. In what follows, we present the results of 39 selected ABR genes exemplifying main trends in the data. The topology of ABR genes across replicon types and genera in the inferred phylogenetic trees confirms the occurrence of lateral gene transfer among genera and also replicon types in the history of all examined genes (electronic supplementary material, table S5). In the following, we present several prime examples.

Lateral transfer of genes encoding efflux pumps

The genes *oqxA* and *oqxB* belong to the resistance–nodulation–cell division (RND) family of bacterial efflux pumps (X.-Z. Li & Nikaido, 2016) and were first isolated from bacteria in swine manure from a farm where olaquinodox was used as a feed additive (Hansen et al., 2004). The proteins OqxA and OqxB, along with TolC, form an efflux pump that provides resistance against olaquinodox, chloramphenicol, quinolones, trimethoprim, tigecycline, nitrofurantoin, and also several detergents and disinfectants by reducing their concentrations in the cell (X.-Z. Li & Nikaido, 2016; J. Li et al., 2019). Both *oqxA* and *oqxB* are almost strictly chromosomal genes in *Klebsiella*, with *oqxA* occurring in 86% and *oqxB* occurring in 19% of *Klebsiella* isolates. We identified a single *Klebsiella* isolate where *oqxA* is present on both chromosome and plasmid (electronic supplementary material, table S4). The two genes have homologues in *Escherichia* and *Salmonella* plasmids. The phylogenetic trees of both genes indicate lateral transfer between *Klebsiella* chromosomes and plasmids of the other two genera (figure 2a; electronic supplementary material, figure S3a). We note that the *oqxA* protein sequences are nearly identical, and hence the phylogenetic reconstruction does not allow a reliable inference of donors and recipients in the transfer event. To further study the history of the ABR genes, we collected information on the isolation date of strains included in our analysis. The gene *oqxA* was documented early in the 1950s as a chromosomal gene in *Klebsiella* (figure 3). The first documentation of *oqxA* on a plasmid in our dataset is from 2004 in *Escherichia*; in recent years, this gene has been found on both plasmids and chromosomes in all three genera. The gene *oqxB* was first documented in 2004 in *Klebsiella* chromosomes and *Escherichia* plasmids, while in recent years, it has been described for both replicon types in all three genera (figure 3). The example of *oqxA* suggests that barriers to gene transfer of core chromosomal genes may differ between genera. The first report on plasmid-encoded *oqxAB* in *E. coli* in the literature dates back to the late 1990s (Hansen et al., 2005); today, after more than 20 years,

that gene is still very rare in *Escherichia* and *Salmonella*. It is tenable to speculate that barriers to plasmid-mediated *oqxA* acquisition are present also in *Escherichia* and *Salmonella*.

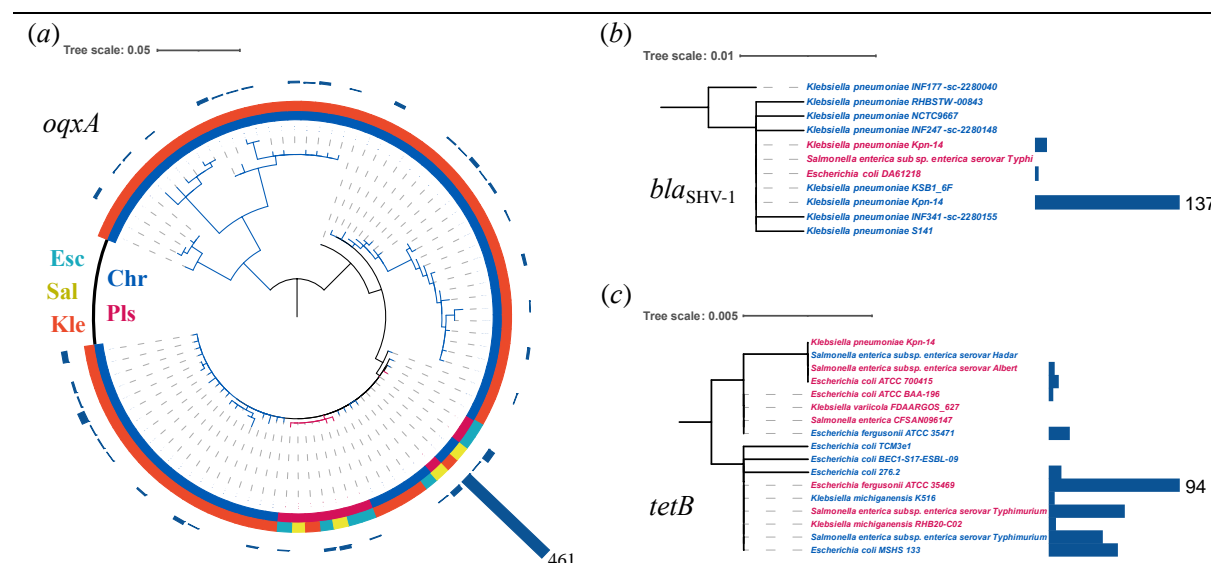


Figure 2. Lateral gene transfer in the evolution of ABR genes. Bars next to the operational taxonomic units show the frequency of redundant amino acid sequences (with highest frequency noted). (a) The outer ring shows the genus of ABR gene origin; the inner ring shows whether the gene is encoded on a plasmid (Pls) or chromosome (Chr). (b,c) Isolate label is coloured according to replicon type (plasmid genes in pink and chromosomal genes in blue). Most phylogenetic trees we reconstructed here for the ABR genes included polytomies, hence donor and recipients in the gene transfer events cannot be reliably inferred. Nonetheless, considering the presence of the examined ABR genes on both plasmids and chromosomes supports the hypothesis that they have been transferred at least once in their evolutionary history.

The Tet efflux pumps were the first efflux systems discovered in bacteria (McMurry et al., 1980; Ball et al., 1980). These major facilitator (MF) family exporters are membrane-associated proteins that pump out tetracycline from the cell in an energy-dependent manner (Roberts, 1996). Unlike the OqxAB–TolC system, the Tet efflux pumps provide resistance only against tetracyclines. Since the original report of *tetA*, many more *tet* genes have been discovered, the vast majority in Gram-negative bacteria (Poole, 2007). For our analysis, we focused on the *tetA* and *tetB* genes, since both genes are enriched on plasmids and rarely found on chromosomes (fewer than 6% in all genera). The phylogenetic reconstruction shows that the protein sequence of *tetA* is more diverged in comparison with that of *tetB*, and that for both genes, multiple transfer events occurred between chromosomes and plasmids, and also across bacterial genera (figure 2b; electronic supplementary material, figure S3d). Furthermore, we identified 13 isolates (from all genera) where *tetA* is encoded on both replicon types and four *Escherichia* isolates that encode *tetB* on both replicon types. The gene *tetA* in our data was documented firstly on a plasmid in *Salmonella* in 1974, while *tetB* was identified on a plasmid in *Escherichia* at the same time (figure 3). In recent years, both genes have been

found on both replicon types and in all three genera, yet *tetB* is seldom documented in *Klebsiella*.

Another efflux gene that is mainly found on plasmids is *qacEΔ1*, a truncated variant of the *qacE* gene (Kazama et al., 1999). The *qac* genes are antiseptic efflux systems primarily exporting quaternary ammonium compounds. Their substrate range, however, is quite large and includes intercalating dyes, biguanidines, diamidines and xanthenes. The gene *qacEΔ1* is frequently found in ABR gene cassettes isolated from bacteria exposed to these compounds compared with those that are not (De Oliveira et al., 2020). In our dataset, we found that *qacEΔ1* is very often encoded on both plasmids and chromosomes in the same isolate (71 *Klebsiella*, 11 *Escherichia* and 7 *Salmonella* isolates). The protein sequence is highly conserved between replicons and across all three genera. In our dataset, *qacEΔ1* was documented first on a plasmid in *Escherichia* in 1969, and then a *Klebsiella* plasmid in 1981; the first chromosomal *qacEΔ1* was identified in *Escherichia* in 1990 (figure 3).

Lateral transfer of genes encoding antibiotic inactivation proteins

A prime example of antibiotic inactivation enzymes are the β -lactamases. These enzymes confer resistance to β -lactam antibiotics by cleaving their central β -lactam ring (Livermore, 1995). The gene *bla*_{TEM-1} was the first TEM β -lactamase gene identified; TEM β -lactamases are among the best-studied ABR enzymes (Datta & Kontomichalou, 1965; Salverda et al., 2010). Although the original *bla*_{TEM-1} gene confers resistance only to penicillins and early cephalosporins, its later variants also provide resistance against second-, third- and fourth-generation cephalosporins, monobactams and β -lactamase inhibitors (Salverda et al., 2010). The *bla*_{TEM-1} gene is mainly found on *Klebsiella* and *Escherichia* plasmids. The phylogenetic reconstruction indicates multiple transfer events and high protein sequence conservation of *bla*_{TEM-1} between replicons and across genera (figure 2c). It was first documented on plasmids in both *Escherichia* and *Salmonella* in 1974, and first reported on chromosomes in *Salmonella* in 1981 (figure 3).

The AmpC β -lactamase of *Escherichia coli* was the first reported bacterial enzyme capable of degrading penicillin (Abraham & Chain, 1940). Although active against penicillins, this inducible enzyme has higher activity against cephalosporins and is an important determinant of resistance against cephalosporins in clinical settings (Jacoby, 2009). The *ampC* gene in *Escherichia* is nearly exclusively a chromosomal gene. Additionally, homologues of *ampC* were found on a plasmid in one *Escherichia* isolate and on the chromosome of two *Salmonella* isolates. The phylogenetic reconstruction indicates that the *Salmonella* homologues were acquired in two independent transfer events (electronic supplementary material, figure S3h).

Lateral transfer of ABR genes involved in target alteration, replacement and protection

The *mcr-1* gene was the first plasmid-mediated colistin resistance determinant reported, signalling the breach of the last resort polymyxin antibiotics (Liu et al., 2016). MCR-1 is a phosphoethanolamine transferase that alters lipid A in the outer membrane of Gram-negative bacteria, resulting in reduced affinity to colistin (Hinchliffe et al., 2017). Homologues of MCR-1 are mainly found on plasmids in all three genera, with some rare homologues on chromosomes in *Escherichia* (1%). Three *Escherichia coli* isolates were identified with *mcr-1* on both replicons. The nearly identical protein sequence of the homologues suggests that *mcr-1* is laterally transferred among the genera by plasmids (electronic supplementary material, figure S3e). The gene was first documented on a plasmid in *Escherichia* in 2007 and on the chromosome in 2014; in recent years it has been found in all three genera (figure 3).

The gene *sul2* confers resistance against the sulfonamide antibiotics (Swedberg & Skold, 1983; Radstrom & Swedberg, 1988). Sulfonamides were among the first synthetic compounds for which specific antibacterial activities were discovered, and they were also among the first antibiotics to be used for medical treatment. They block folic acid synthesis in bacteria by competitively inhibiting dihydropteroate synthase (DHPS) (Sköld, 2000). Sul2 is an alternative drug-resistant variant of the DHPS enzyme that replaces the original drug-sensitive variant to provide resistance (Skold, 1976). The gene is mainly found on plasmids and may occur also on chromosomes (6% of isolates). A total of 34 isolates encode a *sul2* on both replicons, and the phylogenetic tree of *sul2* indicates multiple transfer events between replicons and across genera (electronic supplementary material, figure S3f). Although typically a plasmid gene, *sul2* was first documented on the chromosome in *Klebsiella* in the 1950s and only later on plasmids in *Escherichia* and *Salmonella*; in recent years, it has been reported in all three genera (figure 3).

The gene *qnrS1* belongs to the Qnr family of proteins, which confer resistance to quinolones and fluoroquinolones in Gram-negative bacteria by binding and protecting the target of these antibiotics, type II topoisomerase (Wilson et al., 2020). The binding of the target seems to destabilize the complex of topoisomerase with the drug, resulting in regeneration of the active enzyme (Jacoby et al., 2014). Homologues of *qnrS1* are mainly distributed on plasmids. We identified two isolates that carry this gene on both replicons, including *E. coli* strain LHM10-1 (isolated in 2017 in China from swine faeces) and *Klebsiella pneumoniae* strain C2421 (isolated in 2017 in China from human urine sample). The QnrS1 protein sequences are nearly identical in all plasmids and the phylogenetic tree topology indicates lateral gene transfer between replicons and in all three genera (electronic supplementary material, figure S3g). The gene was first documented on plasmids in *Escherichia* and *Klebsiella* in 2007 and in recent years it has been found on both replicons in all three genera (figure 3).

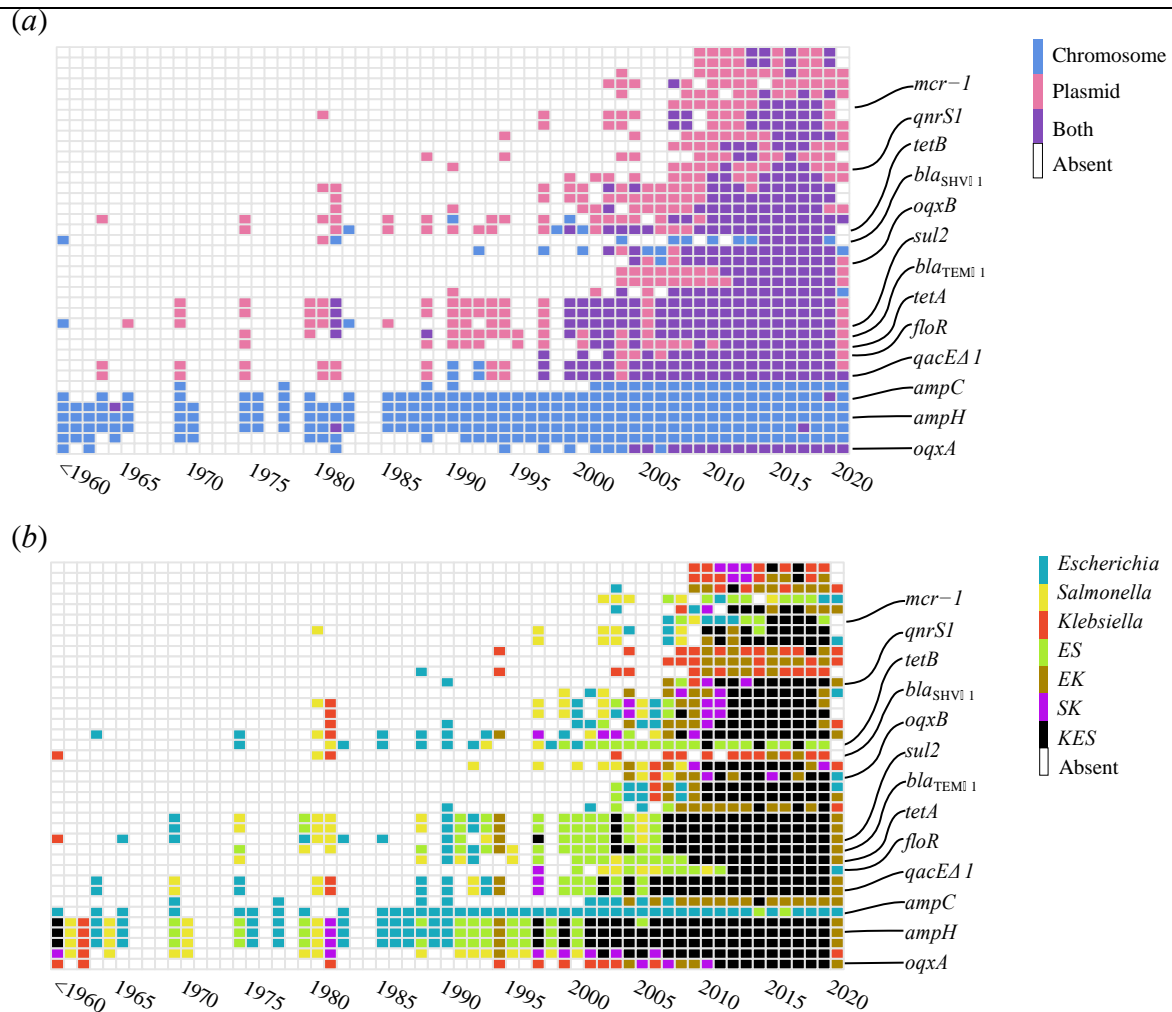


Figure 3. Temporal pattern of ABR gene documentation in replicons and genera. Gene presence according to the isolation date of the strains is summarized. The information on the strains isolated between 1900 and 1960 is condensed in '<1960'. The figure shows data for 39 genes whose phylogenetic trees were closely examined (ordered as in electronic supplementary material, table S4). The names of 13 genes highlighted in the manuscript are marked. Different colours correspond to gene locations. (a) Temporal ABR gene pattern according to replicon type. (b) Temporal ABR gene pattern according to genus. ABR genes found in multiple genera are annotated by capitalized abbreviations of the three genera, i.e. *E* for *Escherichia*, *K* for *Klebsiella*, *S* for *Salmonella*.

Temporal pattern of gene presence in replicons supports the phylogenetic inference

To further study the temporal pattern of ABR gene occurrence on specific replicon type - plasmids or chromosomes - we examined the collection date of isolate genomes encoding ABR genes. The available data reveal two general trends: core ABR genes were typically documented on chromosomes, with only rare findings on plasmids throughout the time range covered by the data. Notably, of the 75 core ABR genes that are enriched on chromosomes, 74 (99%) were documented first on chromosomes, with *oqxB* as the sole exception. Additionally, of the 226 genes found to be enriched on plasmids, 208 (92%) were first identified on plasmids. The temporal pattern of plasmid genes shows a trend of early documentation on plasmids and later presence on both replicons, which is in agreement with the phylogenetic

trees reconstructed in this study (figure 2). The temporal pattern of ABR gene occurrence as observed in isolate genomes provides further support for the role of historical contingency in the bias of ABR gene occurrence on plasmids or chromosomes.

Data and methods

To study the pattern of ABR gene occurrence in bacterial replicons (chromosomes and plasmids), we examined all complete genomes of plasmid-carrying isolates in the genera *Escherichia*, *Salmonella* and *Klebsiella* from the NCBI genome database (electronic supplementary material, table S1). Assembly and annotation files of 2580 genomes were downloaded from the NCBI RefSeq database (O’Leary et al., 2016) (v. 01/2021). An additional 78 complete genomes lacking RefSeq annotation were downloaded from the GenBank database. Protein-coding genes of the 78 genomes were annotated using Prokka (Seemann, 2014) (v. 1.14.6, with parameter `--kingdom Bacteria --gcode 11 --genus (Escherichia/Salmonella/Klebsiella) --usegenus --evaluate 1 × 10-9`). Isolate metadata were downloaded from the BioSample database. Samples lacking host information were examined in detail. A total of 23 strains sequenced within the framework of laboratory experiments were excluded. The surveyed genomes include: 1169 chromosomes and 3286 plasmids from *Escherichia*; 846 chromosomes and 3020 plasmids from *Klebsiella*; 620 chromosomes and 1090 plasmids from *Salmonella*. Homologues to genes that may provide ABR were identified based on the Comprehensive Antibiotic Resistance Database (CARD; v. 3.1.0) using the Resistance Gene Identifier tool (RGI; v. 5.1.1, with parameter `--clean`) (Alcock et al., 2020). The application of RGI uncovered 126 910 protein-coding genes; of these, we excluded 4774 protein-coding genes whose sequence had less than 70% identity or was less than 90% shorter or greater than 30% longer in comparison with the CARD ABR genes. Of the total 3044 genes in CARD, 416 (14%) ABR genes had 122 136 homologues in the examined isolates. Of those, we further excluded 9 genes belonging to multiple ABR mechanism classes.

The inference of protein families in *Escherichia* was performed using a dataset of 599 completely sequenced *Escherichia* strains downloaded from the NCBI RefSeq database (v. 2018) (previously reported in (Wein et al., 2021)). Of the total strains, 416 plasmid-carrying isolates were retained for further analysis (electronic supplementary material, table S2). Genome-wise reciprocal best hits (RBHs) of protein sequences were identified using MMseqs2 (Steinegger & Söding, 2017) (v. 13.45111, with module `easy-rbh`, applying a threshold of $E\text{-value} \leq 1 \times 10^{-10}$). RBHs were further compared by global alignment using Parasail-python (Daily, 2016) (v. 1.2.4, with the Needleman–Wunsch algorithm). Sequence pairs with greater than or equal to 30% identical amino acids were clustered into protein families using the Markov clustering algorithm (MCL) (Enright et al., 2002) (v. 12-135, with parameter `--abc -l 2.0`).

Data analysis, statistical inference and visualization were performed with R (v. 4.0) and MATLAB® (v. 2018b). Testing for a bias in ABR gene occurrence towards plasmids or chromosomes was performed by comparing the frequency of ABR gene occurrence on both replicons with the frequency of the remaining genes on both replicons (using Fisher's exact test with `fisher.test` function in R).

For the phylogenetic reconstruction, non-redundant protein sequences of selected ABR genes were aligned using MAFFT (Kato & Standley, 2013) (v. 7.475). Owing to the large number of identical protein sequences among the ABR gene homologues, the sequences were filtered to include non-redundant amino acid sequences at the resolution of replicon and genus. Phylogenetic trees were reconstructed using IQ-TREE (Minh et al., 2020) (v. 1.6.12) with restricted automatic model selection to the Le & Gascuel (LG) model (Le & Gascuel, 2008) and the LG4X model (Le et al., 2012) as additional alternative (with parameter `-mset LG -madd LG4X`). The LG model was chosen owing to its suitability for the inference of phylogenetic trees using protein sequences from bacterial organisms by means of maximum likelihood methods. The resulting trees were rooted using the midpoint criterion and visualized using iTOL (Letunic & Bork, 2019) v. 6 (<https://itol.embl.de/>).

Discussion

In this study, we performed a comprehensive ABR gene analysis across 2635 complete genome sequences in the related bacterial genera *Escherichia*, *Salmonella* and *Klebsiella*, and identified an inverse association between ABR gene prevalence in plasmids and chromosomes. This inverse association strongly suggests that the presence of a homologous gene in the chromosome poses a barrier to plasmid-mediated ABR gene acquisition. Several non-exclusive mechanisms may account for the barrier to lateral gene transfer, as explained in detail below.

The bias of ABR genes towards specific replicon type—chromosomes or plasmids—is clearly associated with the ABR underlying mechanism (Rodríguez-Beltrán et al., 2020). ABR genes associated with efflux pumps are primarily encoded on chromosomes, those involved in target alteration can be encoded on both replicon types, and those for the remaining three mechanisms are almost exclusively encoded on plasmids. ABR genes encoding efflux pumps are frequently part of the core genome (e.g. in the genera we studied here), where they code for proteins having multiple functions in the cell outside the context of ABR. Examples are proteins involved in export of heavy metals (Nies, 2003), organic pollutants (Rojas et al., 2001) and endogenous metabolites (C. Ruiz & Levy, 2014), and cell signalling molecules (Evans et al., 1998). Furthermore, several efflux pumps are important for biofilm formation (Fahmy et al., 2016) and bacteria–plant interactions (García-León et al., 2014). Core genes are expected to be highly integrated in the cellular transcriptional regulation network and their products are

likely well integrated in protein–protein interaction networks. Proteins encoded by the core genome are therefore expected to have a high level of pleiotropic effects on bacterial physiology. Barriers for lateral transfer of core genes are expected to be high, e.g. owing to genetic conflicts (Rodríguez-Beltrán et al., 2020) or the cost of their interference with well-coordinated cellular processes (Baltrus, 2013). The same logic applies to core genes in the other classes of resistance mechanisms, including target alteration and antibiotic inactivation. Core ABR genes that code for antibiotic target alteration in our dataset include *gyrA*, *parC* and *parE*. These genes may provide resistance to antibiotics that target, e.g., type II topoisomerases, DNA gyrase and topoisomerase IV, which control topological transitions of DNA (J. Ruiz, 2003); all of these are central components in the bacterial information processing mechanisms that are targets for drug therapy (e.g. with fluoroquinolones (Baquero & Levin, 2021)). The frequency of ABR gene transfer in microbial evolution (i.e. transferability) is therefore expected to depend on the complexity of their function within the cell (Jain et al., 1999b; Cohen et al., 2011b) rather than the exact mechanism of resistance, similarly to other genes that do not code for ABR mechanisms.

So far, we have discussed barriers to lateral transfer of ABR genes depending on the effect of gene acquisition on the bacterial organism. Notwithstanding, the acquisition of ABR genes may have consequences also for the evolution of stably inherited plasmids. Plasmids encoding ABR genes may also evolve a stable inheritance in a newly colonized host population in the absence of selective conditions for ABR (Wein et al., 2019). By contrast, the presence of selective conditions for ABR can maintain in the population plasmids whose inheritance is unstable; such plasmids are at risk of extinction under non-selective conditions (Wein et al., 2020). Indeed, plasmids equipped with mobility mechanisms have the capacity to persist via transfer to alternative host populations (Hall et al., 2016). Nonetheless, the stability of plasmid vertical inheritance is paramount for plasmid survival and long-term evolution (reviewed in (Wein & Dagan, 2020)). The potential deleterious effect of ABR gene acquisition in plasmid genomes on the plasmid fitness (Hülter et al., 2020) may explain the rarity of ABR genes on plasmids (figure 1), specifically in non-mobile plasmids (Wein et al., 2020).

Plasmid-encoded ABR genes in our dataset are mostly classified as antibiotic target inactivation (figure 1). ABR genes coding for antibiotic inactivation frequently vary in their relative frequency on plasmids and chromosomes across the three compared genera. Many of those genes correspond to accessory chromosomal genes that are also prevalent on plasmids, indicating that they may be transferred between both replicon types (**table 2**). Indeed, accessory genes may prove essential depending on the isolate genetic background and environmental conditions (Rousset et al., 2021); nonetheless, their variable presence in a genus suggests that they are often dispensable. Genes whose evolution is characterized by rapid gain and loss dynamics are expected to encode proteins whose function is weakly

dependent on specific genetic backgrounds, and are therefore more likely to evolve by lateral gene transfer (Novick & Doolittle, 2020). The ABR gene content of plasmids reflects the propensity of specific ABR genes to be successfully transferred among donors and recipients within the plasmid host range.

Data accessibility

The data are provided in electronic supplementary material. Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.5677445>.

Author Contributions

Yiqing Wang and Tal Dagan conceived the study.

Yiqing Wang performed data analysis and visualization.

Aditi Batra analysed distribution patterns in ABR gene families.

All authors interpreted the results and wrote the manuscript.

Acknowledgements

We thank Giddy Landan for advice on data analysis. All calculations were performed on the HPC-system in Kiel University Computing Centre (Rechenzentrum).

References

- Abraham, E. P., & Chain, E. (1940). An enzyme from bacteria able to destroy penicillin. *Nature*, 146(3713), 837. <https://doi.org/10.1038/146837A0>
- Acar Kirit, H., Lagator, M., & Bollback, J. P. (2020). Experimental determination of evolutionary barriers to horizontal gene transfer. *BMC Microbiology*, 20(1), 326. <https://doi.org/10.1186/S12866-020-01983-5>
- Alcock, B. P., Raphenya, A. R., Lau, T. T. Y., Tsang, K. K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen, A. L. V., Cheng, A. A., Liu, S., Min, S. Y., Miroshnichenko, A., Tran, H. K., Werfalli, R. E., Nasir, J. A., Oloni, M., Speicher, D. J., Florescu, A., Singh, B., ... McArthur, A. G. (2020). CARD 2020: Antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Research*, 48(D1), D517–D525. <https://doi.org/10.1093/NAR/GKZ935>
- Andersson, D. I., Balaban, N. Q., Baquero, F., Courvalin, P., Glaser, P., Gophna, U., Kishony, R., Molin, S., & Tønjum, T. (2021). Antibiotic resistance: Turning evolutionary principles into clinical reality. *FEMS Microbiology Reviews*, 44(2), 171-188. <https://doi.org/10.1093/FEMSRE/FUAA001>
- Ball, P. R., Shales, S. W., & Chopra, I. (1980). Plasmid-mediated tetracycline resistance in *Escherichia coli* involves increased efflux of the antibiotic. *Biochemical and Biophysical*

- Research Communications*, 93(1), 74–81. [https://doi.org/10.1016/S0006-291X\(80\)80247-6](https://doi.org/10.1016/S0006-291X(80)80247-6)
- Baltrus, D. A. (2013). Exploring the costs of horizontal gene transfer. *Trends in Ecology and Evolution*, 28(8), 489–495. <https://doi.org/10.1016/J.TREE.2013.04.002>
- Baquero, F., & Levin, B. R. (2021). Proximate and ultimate causes of the bactericidal action of antibiotics. *Nature Reviews Microbiology*, 19(2), 123–132. <https://doi.org/10.1038/S41579-020-00443-1>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289–300. <https://doi.org/10.1111/J.2517-6161.1995.TB02031.X>
- Blanco, P., Hernando-Amado, S., Reales-Calderon, J. A., Corona, F., Lira, F., Alcalde-Rico, M., Bernardini, A., Sanchez, M. B., & Martinez, J. L. (2016). Bacterial multidrug efflux pumps: Much more than antibiotic resistance determinants. *Microorganisms*, 4(1), 14. <https://doi.org/10.3390/MICROORGANISMS4010014>
- Blau, K., Bettermann, A., Jechalke, S., Fornefeld, E., Vanrobaeys, Y., Stalder, T., Top, E. M., & Smalla, K. (2018). The transferable resistome of produce. *MBio*, 9(6), 1–15. <https://doi.org/10.1128/MBIO.01300-18>
- Che, Y., Yang, Y., Xu, X., Brinda, K., Polz, M. F., Hanage, W. P., & Zhang, T. (2021). Conjugative plasmids interact with insertion sequences to shape the horizontal transfer of antimicrobial resistance genes. *Proceedings of the National Academy of Sciences of the United States of America*, 118(6), e2008731118. <https://doi.org/10.1073/PNAS.2008731118>
- Cohen, O., Gophna, U., & Pupko, T. (2011a). The complexity hypothesis revisited: Connectivity rather than function constitutes a barrier to horizontal gene transfer. *Molecular Biology and Evolution*, 28(4), 1481–1489. <https://doi.org/10.1093/MOLBEV/MSQ333>
- Daily, J. (2016). Parasail: SIMD C library for global, semi-global, and local pairwise sequence alignments. *BMC Bioinformatics*, 17(1). <https://doi.org/10.1186/S12859-016-0930-Z>
- Datta, N., & Kontomichalou, P. (1965). Penicillinase synthesis controlled by infectious R factors in Enterobacteriaceae. *Nature*, 208(5007), 239–241. <https://doi.org/10.1038/208239A0>
- Dcosta, V. M., King, C. E., Kalan, L., Morar, M., Sung, W. W. L., Schwarz, C., Froese, D., Zazula, G., Calmels, F., Debruyne, R., Golding, G. B., Poinar, H. N., & Wright, G. D. (2011). Antibiotic resistance is ancient. *Nature*, 477(7365), 457–461. <https://doi.org/10.1038/NATURE10388>
- De Oliveira, D. M. P., Forde, B. M., Kidd, T. J., Harris, P. N. A., Schembri, M. A., Beatson, S. A., Paterson, D. L., & Walker, M. J. (2020). Antimicrobial resistance in ESKAPE

- pathogens. *Clinical Microbiology Reviews*, 33(3), e00181-19.
<https://doi.org/10.1128/CMR.00181-19>
- Domingues, S., Harms, K., Fricke, W. F., Johnsen, P. J., da Silva, G. J., & Nielsen, K. M. (2012). Natural Transformation Facilitates Transfer of Transposons, Integrons and Gene Cassettes between Bacterial Species. *PLoS Pathogens*, 8(8), e1002837.
<https://doi.org/10.1371/JOURNAL.PPAT.1002837>
- El Ghachi, M., Bouhss, A., Blanot, D., & Mengin-Lecreulx, D. (2004). The bacA gene of *Escherichia coli* encodes an undecaprenyl pyrophosphate phosphatase activity. *Journal of Biological Chemistry*, 279(29), 30106–30113.
<https://doi.org/10.1074/JBC.M401701200>
- Enright, A. J., Van Dongen, S., & Ouzounis, C. A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research*, 30(7), 1575–1584.
<https://doi.org/10.1093/NAR/30.7.1575>
- Evans, K., Passador, L., Srikumar, R., Tsang, E., Nezezon, J., & Poole, K. (1998). Influence of the MexAB-OprM multidrug efflux system on quorum sensing in *Pseudomonas aeruginosa*. *Journal of Bacteriology*, 180(20), 5443–5447.
<https://doi.org/10.1128/JB.180.20.5443-5447.1998>
- Fahmy, A., Srinivasan, A., & Webber, M. A. (2016). The relationship between bacterial multidrug efflux pumps and biofilm formation. In *Efflux-Mediated Antimicrobial Resistance in Bacteria* (pp. 651–663). Springer International Publishing.
https://doi.org/10.1007/978-3-319-39658-3_25
- Gallagher, L. A., Lee, S. A., & Manoil, C. (2017). Importance of core genome functions for an extreme antibiotic resistance trait. *MBio*, 8(6), e01655-17.
<https://doi.org/10.1128/MBIO.01655-17>
- García-León, G., Hernández, A., Hernando-Amado, S., Alavi, P., Berg, G., & Martínez, J. L. (2014). A function of SmeDEF, the major quinolone resistance determinant of *Stenotrophomonas maltophilia*, is the colonization of plant roots. *Applied and Environmental Microbiology*, 80(15), 4559–4565. <https://doi.org/10.1128/AEM.01058-14>
- González-Leiza, S. M., de Pedro, M. A., & Ayala, J. A. (2011). Amph, a bifunctional DD-endopeptidase and DD-carboxypeptidase of *Escherichia coli*. *Journal of Bacteriology*, 193(24), 6887–6894. <https://doi.org/10.1128/JB.05764-11>
- Guo, Q., Tomich, A. D., McElheny, C. L., Cooper, V. S., Stoesser, N., Wang, M., Sluis-Cremer, N., & Doi, Y. (2016). Glutathione-S-transferase FosA6 of *Klebsiella pneumoniae* origin conferring fosfomycin resistance in ESBL-producing *Escherichia coli*. *Journal of Antimicrobial Chemotherapy*, 71(9), 2460–2465.
<https://doi.org/10.1093/JAC/DKW177>

- Hall, J. P. J., Wood, A. J., Harrison, E., & Brockhurst, M. A. (2016). Source-sink plasmid transfer dynamics maintain gene mobility in soil bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America*, 113(29), 8260–8265. <https://doi.org/10.1073/PNAS.1600974113>
- Hansen, L. H., Johannesen, E., Burmølle, M., Sørensen, A. H., & Sørensen, S. J. (2004). Plasmid-encoded multidrug efflux pump conferring resistance to olaquinox in *Escherichia coli*. *Antimicrobial Agents and Chemotherapy*, 48(9), 3332–3337. <https://doi.org/10.1128/AAC.48.9.3332-3337.2004>
- Hansen, L. H., Sørensen, S. J., Jørgensen, H. S., & Jensen, L. B. (2005). The prevalence of the OqxAB multidrug efflux pump amongst olaquinox-resistant *Escherichia coli* in pigs. *Microbial Drug Resistance*, 11(4), 378–382. <https://doi.org/10.1089/MDR.2005.11.378>
- Heuer, H., Schmitt, H., & Smalla, K. (2011). Antibiotic resistance gene spread due to manure application on agricultural fields. *Current Opinion in Microbiology*, 14(3), 236–243. <https://doi.org/10.1016/J.MIB.2011.04.009>
- Hinchliffe, P., Yang, Q. E., Portal, E., Young, T., Li, H., Tooke, C. L., Carvalho, M. J., Paterson, N. G., Brem, J., Niumsup, P. R., Tansawai, U., Lei, L., Li, M., Shen, Z., Wang, Y., Schofield, C. J., Mulholland, A. J., Shen, J., Fey, N., ... Spencer, J. (2017). Insights into the mechanistic basis of plasmid-mediated colistin resistance from crystal structures of the catalytic domain of MCR-1. *Scientific Reports*, 7, 39392. <https://doi.org/10.1038/SREP39392>
- Hülter, N. F., Wein, T., Effe, J., Garoña, A., & Dagan, T. (2020). Intracellular Competitions Reveal Determinants of Plasmid Evolutionary Success. *Frontiers in Microbiology*, 11, 2062. <https://doi.org/10.3389/FMICB.2020.02062>
- Jacoby, G. A. (2009). AmpC B-Lactamases. *Clinical Microbiology Reviews*, 22(1), 161–182. <https://doi.org/10.1128/CMR.00036-08>
- Jacoby, G. A., Strahilevitz, J., & Hooper, D. C. (2014). Plasmid-Mediated Quinolone Resistance. *Microbiology Spectrum*, 2(5), 33. <https://doi.org/10.1128/MICROBIOLSPEC.PLAS-0006-2013>
- Jain, R., Rivera, M. C., & Lake, J. A. (1999a). Horizontal gene transfer among genomes: The complexity hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, 96(7), 3801–3806. <https://doi.org/10.1073/PNAS.96.7.3801>
- Jørgensen, T. S., Xu, Z., Hansen, M. A., Sørensen, S. J., & Hansen, L. H. (2014). Hundreds of circular novel plasmids and DNA elements identified in a rat cecum metagenome. *PLoS ONE*, 9(2), e87924. <https://doi.org/10.1371/JOURNAL.PONE.0087924>
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772–780. <https://doi.org/10.1093/MOLBEV/MST010>

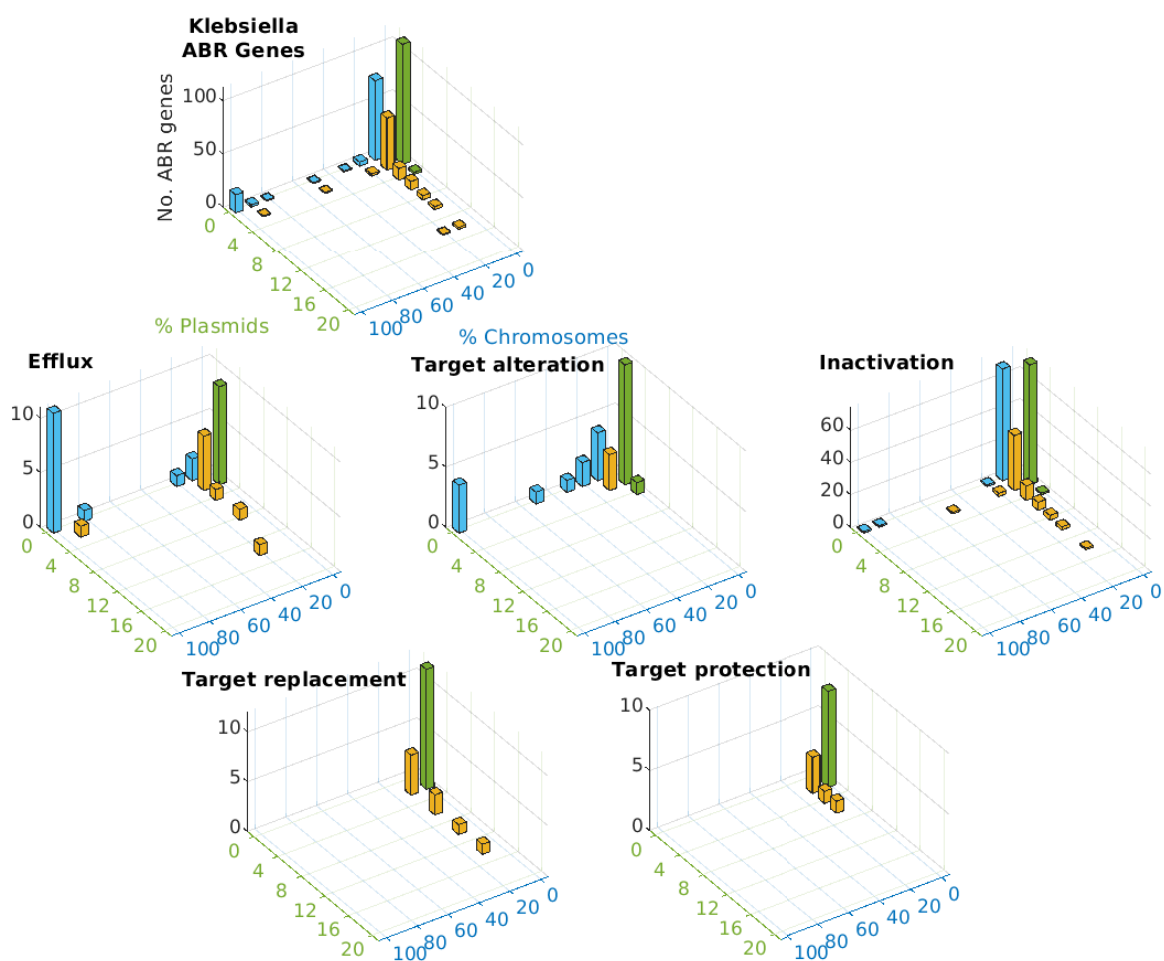
- Kazama, H., Hamashima, H., Sasatsu, M., & Arai, T. (1999). Characterization of the antiseptic-resistance gene *qacEΔ1* isolated from clinical and environmental isolates of *Vibrio parahaemolyticus* and *Vibrio cholerae* non-O1. *FEMS Microbiology Letters*, 174(2), 379–384. <https://doi.org/10.1111/J.1574-6968.1999.TB13593.X>
- Laxminarayan, R., Duse, A., Wattal, C., Zaidi, A. K. M., Wertheim, H. F. L., Sumpradit, N., Vlieghe, E., Hara, G. L., Gould, I. M., Goossens, H., Greko, C., So, A. D., Bigdeli, M., Tomson, G., Woodhouse, W., Ombaka, E., Peralta, A. Q., Qamar, F. N., Mir, F., ... Cars, O. (2013). Antibiotic resistance-the need for global solutions. *The Lancet Infectious Diseases*, 13(12), 1057–1098. [https://doi.org/10.1016/S1473-3099\(13\)70318-9](https://doi.org/10.1016/S1473-3099(13)70318-9)
- Le, S. Q., Dang, C. C., & Gascuel, O. (2012). Modeling protein evolution with several amino acid replacement matrices depending on site rates. *Molecular Biology and Evolution*, 29(10), 2921–2936. <https://doi.org/10.1093/MOLBEV/MSS112>
- Le, S. Q., & Gascuel, O. (2008). An improved general amino acid replacement matrix. *Molecular Biology and Evolution*, 25(7), 1307–1320. <https://doi.org/10.1093/MOLBEV/MSN067>
- Letunic, I., & Bork, P. (2019). Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Research*, 47(W1), W256–W259. <https://doi.org/10.1093/NAR/GKZ239>
- Li, J., Zhang, H., Ning, J., Sajid, A., Cheng, G., Yuan, Z., & Hao, H. (2019). The nature and epidemiology of OqxAB, a multidrug efflux pump. *Antimicrobial Resistance and Infection Control*, 8(1), 44. <https://doi.org/10.1186/S13756-019-0489-3>
- Li, X.-Z., & Nikaido, H. (2016). Antimicrobial Drug Efflux Pumps in *Escherichia coli*. In *Efflux-Mediated Antimicrobial Resistance in Bacteria* (pp. 219–259). Springer International Publishing. https://doi.org/10.1007/978-3-319-39658-3_9
- Liu, Y. Y., Wang, Y., Walsh, T. R., Yi, L. X., Zhang, R., Spencer, J., Doi, Y., Tian, G., Dong, B., Huang, X., Yu, L. F., Gu, D., Ren, H., Chen, X., Lv, L., He, D., Zhou, H., Liang, Z., Liu, J. H., & Shen, J. (2016). Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: A microbiological and molecular biological study. *The Lancet Infectious Diseases*, 16(2), 161–168. [https://doi.org/10.1016/S1473-3099\(15\)00424-7](https://doi.org/10.1016/S1473-3099(15)00424-7)
- Livermore, D. M. (1995). Beta-Lactamases in laboratory and clinical resistance. *Clinical Microbiology Reviews*, 8(4), 557–584. <https://doi.org/10.1128/CMR.8.4.557>
- Mazaheri Nezhad Fard, R., Barton, M. D., & Heuzenroeder, M. W. (2011). Bacteriophage-mediated transduction of antibiotic resistance in enterococci. *Letters in Applied Microbiology*, 52(6), 559–564. <https://doi.org/10.1111/J.1472-765X.2011.03043.X>

- McMurry, L., Petrucci, R. E., & Levy, S. B. (1980). Active efflux of tetracycline encoded by four genetically different tetracycline resistance determinants in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, 77(7 II), 3974–3977. <https://doi.org/10.1073/PNAS.77.7.3974>
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., Lanfear, R., & Teeling, E. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, 37(5), 1530–1534. <https://doi.org/10.1093/MOLBEV/MSAA015>
- Nies, D. H. (2003). Efflux-mediated heavy metal resistance in prokaryotes. *FEMS Microbiology Reviews*, 27(2–3), 313–339. [https://doi.org/10.1016/S0168-6445\(03\)00048-2](https://doi.org/10.1016/S0168-6445(03)00048-2)
- Novick, A., & Doolittle, W. F. (2020). Horizontal persistence and the complexity hypothesis. *Biology and Philosophy*, 35(1), 2. <https://doi.org/10.1007/S10539-019-9727-6>
- O’Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., ... Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, 44(D1), D733–D745. <https://doi.org/10.1093/NAR/GKV1189>
- Poole, K. (2007). Efflux pumps as antimicrobial resistance mechanisms. *Annals of Medicine*, 39(3), 162–176. <https://doi.org/10.1080/07853890701195262>
- Popa, O., Landan, G., & Dagan, T. (2017). Phylogenomic networks reveal limited phylogenetic range of lateral gene transfer by transduction. *ISME Journal*, 11(2), 543–554. <https://doi.org/10.1038/ISMEJ.2016.116>
- Radstrom, P., & Swedberg, G. (1988). RSF1010 and a conjugative plasmid contain *sulIII*, one of two known genes for plasmid-borne sulfonamide resistance dihydropteroate synthase. *Antimicrobial Agents and Chemotherapy*, 32(11), 1684–1692. <https://doi.org/10.1128/AAC.32.11.1684>
- Redondo-Salvo, S., Fernández-López, R., Ruiz, R., Vielva, L., de Toro, M., Rocha, E. P. C., Garcillán-Barcia, M. P., & de la Cruz, F. (2020). Pathways for horizontal gene transfer in bacteria revealed by a global map of their plasmids. *Nature Communications*, 11(1), 3602. <https://doi.org/10.1038/S41467-020-17278-2>
- Roberts, M. C. (1996). Tetracycline resistance determinants: mechanisms of action, regulation of expression, genetic mobility, and distribution. *FEMS Microbiology Reviews*, 19(1), 1–24. <https://doi.org/10.1111/J.1574-6976.1996.TB00251.X>
- Rodríguez-Beltrán, J., Sørum, V., Toll-Riera, M., de la Vega, C., Peña-Miller, R., & Millán, Á. S. (2020). Genetic dominance governs the evolution and spread of mobile genetic

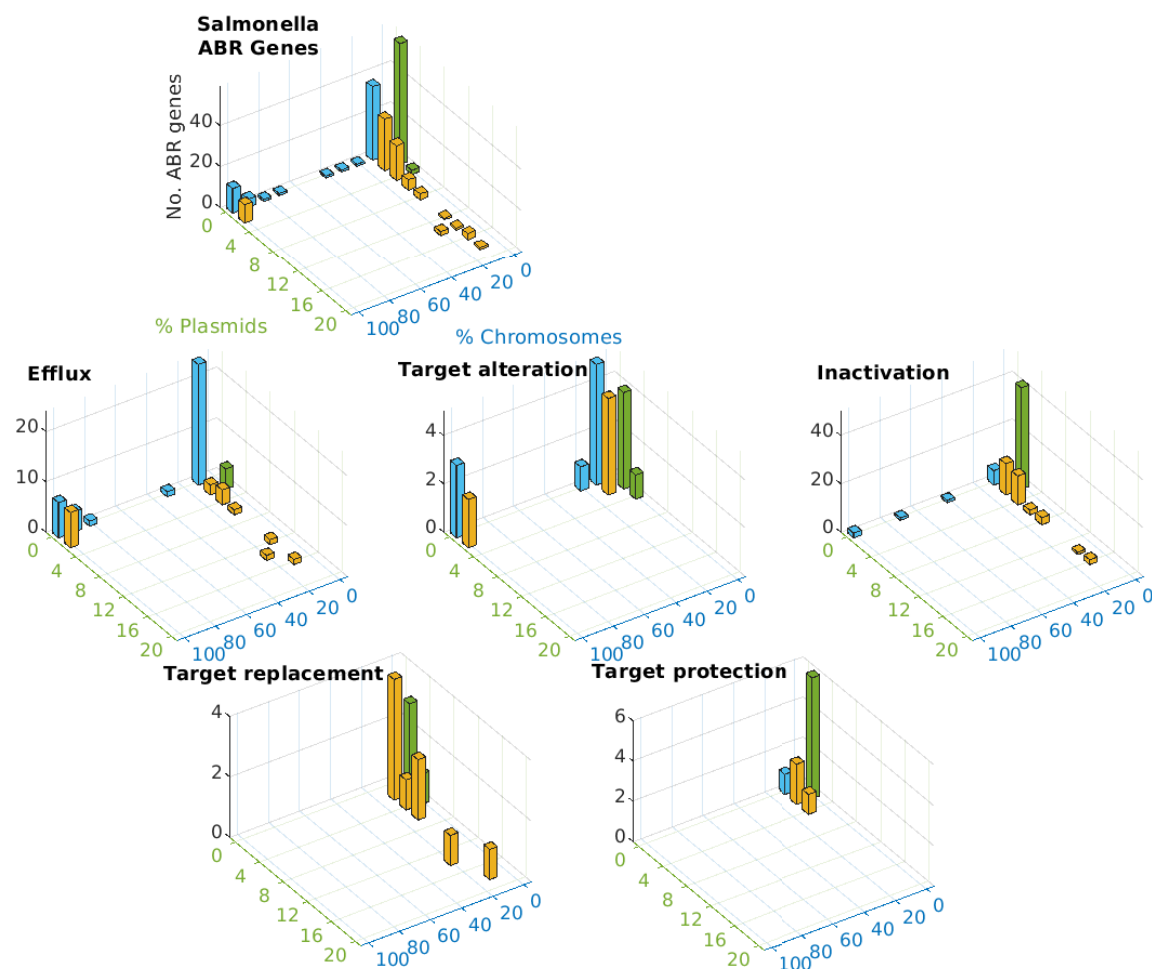
- elements in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 117(27), 15755–15762. <https://doi.org/10.1073/PNAS.2001240117>
- Rojas, A., Duque, E., Mosqueda, G., Golden, G., Hurtado, A., Ramos, J. L., & Segura, A. (2001). Three efflux pumps are required to provide efficient tolerance to toluene in *Pseudomonas putida* DOT-T1E. *Journal of Bacteriology*, 183(13), 3967–3973. <https://doi.org/10.1128/JB.183.13.3967-3973.2001>
- Rousset, F., Cabezas-Caballero, J., Piastra-Facon, F., Fernández-Rodríguez, J., Clermont, O., Denamur, E., Rocha, E. P. C., & Bikard, D. (2021). The impact of genetic diversity on gene essentiality within the *Escherichia coli* species. *Nature Microbiology*, 6(3), 301–312. <https://doi.org/10.1038/S41564-020-00839-Y>
- Ruiz, C., & Levy, S. B. (2014). Regulation of *acrAB* expression by cellular metabolites in *Escherichia coli*. *Journal of Antimicrobial Chemotherapy*, 69(2), 390–399. <https://doi.org/10.1093/JAC/DKT352>
- Ruiz, J. (2003). Mechanisms of resistance to quinolones: Target alterations, decreased accumulation and DNA gyrase protection. *Journal of Antimicrobial Chemotherapy*, 51(5), 1109–1117. <https://doi.org/10.1093/JAC/DKG222>
- Salverda, M. L. M., de Visser, J. A. G. M., & Barlow, M. (2010). Natural evolution of TEM-1 β -lactamase: Experimental reconstruction and clinical relevance. *FEMS Microbiology Reviews*, 34(6), 1015–1036. <https://doi.org/10.1111/J.1574-6976.2010.00222.X>
- Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, 30(14), 2068–2069. <https://doi.org/10.1093/BIOINFORMATICS/BTU153>
- Skold, O. (1976). R factor mediated resistance to sulfonamides by a plasmid borne, drug resistant dihydropteroate synthase. *Antimicrobial Agents and Chemotherapy*, 9(1), 49–54. <https://doi.org/10.1128/AAC.9.1.49>
- Sköld, O. (2000). Sulfonamide resistance: Mechanisms and trends. *Drug Resistance Updates*, 3(3), 155–160. <https://doi.org/10.1054/DRUP.2000.0146>
- Socha, R. D., Chen, J., & Tokuriki, N. (2019). The Molecular Mechanisms Underlying Hidden Phenotypic Variation among Metallo- β -Lactamases. *Journal of Molecular Biology*, 431(6), 1172–1185. <https://doi.org/10.1016/J.JMB.2019.01.041>
- Sorek, R., Zhu, Y., Creevey, C. J., Francino, M. P., Bork, P., & Rubin, E. M. (2007). Genome-wide experimental determination of barriers to horizontal gene transfer. *Science*, 318(5855), 1449–1452. <https://doi.org/10.1126/SCIENCE.1147112>
- Steinegger, M., & Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, 35(11), 1026–1028. <https://doi.org/10.1038/NBT.3988>

- Swedberg, G., & Skold, O. (1983). Plasmid-borne sulfonamide resistance determinants studied by restriction enzyme analysis. *Journal of Bacteriology*, 153(3), 1228–1237. <https://doi.org/10.1128/JB.153.3.1228-1237.1983>
- Turner, P. E., Williams, E. S. C. P., Okeke, C., Cooper, V. S., Duffy, S., & Wertz, J. E. (2014). Antibiotic resistance correlates with transmission in plasmid evolution. *Evolution*, 68(12), 3368–3380. <https://doi.org/10.1111/EVO.12537>
- Varga, M., Kuntová, L., Pantůček, R., Mašlaňová, I., Růžicková, V., & Doškař, J. (2012). Efficient transfer of antibiotic resistance plasmids by transduction within methicillin-resistant *Staphylococcus aureus* USA300 clone. *FEMS Microbiology Letters*, 332(2), 146–152. <https://doi.org/10.1111/J.1574-6968.2012.02589.X>
- Warinner, C., Rodrigues, J. F. M., Vyas, R., Trachsel, C., Shved, N., Grossmann, J., Radini, A., Hancock, Y., Tito, R. Y., Fiddyment, S., Speller, C., Hendy, J., Charlton, S., Luder, H. U., Salazar-García, D. C., Eppler, E., Seiler, R., Hansen, L. H., Castruita, J. A. S., ... Cappellini, E. (2014). Pathogens and host immunity in the ancient human oral cavity. *Nature Genetics*, 46(4), 336–344. <https://doi.org/10.1038/NG.2906>
- Wein, T., & Dagan, T. (2020). Plasmid evolution. *Current Biology*, 30(19), R1158–R1163. <https://doi.org/10.1016/J.CUB.2020.07.003>
- Wein, T., Hülter, N. F., Mizrahi, I., & Dagan, T. (2019). Emergence of plasmid stability under non-selective conditions maintains antibiotic resistance. *Nature Communications*, 10(1), 2695. <https://doi.org/10.1038/S41467-019-10600-7>
- Wein, T., Wang, Y., Barz, M., Stücker, F. T., Hammerschmidt, K., & Dagan, T. (2021). Essential gene acquisition destabilizes plasmid inheritance. *PLoS Genetics*, 17(7), e1009656. <https://doi.org/10.1371/JOURNAL.PGEN.1009656>
- Wein, T., Wang, Y., Hülter, N. F., Hammerschmidt, K., & Dagan, T. (2020). Antibiotics Interfere with the Evolution of Plasmid Stability. *Current Biology*, 30(19), 3841–3847.e4. <https://doi.org/10.1016/J.CUB.2020.07.019>
- Werren, J. H. (2011). Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proceedings of the National Academy of Sciences of the United States of America*, 108(SUPPL. 2), 10863–10870. <https://doi.org/10.1073/PNAS.1102343108>
- Wilson, D. N., Hauryliuk, V., Atkinson, G. C., & O'Neill, A. J. (2020). Target protection as a key antibiotic resistance mechanism. *Nature Reviews Microbiology*, 18(11), 637–648. <https://doi.org/10.1038/S41579-020-0386-Z>

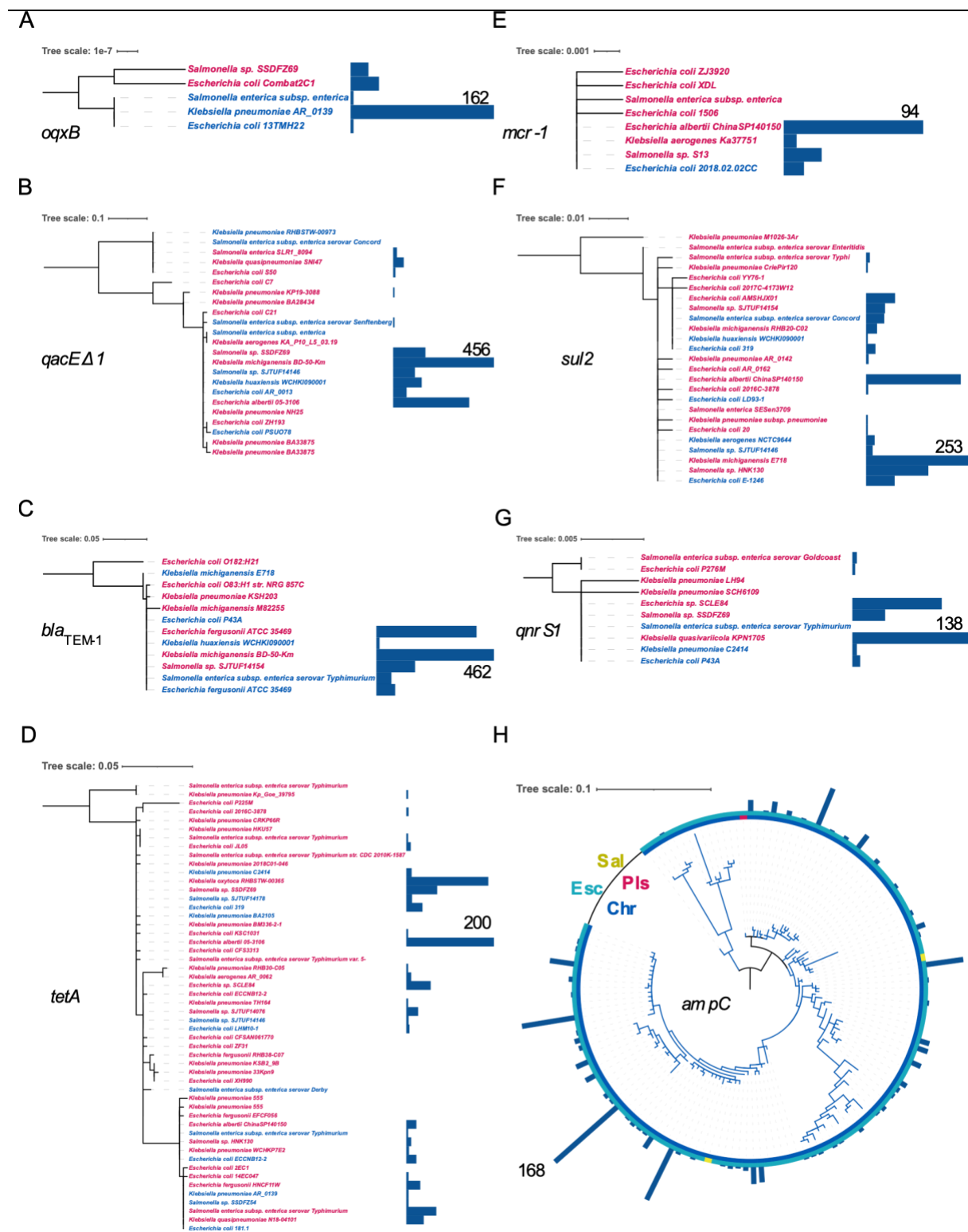
Supporting information



Supplementary Figure S1. A comparison of gene prevalence in plasmids and chromosomes in *Klebsiella* isolates. The distribution of all ABR gene homologs in the examined isolates is shown on the top left. The distribution of ABR gene homologs according to their classification into ABR mechanism is shown below (data supplied in Table S2).



Supplementary Figure S2. A comparison of gene prevalence in plasmids and chromosomes in *Salmonella* isolates. The distribution of all ABR gene homologs in the examined isolates is shown on the top left. The distribution of ABR gene homologs according to their classification into ABR mechanism is shown below (data supplied in Table S2).



Supplementary Figure S3. Phylogenetic reconstruction of 8 ABR genes. ABR gene name is written next to the phylogenetic tree. Bars next to the operational taxonomic units show the frequency of redundant amino acid sequences, the highest frequency is noted. A-G) Isolate label is colored according to replicon type (plasmid genes in pink and chromosomal genes in blue). H) The outer ring shows the genus of ABR gene origin, the inner ring shows whether the gene is encoded on a plasmid or chromosome.

Chapter III - Transposable elements drive the evolution of multiple drug resistance plasmids

Yiqing Wang, Tal Dagan*

Institute of General Microbiology, Kiel University, Kiel, Germany

*Corresponding author

Manuscript submitted.

Abstract

Multidrug resistant (MDR) pathogens are a major threat to human health. The emergence of MDR bacteria is often facilitated by acquisition of plasmids encoding multiple antibiotic resistance genes (ARGs), which are widespread in clinical environments. However, the genetic mechanisms involved in the accumulation of diverse resistance genes in plasmid genomes remain elusive. Here we show that 80% of the ARGs in multidrug resistance plasmids are encoded in transposable ARG islands (termed here *tARGis*). Analyzing the antibiotic resistance accessory genome in 6,784 plasmids from *Escherichia*, *Salmonella*, and *Klebsiella*, we find frequent coincidence of ARGs in diverse combinations. The majority of co-occurring ARGs are densely clustered in genomic islands that correspond to transposable elements. Our results suggest frequent transposition of *tARGis* among large plasmids but rarely into small plasmids, which appear evolutionary diverged. The evolutionary history of *tARGis* is distinct from their vehicle plasmids, implicating transposable elements as the main drivers of antibiotic resistance gene transfer.

Introduction

Plasmids are autonomously replicating genetic elements that reside in prokaryotic populations. The mobility of plasmids between microbial populations renders them a potent driver of antibiotic gene dispersal. The ability of plasmids to persist in their host population may depend on auxiliary functions encoded in their genome. Accessory plasmid genes often encode functions that mediate the host adaptation to growth-limiting conditions, for example, resistance to antibiotics (Rodríguez-Beltrán et al., 2021). Indeed, the emergence of antibiotic resistant bacteria in clinical environments is often linked to the acquisition of plasmids encoding antibiotic resistance genes (ARGs; e.g. (Kumarasamy et al., 2010; Liu et al., 2016; León-Sampedro et al., 2021)). Plasmids that encode diverse ARGs, termed multiple drug resistance (MDR) plasmids, provide their host a combination of resistances to different antibiotics (Nikaido, 2009). Pathogens carrying MDR plasmids are one of the major threats to human health, including pathogenic *E. coli* and *K. pneumoniae* strains. An epidemiological study performed in 2019 estimated that the contribution of bacterial antibiotic resistance to human mortality was as large as other major diseases such as HIV or malaria (Murray et al., 2022).

The evolution of antibiotic resistant bacteria is often facilitated by diverse mobile genetic elements (MGEs); in addition to plasmid replicons, this includes transposable elements (TEs), integrons as well as integrative and conjugative elements (Partridge et al., 2018). Transposable elements are the most abundant mobile genetic elements in bacterial genomes (Khedkar et al., 2022). Transposition events are a major source of genetic diversity in bacterial evolution as they induce novel mutations and generate genomic structure variation. Transposable elements further contribute to translocation and mobilization of antibiotic

resistance genes. Example is Tn1549-mediated mobilization of Vancomycin resistance gene between chromosomes and plasmids previously reported in *Enterococcus faecalis* (Paulsen et al., 2003). Transposition and recombination among transposable elements can lead to clustered ARGs in bacterial genomes that are also termed antimicrobial resistance island (Miriagou et al., 2006) or multi-resistance regions (MRRs) (Partridge, 2011). Serial acquisition of resistance genes may drive genome expansion in MDR plasmids, which can encode up to ca. 30 ARGs (Wein et al., 2020). Nonetheless, the evolution of diverse ARG combinations in plasmid genomes and the genetic mechanisms at play remain understudied.

Here we study the evolution of antibiotic resistance gene content in plasmids. The co-occurrence of accessory genes in bacterial genome has been suggested to evolve under selection for specific gene combinations (Whelan et al., 2021). Indeed, accessory antibiotic resistance genes in plasmids are expected to evolve under positive selection. Hence, antibiotic resistance genes that co-occur in plasmid genomes may be related to common selective pressures and concurrent acquisition or loss of ARGs. Using a large-scale analysis of bacterial genomes, we examine the repertoire of ARG co-occurrence in plasmids, with a focus on three genera: *Klebsiella*, *Escherichia* and *Salmonella* (KES). Plasmids in KES strains share gene content, suggesting low barriers for plasmid transfer among hosts in the three genera as well as common plasmid origins (Redondo-Salvo et al., 2020). To explore the evolutionary history of MDR plasmids, we examine the genetic background of co-occurring ARGs while considering neighboring transposases, plasmid type and size as well as plasmid mobility.

Results

ARG coincidence is more frequent in plasmids in comparison to chromosomes

For the purpose of our study, we clustered 11,995,860 protein sequences in chromosomes and plasmids present in 2,441 KES isolates into 32,623 clusters of homologous protein families. Using the comprehensive antibiotic resistance database (CARD) (Alcock et al., 2020), we identified 114,457 homologs of 397 ARGs in 138 protein families. Members of those 138 families not identified by CARD as ARGs were excluded from further analysis. The results show that all 2,441 chromosomes in the dataset encode ARGs where the number of ARGs ranged between 14 and 73. A total of 2,591 (38%) plasmids encode ARG homologs. The number of ARGs per plasmid ranged between 1 and 38 with the majority of plasmids (78%) encoding multiple ARGs.

To examine the pattern of ARG coincidence, we performed a statistical analysis of all possible pairs of ARG protein families, and tested if their co-occurrence in the same replicon (plasmid or chromosome) is significantly more frequent than expected by chance alone. The analysis was conducted separately for plasmids and chromosomes, for each of the three genera. Our analysis of plasmid genomes revealed an average of 30% (ca. 260 pairs)

significantly co-occurring combinations of paired ARGs (termed coARG), with little variation among the genera (Table 1A). The mean proportion of coARGs in chromosomes was only 13% (Table 1A). Furthermore, the proportion of coARG instances (i.e., copies) in chromosomes was much lower in comparison to plasmids (Table 1B). The paucity of coARGs in chromosomes is well explained by a majority of chromosomal ARGs within protein families that are core universal genes (Wang et al., 2022).

Genes encoding antibiotic resistance were classified according to their mechanism of action. Integrating that information with the coARGs shows that the most frequent coARGs in plasmids were among the pairs encoding antibiotic inactivation mechanisms (Fig. 1B). This result is in agreement with the enrichment of antibiotic inactivation mechanisms in plasmid-encoded genes (Rodríguez-Beltrán et al., 2020; Wang et al., 2022). CoARGs encoding antibiotic inactivation mechanisms were frequently observed also in chromosomes (Supplementary Fig. S1). Additionally, we observed frequent coARGs among the mechanism pairs of antibiotic inactivation-efflux and inactivation-target replacement both in plasmids (Fig. 1B) and in chromosomes (Supplementary Fig. S1). Genes encoding inactivation mechanisms comprise mostly accessory genes, and their distribution is biased towards plasmids. The most frequent coARG were included antibiotic inactivation genes. In contrast, despite the high proportion (57%) of antibiotic efflux genes in our data, coARGs encoding efflux mechanisms were sparse in both plasmids and chromosomes (Fig. 1B, Supplementary Fig. S1). Genes encoding antibiotic efflux are mainly core chromosomal genes. For example, chromosomes in *Escherichia* encode, on average, 38 homologs of 53 efflux ARGs. Indeed, frequent co-occurrence of efflux genes was observed only in *Escherichia* chromosomes (Supplementary Fig. S1). Our results thus demonstrate extensive ARG coincidences in enterobacterial plasmids, most of which are in combination with inactivation mechanisms.

Our analysis revealed several combinations of ARG families whose co-occurrence was significantly less frequent than expected by chance alone. This included, ca. 2% of the ARG pairs in chromosomes and ca. 4% of the ARG pairs in plasmids (Supplementary Fig. S2). Here we highlight several examples of these antagonistic ARG pairs, which are found on plasmids in all three genera. The first pair includes *ampC* and *bla*_{CTX-M-15}, which encode antibiotic inactivation mechanisms that confer resistance against drug class of cephalosporin and penicillin. *ampC* is encoded universally in *Escherichia* chromosomes and *bla*_{CTX-M-15} is encoded mainly on plasmids in all three genera. The rare co-occurrence of the two genes may be due to barriers to transfer of chromosome core genes to plasmids (Wang et al., 2022). Four additional ARG pairs in *Escherichia* and *Klebsiella* plasmids include combinations of *bla*_{NDM-1} or *bla*_{OXA-48}, and *bla*_{TEM-1} or *bla*_{CTX-M-15} (Supplementary Table S2). None of these combinations was, however, antagonistic at the isolate level. Taken together, antagonistic ARG pairs are rare in our data. Several of the ARG pairs we detected are functional orthologs that are not

encoded in the same plasmid but co-occur in the same isolate. Our results thus suggest that infrequent ARG co-occurrence is the result of barriers to gene transfer rather than conflicts among ARG mechanisms.

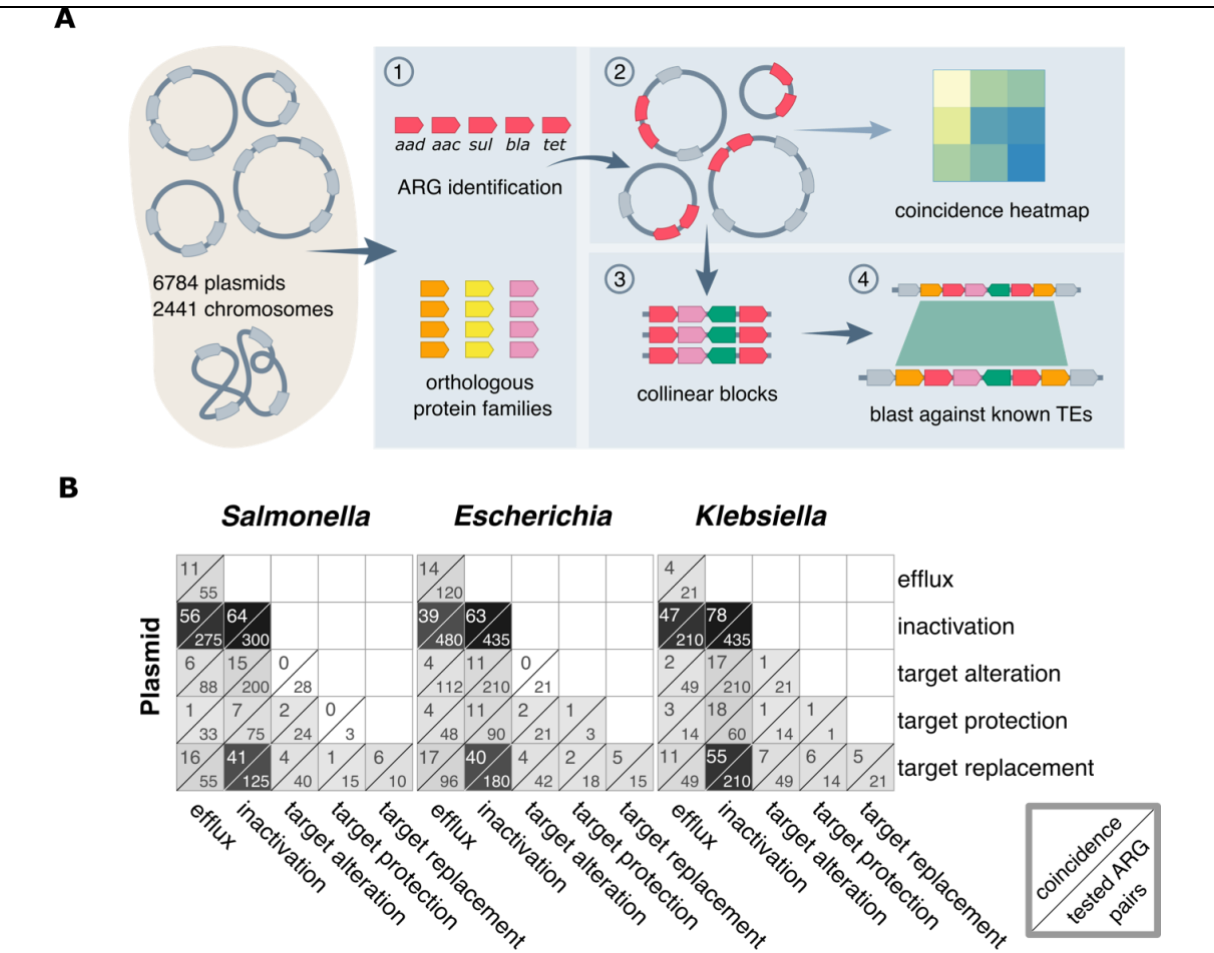


Figure 1. The pattern of antibiotic resistance gene co-occurrence in plasmid genomes. A) An illustration of our workflow for detecting ARG coincidences in transposable elements (TEs). Arrow boxes correspond to protein encoding sequences. (1) Proteins in bacterial genomes were clustered into protein families and ARGs were identified. (2) Coincidence tests for ARG pairs were conducted per replicon type in each genus. (3) Collinear syntenic blocks (CSBs) including coARGs were identified across genomes (4) and further compared to intact TE sequences. B) CoARGs on plasmids in *Salmonella*, *Escherichia*, and *Klebsiella*. The significant co-occurring combinations of paired ARGs (termed coARG) is summarized according to the resistance mechanisms. The numbers of coARGs are shown as nominators, and numbers of total tested ARG pairs are shown as denominators in the heatmap. The three mechanism pairs that have the highest number of coARGs are highlighted in black.

Co-occurring ARGs are conserved across plasmids and coincide with transposable elements

To further study the evolution of coARGs we examined whether the coARGs are encoded in the same genomic neighborhood and if their order is conserved across multiple genomes. To answer these two questions, we identified collinear syntenic blocks (CSBs) across all genomes and retained the CSBs that include coARGs for further analysis (Fig. 1A). Our survey revealed

that coARGs in plasmids are more frequently found within CSBs in comparison to chromosomes (Table 1A). Similarly, the proportion of conserved coARG instances is higher in plasmids in comparison to chromosomes (Table 1B). Manual inspection of the CSBs frequently revealed transposable elements in the neighborhood of coARGs, which is in agreement with previous reports of ARG mobilization by TEs (Paulsen et al., 2003). Consequently, we quantified the frequency of coARG containing CSBs that correspond to TEs (Fig. 1A). We found that ABR plasmids encode significantly more transposases in comparison to non-ABR plasmids (using Wilcoxon rank sum test, $p\text{-value}=4.1\times 10^{-15}$, $n_{\text{ABR plasmids}}=2,591$, $n_{\text{non-ABR plasmids}}=4,193$), where the median number of transposases in ABR plasmids is 8, and 6 in non-ABR plasmids. The numbers of transposases and ARGs encoded in plasmids are weakly correlated, but significant so ($r_s=0.39$, $p\text{-value} < 2.2\times 10^{-16}$, using Spearman correlation). The mapping of transposable elements in all genomes revealed that almost all the CSBs containing coARGs in plasmids correspond to TEs, compared to only 37% in chromosomes (Table 1B). CSBs comprising a combination of coARGs and TEs constituted 38% of the MDR plasmid gene content and 80% of the total ARGs in MDR plasmids. In comparison, transposases account for only 11% of the genes in plasmid CSBs, yet they represent 25% of the transposases in MDR plasmids. Taken together, we found that coARGs in plasmids are often organized in conserved genomic regions that correspond to transposable elements. Our results demonstrate that TEs play a major role in the evolution of MDR plasmids and their accessory gene content.

Previous studies suggested that multiple transposition events may occur at specific genomic locus. For example, IS26 elements were shown to translocate with the flanking region and insert adjacently to another IS26 in the form of translocatable unit (Harmer et al., 2014; Harmer & Hall, 2021). Thus, integration and recombination of TEs may generate novel combinations of antibiotic resistance cargo genes. Indeed, our analysis further revealed that the coARGs corresponding to TEs are heterogeneous and furthermore correspond to multiple ARGs (rather than a pair of genes) that co-occur in the same genomic locus. For example, *aph(6)-Id*, *aph(3'')-Ib*, *bla_{TEM-1}*, *tetA* and *sul2* (Supplementary Fig. S3). These five ARGs are often found together within a region which corresponds to TEs in the plasmids. Our observations thus indicate that transposable elements may generate ARG islands in MDR plasmids.

Table 1. Frequency of cooccurring ARGs in replicons, syntenic blocks and transposable elements.

A						
Replicon	Plasmid			Chromosome		
Genus	<i>Salmonella</i>	<i>Escherichia</i>	<i>Klebsiella</i>	<i>Salmonella</i>	<i>Escherichia</i>	<i>Klebsiella</i>
coARG	257 (36%)	251 (28%)	267 (32%)	231 (10%)	561 (12%)	337 (18%)
CSB	188 (73%)	213 (85%)	249 (93%)	130 (56%)	164 (29%)	117 (35%)
TE	188 (100%)	201 (94%)	249 (100%)	117 (90%)	158 (96%)	116 (99%)
fcTE	83 (44%)	90 (45%)	93 (37%)	67 (57%)	58 (37%)	51 (44%)
B						
Replicon	Plasmid			Chromosome		
Genus	<i>Salmonella</i>	<i>Escherichia</i>	<i>Klebsiella</i>	<i>Salmonella</i>	<i>Escherichia</i>	<i>Klebsiella</i>
coARG	8,085 (68%)	15,557 (75%)	16,398 (70%)	14,556 (5%)	244,070 (15%)	52,596 (21%)
CSB	3,983 (49%)	7,616 (49%)	9,594 (59%)	4,845 (33%)	32,714 (13%)	4,164 (8%)
TE	3,977 (100%)	7,585 (100%)	9,586 (100%)	1,408 (29%)	10,998 (34%)	2,983 (72%)
fcTE	1,785 (45%)	3,614 (48%)	4,209 (44%)	657 (47%)	1,799 (16%)	538 (18%)

A) All ARG combinations were tested in plasmids and chromosomes in each genus separately.
 B) All ARG instances, i.e., the sum of all ARG combinations in all replicons. Lines in the table correspond to steps 2-4 in the analysis workflow (described in Fig. 1A). CoARG: the number of ARG combinations that had significant co-occurrence and their proportion of all the tested combinations of ARG pairs. CSB: the number of coARGs that were found in any collinear syntenic block (CSB) and their proportion from the coARGs. TE: the number of CSBs that partially match with any known transposable element (TE) sequences and their proportion from the CSB. fcTE: the number of CSBs fully matched with any TE sequences and their proportion from TE.

TE-driven ARG coincidence form compact resistance island in plasmids

The pattern of coARGs in transposable elements suggests that ARGs are not randomly distributed within the genome, rather they form islands of multiple ARGs. To test for the presence of ARG islands, we examined the distance between consecutive ARGs by counting the number of protein-coding genes between two, and up to ten ARGs. If the ARGs are randomly distributed in the genome, the distance distribution should fit to a negative binomial distribution. Our results show that the distance between consecutive multiple ARGs deviates significantly from the random expectation (Goodness-of-fit test, distance distributions were tested separately with $\alpha=0.05$). We further compared the distance between multiple ARGs to a control group of randomly sampled genes. The distance between multiple ARGs was significantly shorter than the distance of the control group (Wilcoxon rank sum test, distance distributions were tested separately, $\alpha=0.05$). Notably, the median distance between two adjacent ARGs in plasmids is zero, which means that over half of the ARGs have an ARG as neighbor gene on the same strand. The median distance between two sampled genes in plasmids is 11, which is between the median distance of six and seven ARGs (10 and 13). ARGs are thus densely encoded in plasmid genomes. Our results show that ARGs are significantly clustered in genomes and form compact ARG islands in plasmids.

Further examination of the CSBs revealed five transposase families that contribute the most to the ARG clustering pattern in plasmid we observed. This includes IS26, IS91, IS110, IS6100 and Tn3 family transposase. These transposons comprise ARGs in all ABR

mechanism categories (Fig. 2B). We found nearly all possible combinations of the five transposase families and the 24 ARGs in CSBs, nevertheless, the frequency of ARG and transposase combinations varies in our data. IS26 is the most abundant transposase family, and has more than 300 instances with 15 different ARGs. Most of the ARG combinations with IS26 appear frequently in the data, examples for highly frequent ARG with IS26 are *bla*_{TEM-1} and *sul2* (Fig. 2B). Most ARGs occur with multiple transposases in the data, several of those with high frequency; examples are *sul1* occur in four and *ermE* occur in three transposase families with high frequency (IS26 and IS26* are in the same family). The variable frequency of transposase-ARG co-occurrence suggests that the association between ARGs and transposases is rather transient. To test if the abundance is associated with the antiquity of specific TE-ARG combinations, we examined the isolation date of isolates where those TE-ARG co-occurred. Indeed, this approach depends on the sampling density, nonetheless, it has the potential to reveal temporal patterns in the data. Our results reveal that the most frequent combinations also have earlier first isolation date, e.g., Tn3 and *ermE* (first documented in 1963) IS91 and *sul2* (first documented in 1950). Similarly, the least frequent combinations were more recently documented in isolates, e.g., IS110 and *arr-2*, *aph(4)-la* (first report in 2012, 2014, Fig. 2B). Nevertheless, there are several exceptions, where combinations that have high frequency but were first documented rather recently or the other way around, e.g., *qnrS1* and IS6100 or IS110 with 100 or 67 instances since 2010, *dfrA12* and IS110 with 53 instances since 1981, or Tn3 and *catI* with 102 instances since 1963 (Fig. 2B). These exceptions could be explained by either limited spread through the population or by sampling bias in the 1950s-1970s (Fig. 2B, Supplementary Fig. S4).

To further study the evolution of transposable ARG islands (termed tARGis), we examined the evolutionary history of the associated transposases in the major TE families. A previous study suggested that TE evolution is characterized by frequent transfer and periodic extinctions (Wagner, 2006). Here we examine that suggestion using the phylogenetic reconstruction of two DDE transposase families of IS26 and Tn3. In our data, IS26 contains mainly two variants that comprise 6,266 or 1,385 identical protein sequences. The two variants are often encoded in the same replicon, and correspond to IS26 and IS26* variants (later corrected to IS26-v1 (Harmer et al., 2021)). IS26* has a single G184N replacement, which was reported to increase the cointegration efficiency of IS26 (Pong et al., 2019). IS26 is the most abundant transposase and also the most conserved protein family among the five transposase families. The two dominant IS26 sequence variants comprise 92% of the whole protein family and the remaining variants have few substitutions. Due to the extreme sequence conservation of IS26, it is practically impossible to reconstruct the evolution of this protein family. The high frequency of identical IS26 homologs may correspond to either a very recently emerged variant or a highly conserved variant in the population. To gain further insight on the antiquity of the

two main variants, we examined the isolation date of the isolates encoding those variants. The distribution along the isolation timeline as reflected in our data shows that IS26 was reported first in 1954 and IS26* was reported first in 1990 (Fig. 2C). The higher frequency of ARG combinations with IS26 in comparison to combinations with IS26* (Fig. 2B) is in agreement with an earlier origin of IS26 in the three genera *KES*. The positive trend of increasing frequency in time for both IS26 and IS26* suggests that both variants are highly mobile and stable. Previous studies suggested that TEs are mostly transferred by conjugative plasmids (Che et al., 2021). Indeed, the majority (70%) of the IS26 and IS26* are encoded in conjugative plasmids, with sparse presence also in mobilizable as well as non-mobilizable plasmids (Fig. 2C).

The phylogeny of Tn3 transposase family shows four dominant homologs: Tn5403, TnAs1, TnAs2, and TnAs3, in addition to 164 low frequency variants. Tn3 has the highest diversity among the five transposases, where the four dominant variants comprise only 48% of the total family. The distribution of isolate date shows that TnAs1, TnAs2, and TnAs3 were first reported in the early dates between 1963 and 1981, and Tn5403 was reported later in 1990 (Fig. 2D). We observed a similar pattern of correspondence between variant frequency and antiquity in other three transposase protein families (Supplementary Fig. S5). In IS6100 DDE transposase family a one dominant variant is documented between 1981 and 2020. The IS110 DEDD transposase family comprises four dominant variants, while the IS91 HUH transposase family includes two dominant variants, ISCR1 and ISCR2, which are documented between 1965 and 2020. In our data, IS26, Tn3, IS6100, IS110, and IS91 transposase families are more frequently encoded in plasmids, where 83% (13,080 out of 15,761) of the protein homologs are encoded in plasmids. This bias may supply an explanation for the bias of coARGs towards plasmids. The uneven distribution of the five transposase families also indicates that the gene transfer events mediated by those five transposases are mainly between plasmids, and the LGT from plasmid to chromosome could be restricted.

To further examine evidence for tARGis transfer into bacterial chromosomes, we searched for plasmid CSB containing co-occurrence of TnpA and ARGs (as in Fig. 2B) in chromosomes of isolates with or without plasmid in the same host cell. The pattern of such CSBs in chromosomes was similar to what we observed in plasmid genomes, albeit with significantly fewer instances (Fig. S6). The low frequency of TEs carrying ARGs in chromosome is likely due to the barriers to LGT from plasmids to chromosomes. The pattern of TEs frequency in chromosomes residing with and without plasmids in the same cell was similar. Hence, TEs carrying ARGs may transfer to - and persist in - chromosomes, regardless of whether their plasmid vehicle remains as resident in the host cell.

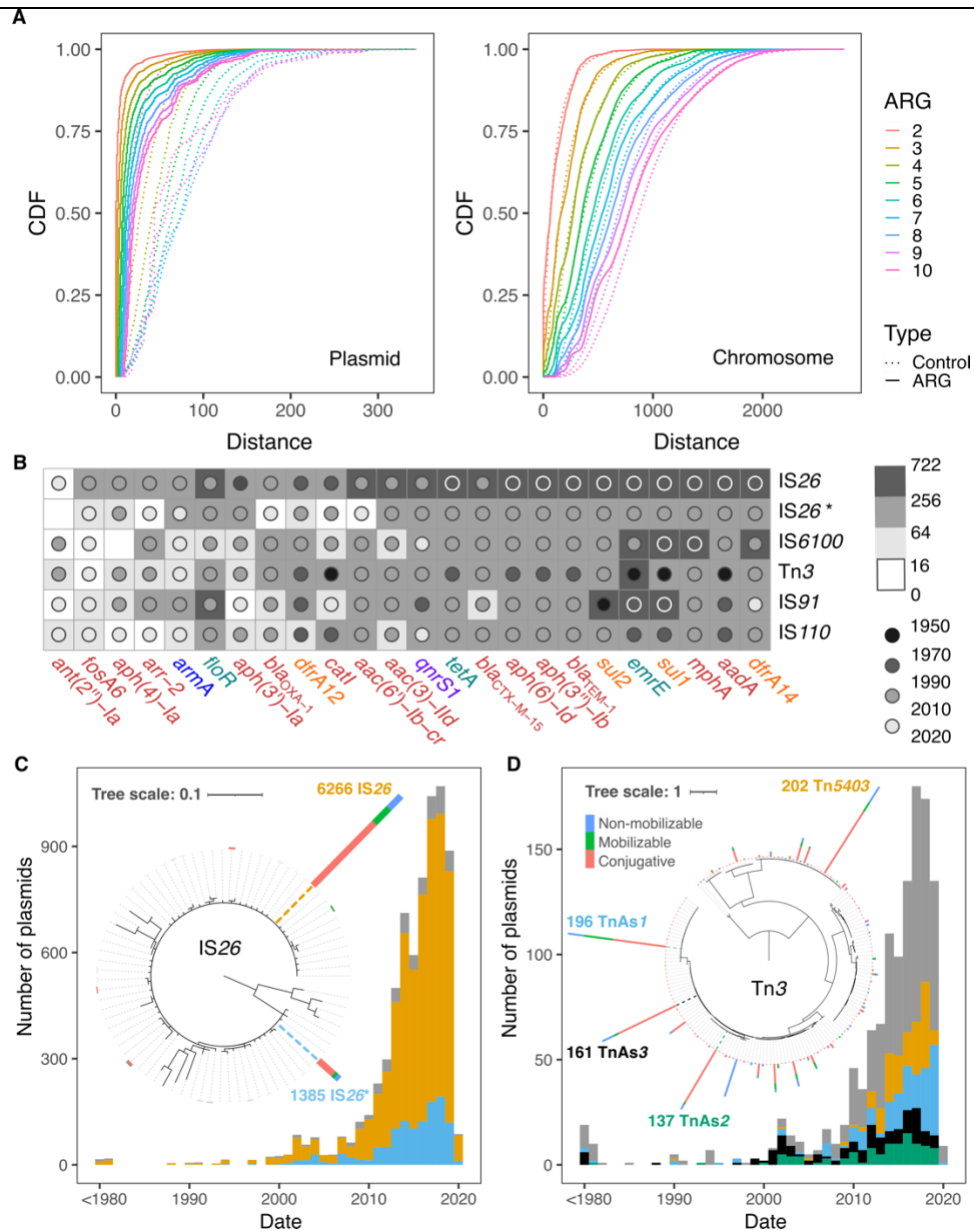


Figure 2. Characteristics and evolution of TE-driven ARG islands in plasmids. **A)** Cumulative Distribution Function (CDF) plots show that the distance between ARGs is shorter than the distance between a control group of randomly sampled genes in plasmids (left), but not in chromosomes (right). Intergenic distances were calculated between two up to ten adjacent genes. **B)** Co-occurrence of 24 ARGs and the common transposase families in plasmid CSBs. Color gradient of the square cells corresponds to the number of ARG-transposase co-occurrence in CSBs in unique plasmids. Color gradient of the circles in cells corresponds to the earliest isolation date where the CSB containing ARG-transposase combination was observed. ARG names are colored according to the antibiotic resistance mechanism (red: antibiotic inactivation, green: antibiotic efflux, orange: antibiotic target replacement, blue: antibiotic target alteration, and purple: antibiotic target protection). **C)** Phylogenetic tree of IS26. Stacked bars at the tree leaves show the number of two dominant IS26 variants, the bar is colored according to the mobility type of the variant carrying plasmid. Histogram below the phylogeny shows the isolation dates of plasmids carrying IS26 variants in different shades. The two dominant variants exist in the population over decades. IS26* may be younger than IS26. **D)** Phylogenetic tree of Tn3. Four dominant variants in Tn3 transposase protein family are highlighted in the phylogeny and distribution of isolation dates as in (C).

Describing ARG coincidence requires a distinction between small and large plasmid types

Previous studies suggested a major role of conjugative plasmids in ARG dissemination (Che et al., 2021). Our results are largely in agreement with those observations as most of the ARG-TEs are present in conjugative plasmids (e.g., Fig. 2CD). Nonetheless, ARG-TEs may occur also in mobilizable and non-mobile plasmids. A common property of MDR plasmids is their large genome size, that is positively associated with the number of ARGs encoded in the plasmid (Wein et al., 2020). Since plasmid mobility type and plasmid size are associated (i.e., conjugative plasmids are typically large (Smillie et al., 2010)), disentangling the plasmid properties associated with the evolution of MDR plasmids is challenging. If the positive association between plasmid size and number of ARGs is trivial, i.e., ARG acquisition leads to plasmid genome expansion, then large and small MDR plasmids are expected to be related (at least in some cases). In other words, large and small plasmids may have a shared gene content, with large plasmids harboring more ARGs in comparison to small plasmids. To test this hypothesis, one has to first find an objective approach to classify our plasmids into ‘large’ and ‘small’ plasmid groups. The distribution of plasmid size has been frequently described as a bi-modal distribution (Smillie et al., 2010). A similar bi-modal plasmid size distribution is observed in our data, with a minimal density at plasmid size of ca. 19kbp (Supplementary Fig. S7). Using the minimum density value as a threshold, we compared the gene content between small plasmids having a genome size <19 kbp and large plasmids with genome size \geq 19 kbp. Our results reveal that plasmids in the two size categories rarely share gene families. The most shared protein families are NikA relaxase (in 4% small and 21% large plasmids), Tnp26 (IS26 transposase; in 2% small and 47% large plasmids) and a Tn3 resolvase (in 2% small and 19% large plasmids). The absence of gene sharing between small and large plasmids indicates that small and large plasmids are not closely evolutionary related and furthermore shows that there is rarely gene flow between the two groups. Previous studies suggested that small plasmids often co-reside with large plasmids that may also contribute to their mobility (San Millan et al., 2013; Barry et al., 2019). To test if gene flow is possible in principle between small and large *KES* plasmids, we tested for co-residence of small and large plasmids in our data. We find that 95% (1861 out of 1961) small plasmids reside in the same isolate with at least one large plasmid in all three genera. Taken together, our results show that gene flow between large and small plasmids is limited and restricted to only rare transposition events.

The pattern of shared genes between small and large plasmids reveals two protein families that are highly frequent in plasmids depending on their size: *rop* (*rom*) and *repB* (Fig. 4). These two proteins are part of two fundamentally different plasmid control or replication mechanisms (del Solar et al., 1998). The Rop protein family includes 903 homologs encoded in 857 plasmids, of which 95% are smaller than 19 kbp. The RepB protein family includes

2,603 homologs encoded in 2,393 plasmids. Only one RepB-encoding plasmid has a genome size smaller than 19 kbp, which we suspect to be a truncated assembly (based on sequence identity to other large plasmids). Taken together, the presence of Rop or RepB in plasmid genomes is well associated with plasmid size, with only few exceptions (Fig. 4; Table S1). The distinct gene content of small and large plasmids indicates that plasmid size is associated with different plasmid replication backbones (i.e., plasmid types, replication mechanisms). Plasmids containing Rop or RepB correspond to 51% (3483/6784) plasmids in the total set, hence additional proteins may be associated with plasmid size (Supplementary Table S4). In the following we compare the distribution of ARGs in small and large plasmids by focusing on Rop-encoding and Rep-encoding plasmids.

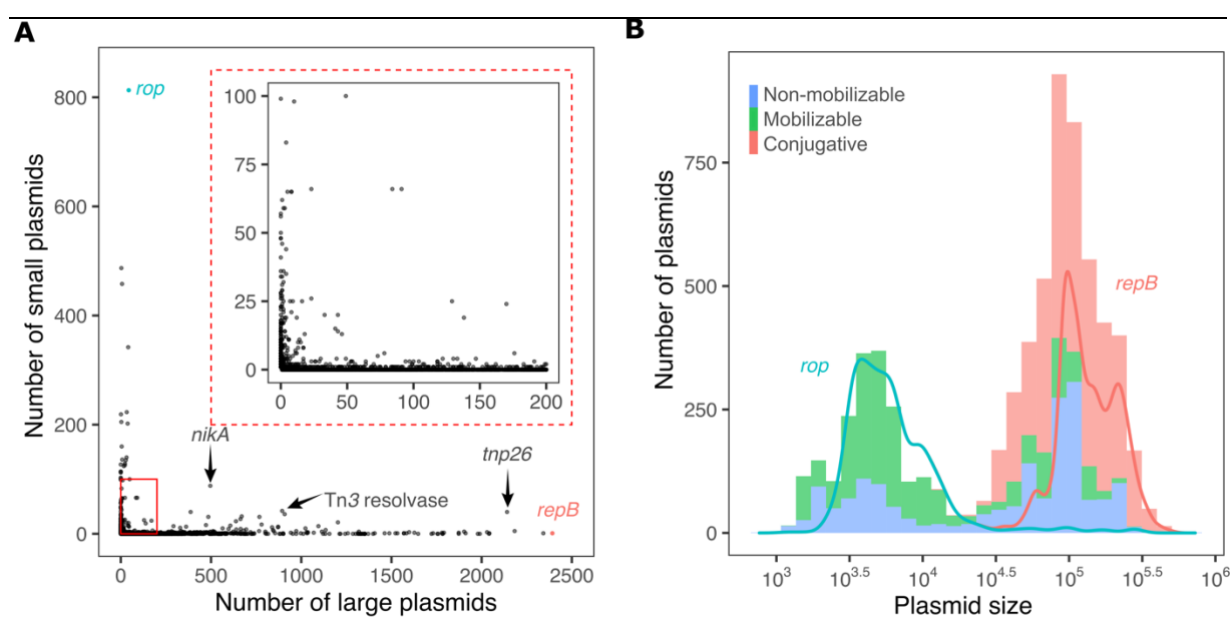


Figure 3. Identification of marker genes of plasmid size. A) Shared genes between small (<19Kb) and large (≥ 19 Kb) plasmids. Each dot corresponds to a protein family with the axes corresponding to the number of small and large plasmids encoding a homolog in the family. Most plasmid protein families comprise homologs either in small or large plasmids and are rarely shared between the two size groups. Rop homologs (in blue) are mostly found in small plasmids and RepB homologs (in red) are exclusively found in large plasmids. The three most frequently shared protein families are pointed out and marked with their names. The distribution of protein families present in less than 100 small plasmids and less than 200 large plasmids is magnified in the inlay figure. B) A histogram of genome size for all plasmids in the dataset colored by plasmid mobility types. The superimposed density curves (solid line) were created from the size distributions of Rop-encoding plasmids (in blue) and RepB-encoding plasmids (in red).

No evidence of TE-driven ARG coincidence in small Rop-plasmids

ARGs are rarely found in small plasmids; of the total 1,961 small plasmids, only 7% harbor ARGs, with 72% of the small plasmids encoding a single ARG. To further investigate the evolution of ARGs in small plasmids we examined the phylogenetic tree of Rop and compared

the properties of closely related plasmids. In our data, Rop proteins are encoded on ColE1-like plasmids (Ares-Arroyo et al., 2021). Their sequence is short and highly conserved, hence

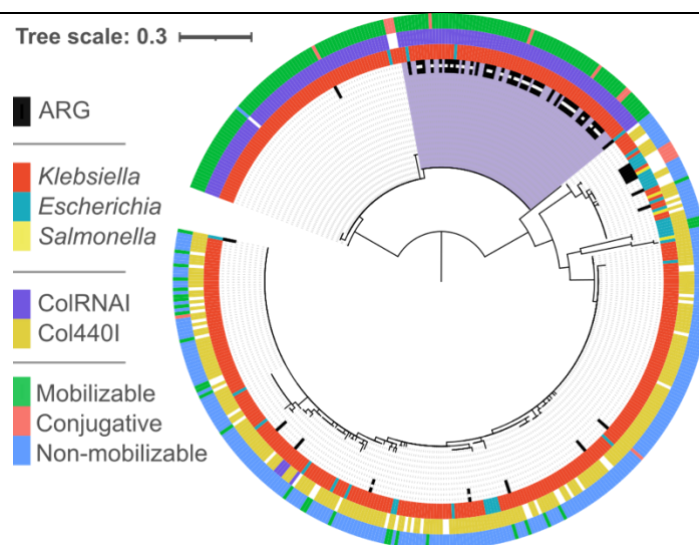


Figure 4. Phylogenetic tree of 469 Rop sequences encoded in 443 plasmids. Rings of the phylogenetic tree show (from inner to outer rings): the presence of ARGs, host genus, plasmid type and plasmid mobility class (see legend). The majority of plasmids reside in *Klebsiella* isolates. The Rop phylogeny constitutes two main Rop clades encoded in non-mobilizable ColRNAI plasmids and mobilizable Col440I plasmids. A subclade of ColRNAI plasmids harboring ARGs is marked in purple, in that clade, small plasmids carrying Tn 1331Δ:IS26 are marked by a white square within the ARG color strip.

our ability to reconstruct a reliable phylogeny using all 903 members of that protein family is limited. Consequently, we split the phylogenetic analysis into two major Rop groups: Rop encoded in mostly non-ABR plasmids in *Escherichia* and *Salmonella* (Supplementary Fig. S8), and Rop encoded in *Klebsiella* plasmids. The *Klebsiella* Rop phylogeny shows a clear split into two clades that correspond to Col440I and ColRNAI plasmids (Fig. 4). The ColRNAI clade includes 21 large plasmids corresponding to co-integrates (of plasmid fusions; Supplementary Table S1) and 27 small ABR plasmids; 24 of the small plasmids in that clade carry one *aac(6')*-*Ib* in the same Tn3-based transposon previously reported as Tn 1331Δ:IS26 (Fig. 4)(Ramirez et al., 2019). Comparing the plasmid sequences reveals that the plasmid backbones of this sub-clade are almost identical to its sister sub-clade, with the only difference between plasmids in the two sub-clades being the presence of Tn 1331Δ:IS26 (Supplementary Fig. S9). The branching pattern suggests a single ancestral transposon insertion event in the clustered ColRNAI plasmids. The ARG-carrying ColRNAI plasmids are reported in isolates from human sampled between 1974 and 2019, suggesting that this plasmid has a stable persistence in *Klebsiella* (Supplementary Table S1). The *Klebsiella* ColRNAI clade further includes three plasmids reported in *Escherichia* isolates, which likely correspond to plasmid transfer from *Klebsiella* to *Escherichia*. The larger plasmids in the ColRNAI clade are putatively mobilizable or conjugative; these plasmids are untypically large for Rop-encoding plasmids (see also Fig.

3B) and likely originated from plasmid fusion events (Supplementary Table S1). Taken together, our analysis shows that ARGs are depleted in small plasmids with only few exceptions.

TE-driven ARG coincidence is restricted to large plasmids

Most of the plasmid ARG homologs (98%) in our data are encoded in large plasmids. Among the large plasmids, RepB is the most commonly shared protein family, which is also a signature gene for the protein-dependent plasmid replication mechanism. RepB-encoding plasmids are the major carriers of ARGs. Indeed, 44% (1045/2392) of the RepB plasmids carry 68% (10,490/15,534) of the plasmid ARG homologs. To further investigate the evolution of tARGs in large plasmids, we examined the RepB phylogeny. The phylogeny comprises several RepB groups that correspond to recognized plasmid incompatibility groups and plasmid mobility classes.

Conjugative plasmids are reported as the major carriers of ARG (Che et al., 2021) and this observation is confirmed in the three genera analyzed here. Nonetheless, to test whether conjugative plasmids carry more ARGs than non-conjugative plasmids, we need to disentangle the association between plasmid size and plasmid mobility type. In the previous section we first examined the differences between large and small plasmids, here we further selected a group of RepB encoding large IncFIB plasmids in *Escherichia* (Fig. 5A, shown in purple area). The selected RepB homologs (748) share very high sequence similarity indicating close evolutionary relationships between the plasmids, hence the plasmids may be comparable in the context of ARG content. The proportion of ABR plasmids among the selected plasmids is higher for conjugative plasmids in comparison to non-conjugative plasmids (Fig. 5B). Nonetheless, there is no significant difference in the number of ARGs between the three mobility classes (p-value=0.094, using Kruskal-Wallis rank sum test, $n_{\text{conjugative}}=245$, $n_{\text{mobilizable}}=23$, $n_{\text{non-mobilizable}}=33$). We next tested whether there is an apparent difference in the TE-mediated ARG coincidence among the three plasmid mobility classes. Comparing plasmids with at least two ARGs within the selected group (purple area in Fig. 5), shows that plasmids in the three mobility groups have similar proportions of coARG, coARG in CSBs, and coARG in CSB matched with TEs (Fig. 5B). Taken together, we show that the TE-ARG islands are observed in MDR plasmids of all mobility groups. That is, TEs have a similar effect on MDR plasmid gene content, regardless of plasmid mobility types. One possible explanation for this observation is the evolution of plasmid mobility. The classification of plasmids into the three mobility classes relies on the presence of functional protein related to plasmid transfer, but ignores the history of the plasmids. Note that the mobilizable or non-mobilizable MDR plasmids in the RepB tree are often neighbors of conjugative plasmids. Hence, the MDR plasmids we

compared here may have the same ancestral conjugative plasmid and their ARG gene content is inherited (rather than transferred).

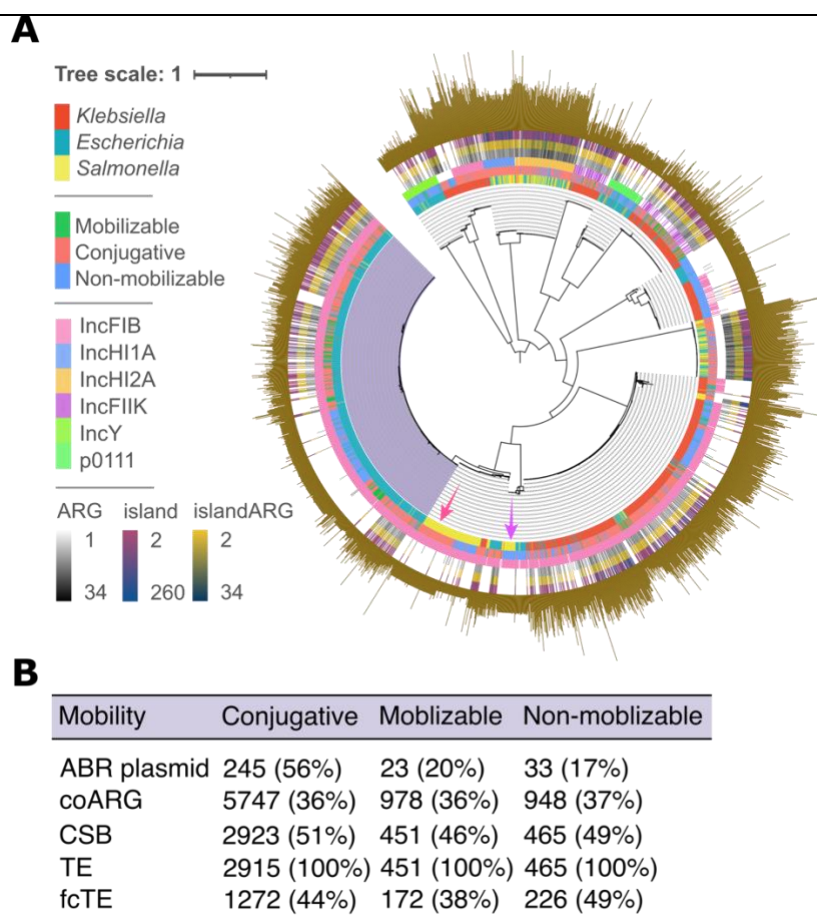


Figure 5. The frequency of tARGs in large ABR plasmids is independent of plasmid mobility. **A)** Phylogenetic tree of 2,529 RepB sequences encoded in 2,354 plasmids. Rings of the phylogenetic tree show (from inner to outer rings): host genus, plasmid mobility class, plasmid type, the frequency of ARGs, the frequency of genes in CBSs containing CoARGs, the frequency of ARGs in conserved TEs (see legend). The outermost ring shows plasmid size in golden bars. A clade of RepFIB plasmids highlighted in purple was selected for TE and ARG comparison between plasmid mobility types. The pink arrow highlights a group of 79 conjugative plasmids where ARGs are rarely found. A purple arrow points out a clade of 38 non-mobilizable plasmids with no ARGs, these plasmids are the only plasmids in their host cell. **B)** A comparison of TE-driven ARG content between three mobility types in IncFIB of plasmids (purple area). Conjugative plasmids carry more often ARGs than the other two types. TE has a similar contribution to ARG coincidence in ARG-carrying plasmid regardless of mobility type. ABR Plasmids: the number of ARG-carrying plasmids and proportion. CoARG: the number of ARG pairs that had significant co-occurrence and their proportion of all the tested ARG pairs. CSB: the number of coARGs that were found in any CSBs and their proportion from the coARG. TE: the number of CSBs that partially match with any known TE sequences and their proportion from the CSB. fcTE: the number of CSBs fully matched with TE sequences and their proportion from TE.

Discussion

The genomes we analyzed here include thousands of isolated from the three genera that span over several decades. Nonetheless, the transposase amino acid sequences within the protein families are identical, even among distantly related plasmids and strains, as well as distant isolation date and geographic location. Previously it has been suggested that the observed low transposase intragenomic diversity is due to the dynamics of their transposition among bacterial hosts. Transfer dynamics that are similar to slow-spreading infectious diseases of cell lineages will comprise periodic extinction of transposable elements, such that only recently acquired elements can be observed in contemporary genomes (Wagner, 2006, 2008). Our results confirm the low intragenomic diversity and further reveal an extreme sequence conservation of transposase genes across the three genera. Combining transposase phylogeny and the timeline of isolate sampling year over 20 years (Fig. 2C; between ca. 2000-2020) reveals the presence of low frequency variants and high frequency dominant variants. In those 20 years that are documented by sequenced genomes, we observe no clear evidence for the periodic extinction model. Alternatively, we propose a constant process of transposase diversification that may be followed by expansion of successful transposase variants, as exemplified by the evolution of IS26*. This recently emerging variant from the ancestral IS26, IS26* has a 10-fold higher cointegration efficiency compared to the ancestral form (Pong et al., 2019). Our results thus suggest that the extreme conservation of transposase sequences is evidence for purifying selection on the transposase function (i.e., the ability to transpose).

Previous studies of plasmid evolution typically compare plasmid gene content on the full set of available plasmids, without much distinction of plasmid origin or properties (Wein et al., 2020; Che et al., 2021). The rarity of shared protein families among small and large plasmids suggests that small and large plasmids are different lineages of plasmid groups, similar to taxonomic groups of bacterial organisms. Thus, a comparison of specific plasmid features between small and large plasmids may lead to an inherent bias of the ensuing conclusions. Example is the comparison of TE frequency between conjugative and non-conjugative plasmids. Here we suggest a practical approach to distinguish between small and large plasmid types. Small plasmids and large plasmids are different in terms of gene repertoire especially ARGs (Fig. 3). The two marker genes for plasmid size, *rop* and *repB*, correspond to two plasmid replication mechanisms – RNA-dependent or protein-dependent. Notably, the two plasmid size classes have limited shared genes, which indicates rare gene transfer events in between those groups. TEs may play a role in plasmid genome expansion, like in bacterial genome expansion (Touchon & Rocha, 2007), yet the analysis of *KES* plasmids, there is no evidence of TE-mediated genome expansion of small plasmids to evolve into large ones. Furthermore, we cannot conclude whether TE was the driving force of large plasmid expansion in the past, due to the lack of evidence for a ‘small’ ancestor of large plasmid. The

origin of large RepB plasmids may be resolved, e.g., by including *repB* plasmid genomes from other genera in the phylogenetic analysis. The differences observed between plasmid size groups imply that comparative plasmid genomics using a mixed plasmid set may reach a biased conclusion, particularly when considering plasmid mobility outside the context of plasmid size. Furthermore, plasmid mobility may be a dynamic feature of plasmid properties (Coluzzi et al., 2022). Studies of plasmid evolution should consider the plasmid intrinsic features, i.e., plasmid size, plasmid incompatibility group and the plasmid backbone.

What are the barriers for TE-mediated transfer of ARG islands? Our results reveal the absence of tARGs in small plasmids and only rare occurrence of TE-ARGs in large non-mobilizable plasmids. Indeed, recently we reported a positive correlation between the number of ARGs encoded in ABR plasmids and plasmid genome size (Wein et al., 2020). Nonetheless, considering the association between plasmid type and size we showed here, required us to revisit that conclusion. The positive association between plasmid size and number of ARGs is significant for large plasmids, albeit with a low correlation coefficient ($r_s=0.53$, $p\text{-value}<2.2\times10^{-16}$, using Spearman correlation). No significant association between plasmid size and number of ARGs is observed for small plasmids ($r_s=0.15$, $p\text{-value}=0.07$, using Spearman correlation). The rarity of TEs in small plasmids may be due to purifying selection against plasmid genome expansion. Yet, the rarity of TEs in large non-mobilizable plasmids suggests that other barriers to TE invasion, besides plasmid size, may be at play. In our analysis, 83% of the five selected transposase families are in plasmids, hence the presence of these five TEs is clearly biased towards plasmids. TEs are also the major drivers of gene flow between replicons (i.e., chromosome and plasmids) in bacterial cells, even so there are barriers for TE transfer of ARGs between plasmids and chromosomes. For example, a core chromosome ARG is unlikely to be transferred to plasmid in the same genus (Wang et al., 2022). The reasons for such barriers are likely related to conflicting regulation of duplicated genes and dose effect (Sorek et al., 2007; Baltrus, 2013).

Comparing to the genomes of MDR plasmids, the sequence of ARG-carrying TEs is highly conserved. This difference suggests that the evolutionary dynamics of plasmids and their resident tARGs are distinct, yet the reasons for that remain elusive. Plasmids that mediate the transfer of TEs between bacterial cells, can be targeted by the bacterial immune system (Marraffini & Sontheimer, 2008; Jaskólska et al., 2022). In contrast, to our knowledge, it remains unknown if defense system against TEs exist in prokaryotes. Notably, the endonuclease of bacterial immune system CRISPR is encoded by IS200/IS605 transposon family (Altae-Tran et al., 2021; Karvelis et al., 2021). In this case, the TE itself is considered an integral component of the bacterial immune system. Such transposable elements may be guarded from the bacterial immune system as they are not recognized as foreign DNA. It is tenable to speculate that even in case of immune response to invading plasmids upon entry

into the recipient cell, the hitchhiking TEs may already start transposition onto another DNA molecule in the recipient cell and thus escape the immune system. Following the degradation of the attacked plasmid, TEs and their cargo gene may thus persist in the new host cell and avoid extinction. Additional research is required to test if this scenario explains the incompatible evolutionary histories of plasmids and tARGs.

Methods

Data collection

We examined all complete genomes of plasmid-carrying isolates in the genera *Escherichia*, *Salmonella* and *Klebsiella* from NCBI RefSeq database. Assembly and annotation files were downloaded from the NCBI RefSeq database (version 01/2021). Isolate metadata was downloaded from BioSample database. Samples lacking host information were examined in detail. A total of 23 strains sequenced within the framework of laboratory experiments were excluded. The surveyed genomes include: 1,114 genomes and 3,098 plasmids from *Escherichia*; 755 genomes and 2,693 plasmids from *Klebsiella*; 572 genomes and 992 plasmids from *Salmonella* (Supplementary Table S1).

ARG identification

Homologs to genes that may provide antibiotic resistance were identified based on the comprehensive antibiotic resistance database (Alcock et al., 2020) (CARD version 3.1.0) using Resistance Gene Identifier (version 5.1.1, with parameter --clean). We excluded protein-coding genes whose sequence had less than 70% identity or was less than 90% shorter or greater than 30% longer in comparison with the CARD ARGs. Of the total 3,044 genes in CARD, 416 (14%) ABR genes have 114,457 homologues in the examined isolates.

Construction of protein families

Reciprocal best hits (RBHs) of protein sequences between all pairs of 9,225 replicons were identified using MMseqs2 (Steinegger & Söding, 2017) (v.13.45111, with module easy-rbh applying a threshold of $E\text{-value} \leq 1 \times 10^{-10}$). RBHs were further compared by global alignment using parasail-python (Daily, 2016) (v. 1.2.4, with the Needleman-Wunsch algorithm). Sequence pairs with greater than or equal to 30% identical amino acids were clustered into protein families using a high-performance parallel implementation of the Markov clustering algorithm (Azad et al., 2018) (HipMCL with parameter --abc -l 2.0). After protein family clustering, 114,457 ARGs are found in 138 protein families; the CARD classification was only partially in agreement with our protein family classification (114,457/203,591). Most of the contradictions were in chromosomal families (97%) with only 3% contradictions with the

plasmid-encoded families. Aiming for a conservative analysis, we excluded members not identified as ARGs by CARD from the analysis.

Coincidence test

Significant co-occurrence of ARG pairs from different protein families was tested using Fisher's exact test with `fisher.test` function in R (version 4.0.3, $\alpha=0.05$ and correction for multiple comparisons with false discovery rate (FDR)). As plasmid and chromosome have different ARG composition in different genera, The test has been done in different replicon types and genera separately. For each test, sample size n is the number of chromosomes or the number of ARG-carrying plasmids. Notably, ARGs in the chromosomal core genome will always be found with all the other ARGs. Owing to the test is only feasible on a 2×2 contingency table, any pairs of chromosome ARGs including core ARGs won't have a full table, hence they were not taken into the test. Test results are summarized into matrices plotted using `pheatmap` function in R (version 4.0.3). Eight families of ARG encoding multiple resistance mechanisms were not shown in the mechanism matrices.

Collinear syntenic blocks

Collinear syntenic blocks (i.e., gene order) were identified using CSBFinder-S (Svetlitsky et al., 2020) (With parameter `-q 4 -ins 0 -lmax 200 -c -cs -alg MATCH_POINTS`). Only the same blocks in at least four genomes were kept, allowing no insertion and maximal length of 200 protein-coding sequences.

TE sequence collection and Blast searching for TE-related CSB sequence in genomes

Intact transposable elements sequences were collected from three TE databases: The Transposon Registry (version 10/2021) (Tansirichaiya et al., 2019), TnCentral (version 10/2021) (Ross et al., 2021), and ISfinder database (version 10/2020) (Siguier, 2006). Partial sequences and complete plasmid sequences were excluded. We searched TE-related CSBs in all 2,441 complete genomes using Blast+ (version: 2.9.0+ parameter: $E\text{-value} \leq 1 \times 10^{-10}$) (Camacho et al., 2009). The blast results were further matched to the CSB results, only the CSBs with at least two proteins completely matched to a single TE sequence were kept. We further examined TE-related CSBs in an additional set of 935 complete genomes of isolates without plasmid in the three genera from NCBI RefSeq database (version 01/2021, Supplementary Table S3). The nucleotide sequences of plasmid CSB containing selected TnpA-ARG combinations (from Fig. 2B) were extracted. The CSB sequences were searched in the 935 chromosomes using Blast+ (version: 2.9.0+, parameter: $E\text{-value} \leq 1 \times 10^{-10}$). CSBs matched to chromosome sequence with coverage $> 80\%$ and identity $> 90\%$ were kept.

ARG island analysis

The distance between ARGs was counted as the number of protein coding genes on the same strand from one ARG to the adjacent n th ARG ($n=2-10$). The randomness of ARG distance was tested using chi square goodness-of-fit test with negative binomial distribution as the null hypothesis. The distance of randomly sampled genes was used as a control for distance comparison. A same number of gene families were sampled. Each sampled protein family has a similar family size of the corresponding ARG family. The distance of randomly sampled gene was counted in the same way as ARG distance, the sampling process was repeated for 100 times. The distance distribution of ARG and control group were compared using Wilcoxon rank-sum test for each step separately ($n=2-10$).

Plasmid size threshold value estimation

The multimodality of plasmid size distribution was tested using R function `modetest`. The minimum and maximum density of plasmid size distribution were estimated using R function `locmodes` with a bimodal distribution.

Plasmid typing

Plasmid incompatibility group type was identified using PlasmidFinder (version 2.1.6 with parameters: `-l 0.6 -t 0.9 -mp blastn`) and pMLST (Carattoli & Hasman, 2020) (version 2.0.3, with parameter `-mp blastn`). Plasmid mobility type was identified using MOB-suite (Robertson & Nash, 2018) (version 3.0.3)

Phylogenetic analysis

Protein sequences were aligned using MAFFT (Katoh & Standley, 2013) (version v7.475). Owing to the large number of identical protein sequence among transposase, the sequences were filtered into non-redundant amino-acid sequence. Maximum likelihood trees were reconstructed using IQ-Tree (L. T. Nguyen et al., 2015) (version 1.6.12) with Le and Gascuel (LG) model and the LG4X model as additional alternative (with parameter `-mset LG -madd LG4X -bb 1000`). The resulting trees were rooted using the midpoint criterion and visualized using iTOL (Letunic & Bork, 2021) (version 6.5.8).

Data accessibility

The data are provided in electronic supplementary material.

Author contributions

Yiqing Wang and Tal Dagan conceived the study.

Yiqing Wang designed and performed the data analysis and visualizations.

Yiqing Wang and Tal Dagan interpreted the results and wrote the manuscript.

Acknowledgments

We thank Hinrich Schulenburg, Giddy Landan, Nils Hülter, Devani Picazo, Dustin Hanke, Ishan Bhatt for fruitful discussion and critical comments on the manuscript. We are grateful for financial support from the German Science Foundation (RTG 2501 TransEvo) and the China Scholarship Council (CSC scholarship to Y.W.). This research was supported in part through high-performance computing resources available at the Kiel University Computing Centre.

References

- Alcock, B. P., Raphenya, A. R., Lau, T. T. Y., Tsang, K. K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen, A. L. v., Cheng, A. A., Liu, S., Min, S. Y., Miroshnichenko, A., Tran, H. K., Werfalli, R. E., Nasir, J. A., Oloni, M., Speicher, D. J., Florescu, A., Singh, B., ... McArthur, A. G. (2020). CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Research*, 48(D1), D517–D525. <https://doi.org/10.1093/NAR/GKZ935>
- Altae-Tran, H., Kannan, S., Demircioglu, F. E., Oshiro, R., Nety, S. P., McKay, L. J., Dlakić, M., Inskeep, W. P., Makarova, K. S., Macrae, R. K., Koonin, E. v., & Zhang, F. (2021). The widespread IS200/IS605 transposon family encodes diverse programmable RNA-guided endonucleases. *Science*, 374(6563), 57–65. <https://doi.org/10.1126/science.abj6856>
- Ares-Arroyo, M., Rocha, E. P. C., & Gonzalez-Zorn, B. (2021). Evolution of ColE1-like plasmids across γ -Proteobacteria: From bacteriocin production to antimicrobial resistance. *PLOS Genetics*, 17(11), e1009919. <https://doi.org/10.1371/JOURNAL.PGEN.1009919>
- Azad, A., Pavlopoulos, G. A., Ouzounis, C. A., Kyrpides, N. C., & Buluç, A. (2018). HipMCL: a high-performance parallel implementation of the Markov clustering algorithm for large-scale networks. *Nucleic Acids Research*, 46(6), e33–e33. <https://doi.org/10.1093/NAR/GKX1313>
- Baltrus, D. A. (2013). Exploring the costs of horizontal gene transfer. *Trends in Ecology & Evolution*, 28(8), 489–495. <https://doi.org/10.1016/J.TREE.2013.04.002>
- Barry, K. E., Wailan, A. M., Sheppard, A. E., Crook, D., Vegesana, K., Stoesser, N., Parikh, H. I., Sebra, R., & Mathers, A. J. (2019). Don't overlook the little guy: An evaluation of the frequency of small plasmids co-conjugating with larger carbapenemase gene containing plasmids. *Plasmid*, 103, 1–8. <https://doi.org/10.1016/J.PLASMID.2019.03.005>

- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, 10(1), 1–9. <https://doi.org/10.1186/1471-2105-10-421>
- Carattoli, A., & Hasman, H. (2020). PlasmidFinder and In Silico pMLST: Identification and typing of plasmid replicons in whole-genome sequencing (WGS). *Methods in Molecular Biology*, 2075, 285–294. https://doi.org/10.1007/978-1-4939-9877-7_20
- Che, Y., Yang, Y., Xu, X., Brinda, K., Polz, M. F., Hanage, W. P., & Zhang, T. (2021). Conjugative plasmids interact with insertion sequences to shape the horizontal transfer of antimicrobial resistance genes. *Proceedings of the National Academy of Sciences of the United States of America*, 118(6), e2008731118. <https://doi.org/10.1073/PNAS.2008731118>
- Coluzzi, C., Garcillán-Barcia, M. P., de la Cruz, F., & Rocha, E. P. C. (2022). Evolution of plasmid mobility: Origin and fate of conjugative and nonconjugative plasmids. *Molecular Biology and Evolution*, 39(6). <https://doi.org/10.1093/MOLBEV/MSAC115>
- Daily, J. (2016). Parasail: SIMD C library for global, semi-global, and local pairwise sequence alignments. *BMC Bioinformatics*, 17(1), 1–11. <https://doi.org/10.1186/S12859-016-0930-Z>
- del Solar, G., Giraldo, R., Ruiz-Echevarría, M. J., Espinosa, M., & Díaz-Orejas, R. (1998). Replication and control of circular bacterial plasmids. *Microbiology and Molecular Biology Reviews*, 62(2), 434–464. <https://doi.org/10.1128/MMBR.62.2.434-464.1998>
- Gilmour, M. W., Thomson, N. R., Sanders, M., Parkhill, J., & Taylor, D. E. (2004). The complete nucleotide sequence of the resistance plasmid R478: defining the backbone components of incompatibility group H conjugative plasmids through comparative genomics. *Plasmid*, 52(3), 182–202. <https://doi.org/10.1016/j.plasmid.2004.06.006>
- Harmer, C. J., & Hall, R. M. (2021). IS26 cannot move alone. *Journal of Antimicrobial Chemotherapy*, 76(6), 1428–1432. <https://doi.org/10.1093/JAC/DKAB055>
- Harmer, C. J., Moran, R. A., & Hall, R. M. (2014). Movement of IS26-Associated antibiotic resistance genes occurs via a translocatable unit that includes a single IS26 and preferentially inserts adjacent to another IS26. *MBio*, 5(5), e01801-14. <https://doi.org/10.1128/MBIO.01801-14>
- Harmer, C. J., Pong, C. H., & Hall, R. M. (2021). A brief guide to correct annotation of IS26 and variants. *Journal of Antimicrobial Chemotherapy*, 76(8), 2213–2215. <https://doi.org/10.1093/JAC/DKAB139>
- Jaskólska, M., Adams, D. W., & Blokesch, M. (2022). Two defence systems eliminate plasmids from seventh pandemic *Vibrio cholerae*. *Nature*, 604(7905), 323–329. <https://doi.org/10.1038/s41586-022-04546-y>

- Karvelis, T., Druteika, G., Bigelyte, G., Budre, K., Zedaveinyte, R., Silanskas, A., Kazlauskas, D., Venclovas, Č., & Siksnyis, V. (2021). Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. *Nature*, 599(7886), 692–696. <https://doi.org/10.1038/s41586-021-04058-1>
- Katoh, K., & Standley, D. M. (2013). MAFFT Multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772–780. <https://doi.org/10.1093/MOLBEV/MST010>
- Khedkar, S., Smyshlyaev, G., Letunic, I., Maistrenko, O. M., Coelho, L. P., Orakov, A., Forslund, S. K., Hildebrand, F., Luetge, M., Schmidt, T. S. B., Barabas, O., & Bork, P. (2022). Landscape of mobile genetic elements and their antibiotic resistance cargo in prokaryotic genomes. *Nucleic Acids Research*, 50(6), 3155–3168. <https://doi.org/10.1093/NAR/GKAC163>
- Kumarasamy, K. K., Toleman, M. A., Walsh, T. R., Bagaria, J., Butt, F., Balakrishnan, R., Chaudhary, U., Doumith, M., Giske, C. G., Irfan, S., Krishnan, P., Kumar, A. v., Maharjan, S., Mushtaq, S., Noorie, T., Paterson, D. L., Pearson, A., Perry, C., Pike, R., ... Woodford, N. (2010). Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: a molecular, biological, and epidemiological study. *The Lancet Infectious Diseases*, 10(9), 597–602. [https://doi.org/10.1016/S1473-3099\(10\)70143-2](https://doi.org/10.1016/S1473-3099(10)70143-2)
- León-Sampedro, R., DelaFuente, J., Díaz-Agero, C., Crellen, T., Musicha, P., Rodríguez-Beltrán, J., de la Vega, C., Hernández-García, M., López-Fresneña, N., Ruiz-Garbajosa, P., Cantón, R., Cooper, B. S., & San Millán, Á. (2021). Pervasive transmission of a carbapenem resistance plasmid in the gut microbiota of hospitalized patients. *Nature Microbiology* 2021 6:5, 6(5), 606–616. <https://doi.org/10.1038/s41564-021-00879-y>
- Letunic, I., & Bork, P. (2021). Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, 49(W1), W293–W296. <https://doi.org/10.1093/NAR/GKAB301>
- Liu, Y. Y., Wang, Y., Walsh, T. R., Yi, L. X., Zhang, R., Spencer, J., Doi, Y., Tian, G., Dong, B., Huang, X., Yu, L. F., Gu, D., Ren, H., Chen, X., Lv, L., He, D., Zhou, H., Liang, Z., Liu, J. H., & Shen, J. (2016). Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. *The Lancet Infectious Diseases*, 16(2), 161–168. [https://doi.org/10.1016/S1473-3099\(15\)00424-7](https://doi.org/10.1016/S1473-3099(15)00424-7)
- Marraffini, L. A., & Sontheimer, E. J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science*, 322(5909), 1843–1845. <https://doi.org/10.1126/science.1165771>

- Miriagou, V., Carattoli, A., & Fanning, S. (2006). Antimicrobial resistance islands: resistance gene clusters in *Salmonella* chromosome and plasmids. *Microbes and Infection*, 8(7), 1923–1930. <https://doi.org/10.1016/J.MICINF.2005.12.027>
- Murray, C. J., Ikuta, K. S., Sharara, F., Swetschinski, L., Robles Aguilar, G., Gray, A., Han, C., Bisignano, C., Rao, P., Wool, E., Johnson, S. C., Browne, A. J., Chipeta, M. G., Fell, F., Hackett, S., Haines-Woodhouse, G., Kashef Hamadani, B. H., Kumaran, E. A. P., McManigal, B., ... Naghavi, M. (2022). Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet*, 399(10325), 629–655. [https://doi.org/10.1016/S0140-6736\(21\)02724-0](https://doi.org/10.1016/S0140-6736(21)02724-0)
- Nguyen, L. T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, 32(1), 268–274. <https://doi.org/10.1093/MOLBEV/MSU300>
- Nguyen, S. v., Harhay, G. P., Bono, J. L., Smith, T. P. L., & Harhay, D. M. (2017). Genome sequence of the thermotolerant foodborne pathogen *Salmonella enterica* Serovar Senftenberg ATCC 43845 and phylogenetic analysis of loci encoding increased protein quality control mechanisms. *MSystems*, 2(1), e00190-16. <https://doi.org/10.1128/mSystems.00190-16>
- Nikaido, H. (2009). Multidrug resistance in bacteria. *Annual Review of Biochemistry*, 78, 119. <https://doi.org/10.1146/annurev.biochem.78.082907.145923>
- Partridge, S. R. (2011). Analysis of antibiotic resistance regions in Gram-negative bacteria. *FEMS Microbiology Reviews*, 35(5), 820–855. <https://doi.org/10.1111/J.1574-6976.2011.00277.X>
- Partridge, S. R., Kwong, S. M., Firth, N., & Jensen, S. O. (2018). Mobile genetic elements associated with antimicrobial resistance. *Clinical Microbiology Reviews*, 31(4), e00088-17. <https://doi.org/10.1128/CMR.00088-17>
- Paulsen, I. T., Banerjee, L., Hyers, G. S. A., Nelson, K. E., Seshadri, R., Read, T. D., Fouts, D. E., Eisen, J. A., Gill, S. R., Heidelberg, J. F., Tettelin, H., Dodson, R. J., Umayam, L., Brinkac, L., Beanan, M., Daugherty, S., DeBoy, R. T., Durkin, S., Kolonay, J., ... Fraser, C. M. (2003). Role of mobile DNA in the evolution of vancomycin-resistant *Enterococcus faecalis*. *Science*, 299(5615), 2071–2074. <https://doi.org/10.1126/SCIENCE.1080613>
- Pong, C. H., Harmer, C. J., Ataide, S. F., & Hall, R. M. (2019). An IS26 variant with enhanced activity. *FEMS Microbiology Letters*, 366(3), 31. <https://doi.org/10.1093/FEMSLE/FNZ031>
- Ramirez, M. S., Iriarte, A., Reyes-Lamothe, R., Sherratt, D. J., & Tolmasky, M. E. (2019). Small *Klebsiella pneumoniae* Plasmids: Neglected Contributors to Antibiotic

- Resistance. *Frontiers in Microbiology*, 10, 2182.
<https://doi.org/10.3389/FMICB.2019.02182>
- Redondo-Salvo, S., Fernández-López, R., Ruiz, R., Vielva, L., de Toro, M., Rocha, E. P. C., Garcillán-Barcia, M. P., & de la Cruz, F. (2020). Pathways for horizontal gene transfer in bacteria revealed by a global map of their plasmids. *Nature Communications* 11(1), 1–13. <https://doi.org/10.1038/s41467-020-17278-2>
- Robertson, J., & Nash, J. H. E. (2018). MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. *Microbial Genomics*, 4(8), e000206. <https://doi.org/10.1099/mgen.0.000206>
- Rodríguez-Beltrán, J., DelaFuente, J., León-Sampedro, R., MacLean, R. C., & San Millán, Á. (2021). Beyond horizontal gene transfer: the role of plasmids in bacterial evolution. *Nature Reviews Microbiology*, 19(6), 347–359. <https://doi.org/10.1038/s41579-020-00497-1>
- Rodríguez-Beltrán, J., Sørum, V., Toll-Riera, M., de la Vega, C., Peña-Miller, R., & Millán, Á. S. (2020). Genetic dominance governs the evolution and spread of mobile genetic elements in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 117(27), 15755–15762. <https://doi.org/10.1073/PNAS.2001240117>
- Ross, K., Varani, A. M., Snesrud, E., Huang, H., Alvarenga, D. O., Zhang, J., Wu, C., McGann, P., & Chandler, M. (2021). TnCentral: a prokaryotic transposable element database and web portal for transposon analysis. *MBio*, 12(5), e0206021. <https://doi.org/10.1128/mBio.02060-21>
- San Millan, A., Heilbron, K., & MacLean, R. C. (2013). Positive epistasis between co-infecting plasmids promotes plasmid survival in bacterial populations. *The ISME Journal*, 8(3), 601–612. <https://doi.org/10.1038/ismej.2013.182>
- Siguier, P. (2006). ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Research*, 34(90001), D32–D36. <https://doi.org/10.1093/nar/gkj014>
- Smillie, C., Garcillán-Barcia, M. P., Francia, M. V., Rocha, E. P. C., & de la Cruz, F. (2010). Mobility of plasmids. *Microbiology and Molecular Biology Reviews*, 74(3), 434–452. <https://doi.org/10.1128/mmbr.00020-10>
- Sorek, R., Zhu, Y., Creevey, C. J., Francino, M. P., Bork, P., & Rubin, E. M. (2007). Genome-wide experimental determination of barriers to horizontal gene transfer. *Science*, 318(5855), 1449–1452. <https://doi.org/10.1126/SCIENCE.1147112>
- Steinegger, M., & Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, 35(11), 1026–1028. <https://doi.org/10.1038/nbt.3988>

- Svetlitsky, D., Dagan, T., & Ziv-Ukelson, M. (2020). Discovery of multi-operon colinear syntenic blocks in microbial genomes. *Bioinformatics*, 36(Suppl_1), i21–i29. <https://doi.org/10.1093/BIOINFORMATICS/BTAA503>
- Tansirichaiya, S., Rahman, Md. A., & Roberts, A. P. (2019). The transposon registry. *Mobile DNA*, 10(1), 40. <https://doi.org/10.1186/s13100-019-0182-3>
- Touchon, M., & Rocha, E. P. C. (2007). Causes of insertion sequences abundance in prokaryotic genomes. *Molecular Biology and Evolution*, 24(4), 969–981. <https://doi.org/10.1093/MOLBEV/MSM014>
- Wagner, A. (2006). Periodic extinctions of transposable elements in bacterial lineages: Evidence from intragenomic variation in multiple genomes. *Molecular Biology and Evolution*, 23(4), 723–733. <https://doi.org/10.1093/MOLBEV/MSJ085>
- Wagner, A. (2008). Transposable elements as genomic diseases. *Molecular BioSystems*, 5(1), 32–35. <https://doi.org/10.1039/B814624C>
- Wang, Y., Batra, A., Schulenburg, H., & Dagan, T. (2022). Gene sharing among plasmids and chromosomes reveals barriers for antibiotic resistance gene transfer. *Philosophical Transactions of the Royal Society B*, 377(1842), 20200467. <https://doi.org/10.1098/RSTB.2020.0467>
- Wein, T., Wang, Y., Hülter, N. F., Hammerschmidt, K., & Dagan, T. (2020). Antibiotics interfere with the evolution of plasmid stability. *Current Biology*, 30(19), 3841–3847.e4. <https://doi.org/10.1016/j.cub.2020.07.019>
- Whelan, F. J., Hall, R. J., & McInerney, J. O. (2021). Evidence for selection in the abundant accessory gene content of a prokaryote pangenome. *Molecular Biology and Evolution*, 38(9), 3697–3708. <https://doi.org/10.1093/MOLBEV/MSAB139>
- Winter, A. R., Stewart, G. F., McFarlane, V. H., & Solowey, M. (1946). Pasteurization of Liquid Egg Products. *American Journal of Public Health and the Nations Health*, 36(5), 451–460. <https://doi.org/10.2105/AJPH.36.5.451>
- Zhang, Z., Chang, J., Xu, X., Hu, M., He, S., Qin, X., Zhou, M., Shi, C., & Shi, X. (2022). Phylogenomic analysis of *Salmonella enterica* Serovar Indiana ST17, an emerging multidrug-resistant clone in China. *Microbiology Spectrum*, 10(4), e0011522. <https://doi.org/10.1128/SPECTRUM.00115-22>
- Zhang, Z., He, S., Yang, B., Shi, C., Liu, Y., & Shi, X. (2021). Genomic characterization of an extensively drug-resistant chicken-borne *Salmonella* Indiana isolate carrying an IncHI2-IncHI2A plasmid. *Food Control*, 125, 107761. <https://doi.org/10.1016/J.FOODCONT.2020.107761>

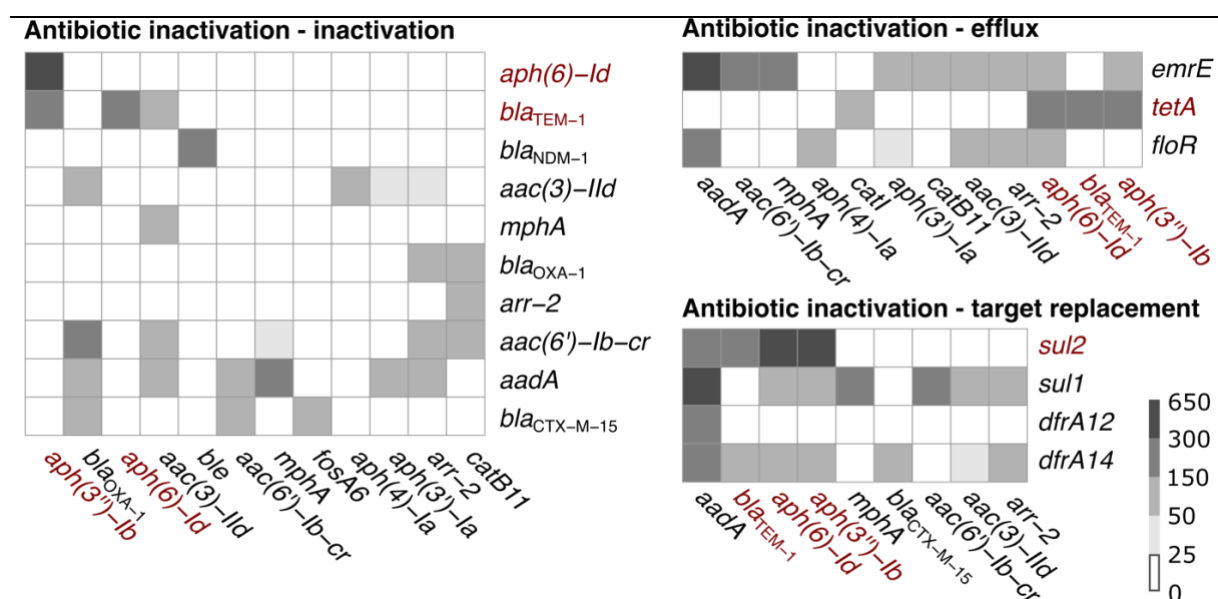
Supporting information

	<i>Salmonella</i>					<i>Escherichia</i>					<i>Klebsiella</i>					
Chromosome	29					182					17					efflux
	46	48				46	69				43	70				inactivation
	5	12	1			109	32	21			26	43	6			target alteration
	1	6	1			3	4				2	15	1			target protection
	13	39	6	1	4	17	42	8		4	10	45	9	5	4	target replacement

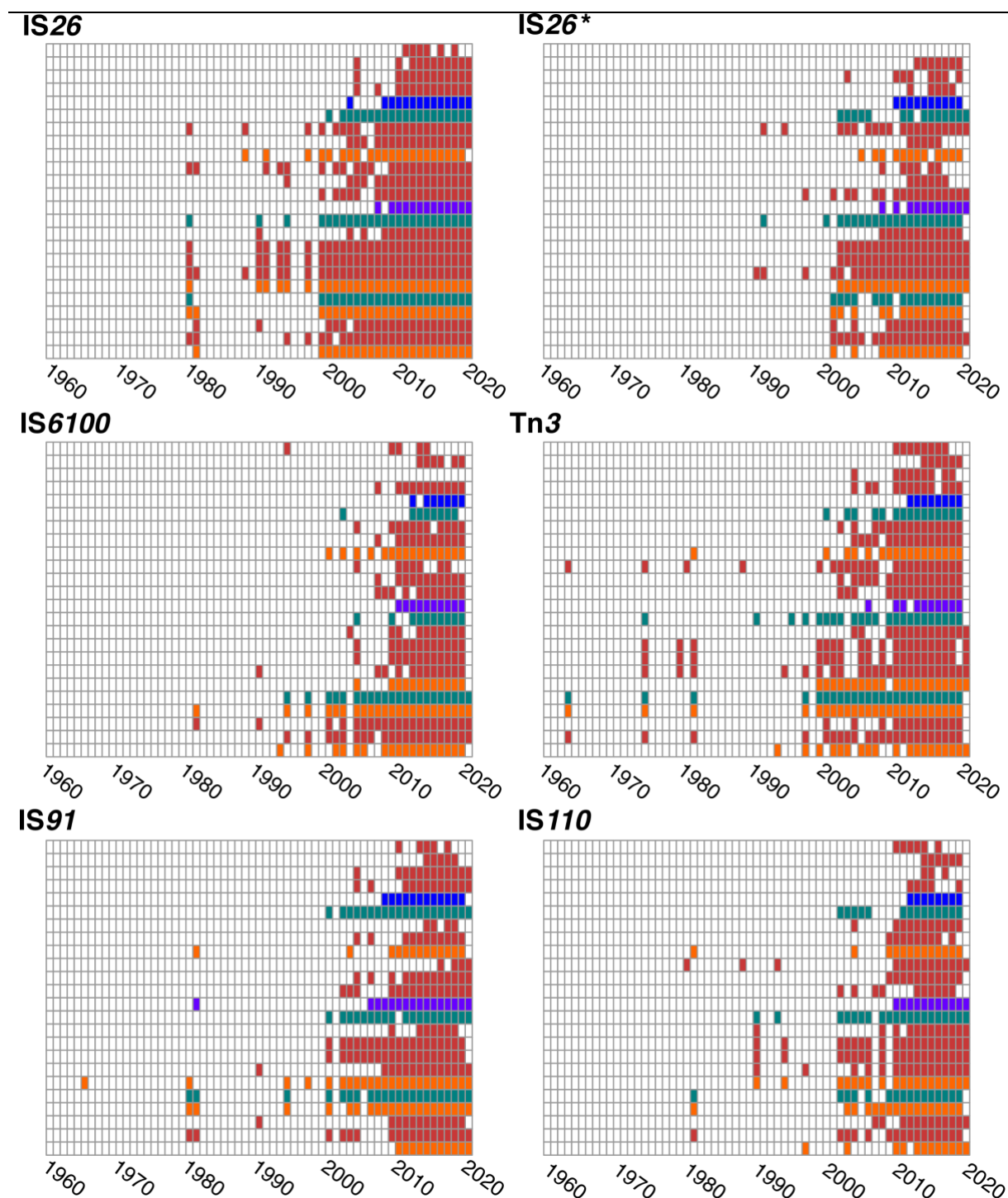
Supplementary Figure S1. CoARGs on chromosomes in *Salmonella*, *Escherichia*, and *Klebsiella*. The significant co-occurring combinations of paired ARGs (termed coARG) is summarized according to the resistance mechanism. The numbers of coARGs are shown in the heatmap. Beside antibiotic efflux in *Escherichia*, resistance mechanism pairs of inactivation-inactivation, inactivation-efflux, and inactivation-target replacement have the highest number of coARGs. ARGs encoding multiple mechanisms are not shown in the heatmap.

	<i>Salmonella</i>			<i>Escherichia</i>					<i>Klebsiella</i>			
Chromosome	9			15					4			efflux
	8			8					19	3		inactivation
			1	43	13	6			13	6	5	target alteration
	1			1					1			target protection
	8			2					4			target replacement
Plasmid		8		6	19				3	26		inactivation
	1	4		1	6				1	3		target alteration
					1	1				2		target protection
		1	1		8	1	1	1		6	1	target replacement
	efflux	inactivation	target alteration	efflux	inactivation	target alteration	target protection	target replacement	efflux	inactivation	target alteration	

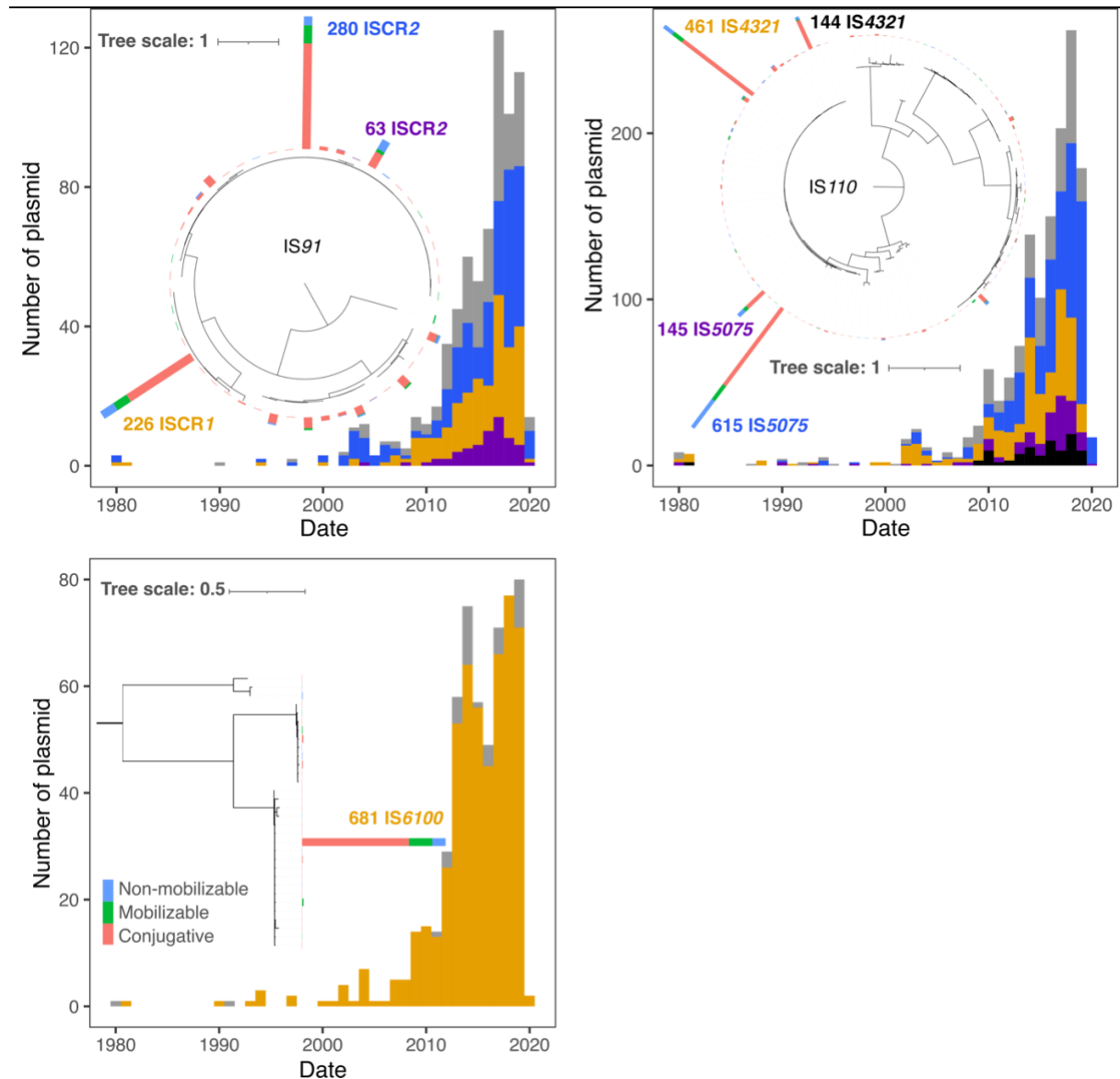
Supplementary Figure S2. Less frequent co-occurrence of ARGs on chromosomes and plasmids in *Salmonella*, *Escherichia*, and *Klebsiella*. The significant less frequent co-occurring combinations of paired ARGs is summarized according to the resistance mechanism. The numbers are shown in the heatmap. ARG pairs including genes encoding multiple resistance mechanisms are not shown in the heatmap. (Data supplied in Supplementary Table S2)



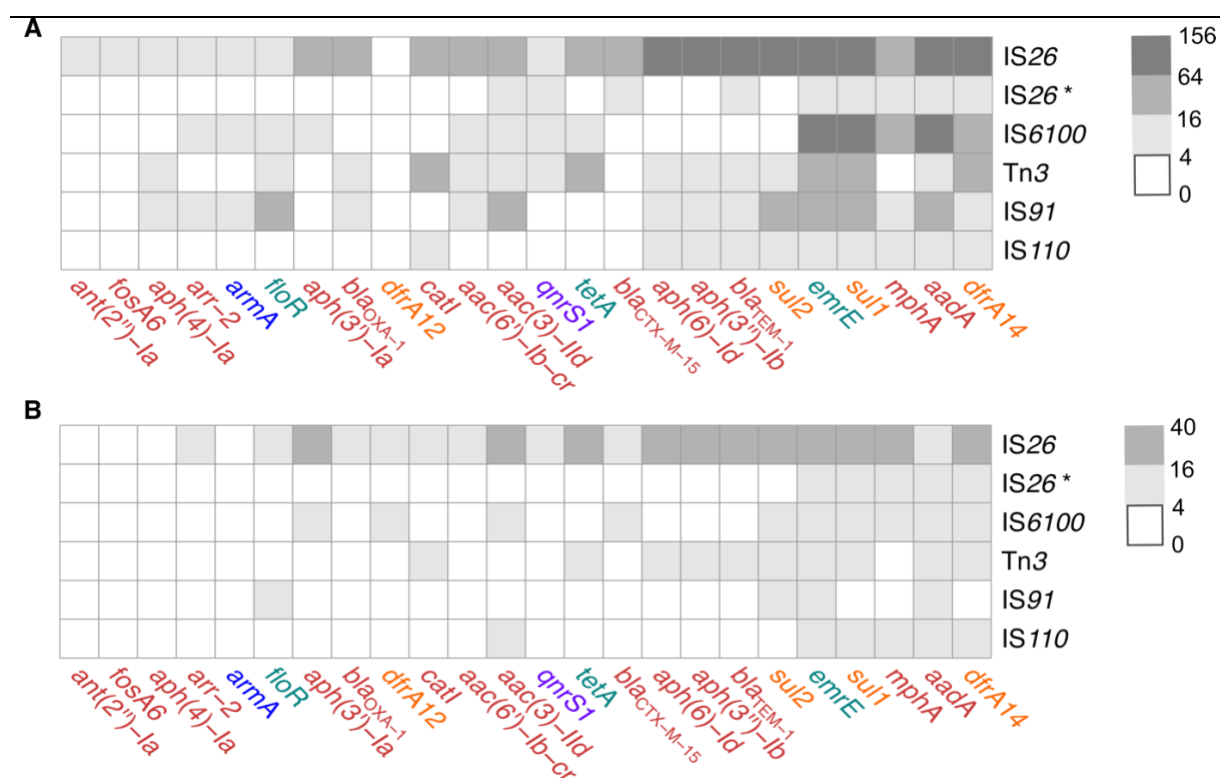
Supplementary Figure S3. Plasmid coARGs found in CSBs that match TEs. CoARGs found in collinear syntenic blocks (CSBs) that match transposable elements (TEs) are summarized in heatmaps. Only the coARGs found in all three genera are presented. Common gene names in three matrices are marked out in red, these genes potentially form ARG islands in plasmid genomes.



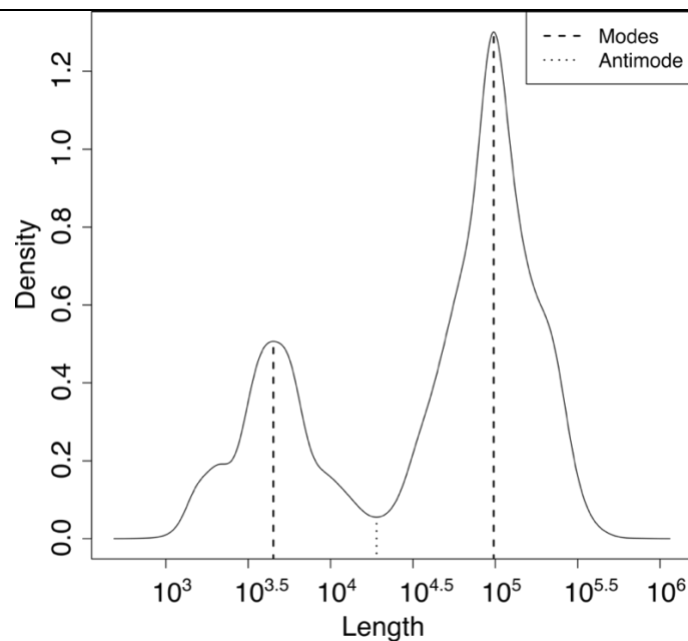
Supplementary Figure S4. Temporal pattern of plasmid CSBs containing five transposase families and 24 ARGs. Collinear syntenic blocks (CSBs) presence according to the isolation date of the strains is summarized. 24 ARGs in rows are ordered as the columns in Fig. 2B and identically colored according to the encoded resistance mechanism (red: antibiotic inactivation, green: antibiotic efflux, orange: antibiotic target replacement, blue: antibiotic target alteration, and purple: antibiotic target protection).



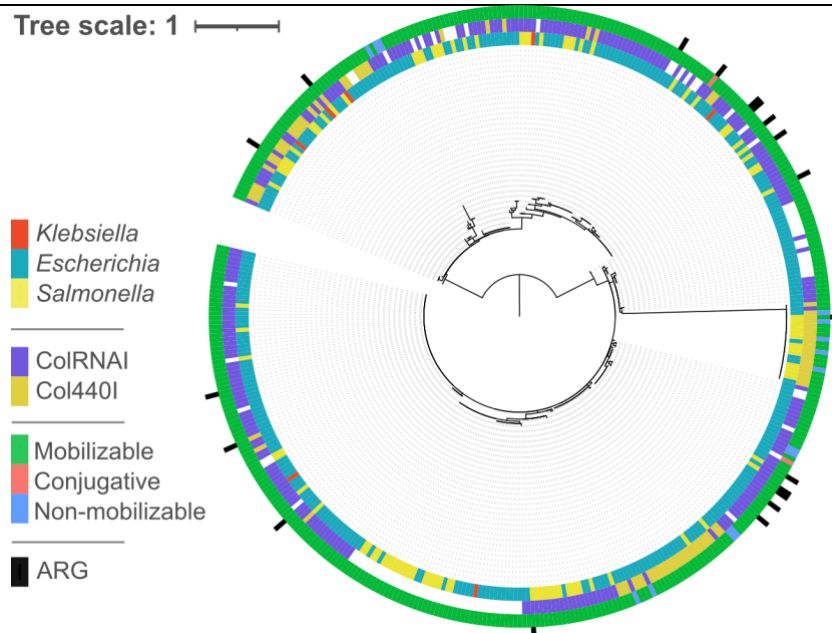
Supplementary Figure S5. Phylogeny and timeline distribution of three transposase families: IS91, IS110, and IS6100. Stacked bars at the tree leaves shows the number of the dominant variants, the bar is colored according to the mobility type of the variant carrying plasmid. Histogram below the phylogeny shows the isolation dates of plasmids carrying variants in different shades.



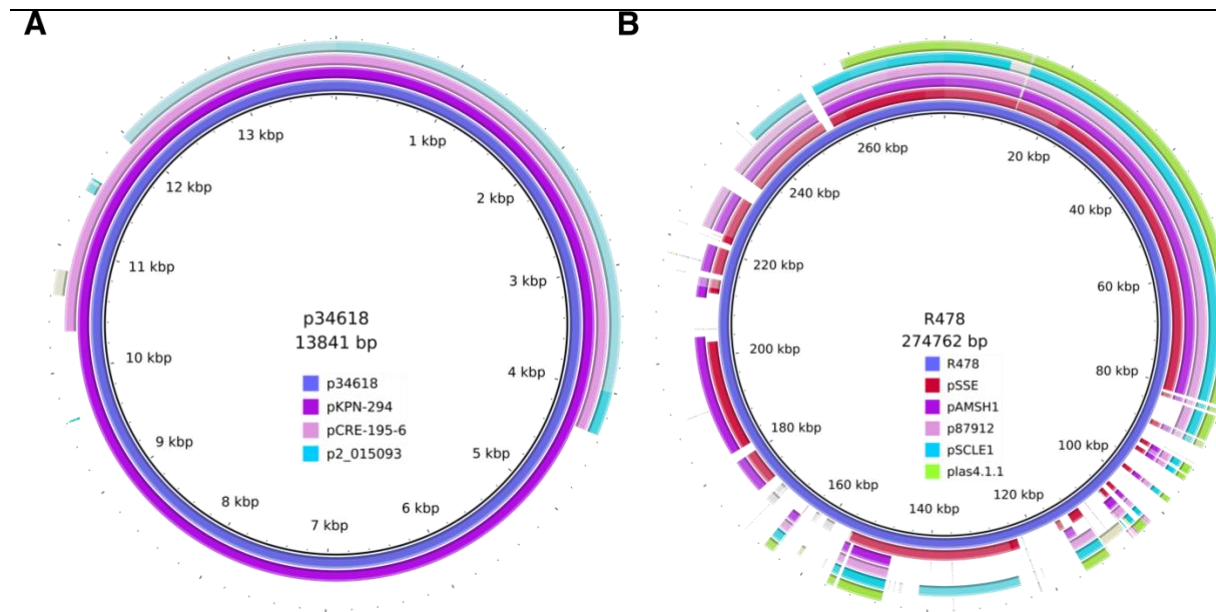
Supplementary Figure S6. Plasmid CSBs containing combinations of 24 ARGs and five transposase families are found in chromosomes. Color gradient of the square cells corresponds to the number of ARG-transposase co-occurrence in CSBs in unique chromosomes. The matrices are ordered as in Fig. 2B, ARG names are colored as in Fig. 2B according to the encoded resistance mechanisms (red: antibiotic inactivation, green: antibiotic efflux, orange: antibiotic target replacement, blue: antibiotic target alteration, and purple: antibiotic target protection). A) The plasmid CSBs containing co-occurrence of ARG-transposase are searched in 2,441 chromosomes with plasmids in the same host cell. B) The plasmids CSBs containing co-occurrence of ARG-transposase are searched in 935 chromosomes without plasmids in the same host cell. These two matrices show a similar pattern, which provides the evidence for lateral transfer of transposable elements from plasmid into chromosomes regardless their previous plasmid vehicles degraded or not.



Supplementary Figure S7. Density plot of genome size for all plasmids in the dataset. Plasmid size is presented in a common logarithm scale. A bi-modal distribution is given for the estimation of the two highest density values (marked out with thick dash lines) and the lowest density value (marked out with a thin dash line).



Supplementary Figure S8. Phylogenetic tree of 422 Rop protein sequences encoded on 405 plasmids in *Escherichia* and *Salmonella*. In *Escherichia* and *Salmonella* Rop-plasmids rarely encode ARGs. Rings of the phylogenetic tree show (from inner to outer ring): host genus, plasmid type, plasmid mobility type (see legend), and the presence of ARGs.



Supplementary Figure S9. Genome comparison of Rop- and RepB-plasmids. A) A comparison between four Rop-plasmids selected from the two sister groups of ColRNAI mobilizable plasmids (in Fig. 4). ABR plasmids p34618 (used as reference genome) and pKPN-294 containing Tn 1331 Δ :IS26 have the almost identical plasmid backbone as the non-ABR plasmids pCRE-195-6 and p2_015093 from the sister group. B) A comparison between six plasmids selected from the IncHI2/IIA group (Fig. 5A in yellow). IncHI2/IIA plasmids carry the most ARGs found in islands (median number of ARGs in island is 12, median ARG per plasmid is 14.5), they represent a prime example of extensive drug resistance (XDR) plasmids found mainly in *Escherichia* and *Salmonella*. Notably, IncHI2/IncHI2A is the only group of plasmids that consistently possesses two RepB homologs in the plasmid backbone. The IncHI2/IncHI2A multi-replicon plasmid backbone was first described in the detailed genomic analysis of plasmid R478 in *Serratia marcescens* (Gilmour et al., 2004). R478 is also the prototype of IncHI2/IncHI2A multi-replicon plasmid which was isolated in the USA in 1969. In our data, the earliest reported IncHI2/IncHI2A plasmid is pSSE ATCC-43845, which was found in *Salmonella* from the USA in 1941 (Supplementary Table S1). Due to the absence of ARGs in that plasmid, it has been hypothesized as potentially a relic before the era of therapeutic antibiotic use (S. v. Nguyen et al., 2017). Conversely, it carries two genetic islands of thermoresistant genes and was reported as extremely thermotolerant in 1946 (Winter et al., 1946). In the past 80 years, IncHI2/IncHI2A plasmids are reported frequently as MDR or XDR plasmids, especially in China, e.g., plasmid p87912 with 27 ARGs (Zhang et al., 2021) and p14154A with 25 ARGs (Zhang et al., 2022), both in *Salmonella* isolates. With plasmid genome comparative analysis, we confirmed the four mentioned IncHI2/IncHI2A plasmids and three additional plasmids (pAMSH1, pSCLE1, and plas4.1.1 encoding 32, 12 and 22 ARGs separately) share a conserved plasmid backbone, which includes the two *repB* and 89 protein-coding genes, in total a 97 kbp sequence.

Concluding remarks

In this thesis, I have investigated the dynamics of plasmid gene content and the evolutionary history of plasmid genes. My study reveals the complexity of plasmid gene content evolution. However, there are still many questions remaining regarding the interactions between mobile genetic elements and their role in shaping bacterial genome evolution. Further studies are required to explore this fascinating field.

Plasmids are nonessential components of the bacterial genome. In chapter one, our study on the essential gene content of plasmids in *E. coli* revealed that true essential genes are rarely present on these plasmids. Instead, we found that many conjugative plasmids carry re-purposed essential genes, which suggests that plasmid genes provide benefits depending on the environmental conditions rather than being essential *per se*. We observed deep divergences in both the phylogenetic trees of the ubiquitous plasmid-encoded Ssb and several GroEL/ES protein sequences. The mobility of conjugative plasmids and the conserved syntenic block of GroE (later confirmed as Tn 125) indicate that the essential genes identified in our study may have originated from another host taxon rather than gene transfer from the *E. coli* chromosome. The emergence of Ssb-coding conjugative plasmids is likely an ancient event, as they are widely prevalent and stably persisting in the *E. coli* genomes. In contrast, GroE-encoding Tn 125 is more likely a recent transfer event that occurred within a limited population. These two examples illustrate that the transition of an essential gene to a beneficial gene could be accompanied by lateral transfer of genes mediated by plasmids and mobile genetic elements.

In chapter two, I investigated the distribution of antibiotic resistance genes across three genera in Enterobacteriaceae. Our findings suggest that the concept of core and accessory genes is not universal across different taxonomic groups. Rather, the transfer of genes between the chromosome and plasmids seems to play a crucial role in their transition from core to accessory status, that status depends on the taxon. Specifically, I observed that *oqxA* considered core in *Klebsiella*, which is less mobile between the chromosome and plasmid, becomes an accessory gene primarily on plasmids in *Escherichia* and *Salmonella*. These results demonstrate that the mobility of plasmid genes is evolvable and dependent on the genomic (or rather chromosomal) composition of the host taxon.

In chapter three, I delved into how transposable elements contribute to the evolution of antibiotic resistance gene content in multiple drug resistance plasmids. Analyzing the distribution of transposases and antibiotic resistance genes in genomic islands, as well as their conserved sequences across different types of plasmids, I found that the evolution of cargo genes is distinct from that of plasmid backbone genes. However, the more detailed interplay between plasmids and transposable elements remains unclear, and further research is needed

to fully understand the evolutionary trajectory of these elements. Considering the accumulating discoveries of various bacterial defense mechanisms against foreign DNAs, it is tenable to speculate the presence of defense mechanism against transposons, hence it would be intriguing to explore how chromosomes and plasmids defend against transposable elements.

Acknowledgements

Thank you, Tal! You are an amazing supervisor and also my friend. Thank you for choosing me to work on these projects. I would benefit from your advice both within and outside of research for the rest of my life. I'm so grateful to be spending my PhD life with you. I couldn't imagine what I would be doing now if I hadn't met you during the second interview with Christian. Every PhD is unique, but our generation is different due to the pandemic. I will always remember these days I spent with you. You made my dream come true.

I want to express my sincere thanks to our collaborator Hinrich, thanks for your guidance and support. Special appreciation goes to you for organizing TransEvo with everyone in this wonderful research graduate school.

My studies wouldn't be possible without the fruitful guidance and discussions from my colleagues in our group (from genomic microbiology group to evolutionary microbiology group). Special thanks to Giddy and Nils. Your professional knowledge in statistical analysis and plasmid biology is always so helpful. I regret that we didn't meet earlier and couldn't work longer together. I extend my heartfelt thanks to all the past and present group members. Working with you was an absolute joy, and the memories of our time together will stay with me forever.

Thanks to my co-authors, Tana and Aditi. Collaborating with you was a great experience, and I feel proud of what we've achieved together. Thank you for your dedication, hard work, and contributions to our project.

I cannot express enough gratitude to my beloved family and friends. Your unwavering emotional support has been the light that shone through the dark and cold winter in Kiel. Your love, encouragement, and kindness have carried me through the toughest times, and I feel so blessed to have you all in my life.

Finally, genuine thanks to the many people who helped me in Germany and in China. Even though we may have never met, I appreciate your help here.