Acquire and rewire: mobile genetic elements co-opt the bacterial core genome

Rebeca Carvalho [1], James P. J. Hall [2#], João Botelho [1#]

[1]Universidade Católica Portuguesa, CBQF - Centro de Biotecnologia e Química Fina – Laboratório Associado, Escola Superior de Biotecnologia, Rua Diogo Botelho 1327, 4169-005 Porto, Portugal

[2]Department of Evolution, Ecology and Behavior, Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool L69 7ZB

#correspondence: J.P.J.Hall@liverpool.ac.uk, jfbotelho@ucp.pt

Summary

It is widely assumed that bacterial genomes are streamlined, with genes encoding core cellular functions tightly integrated with one another, and with aligned fitness interests and evolutionary fate. Such genes stand to gain little from horizontal transfer, as any recipient organism would already carry its own copy better integrated into its resident networks. These putative restrictions motivated the 'complexity hypothesis', which predicts that genes performing core cellular functions, with products highly connected and specialized to cellular physiology, are rarely transferred into new genetic backgrounds. Yet homologues of core genes are repeatedly found outside their native chromosomal context, carried by mobile genetic elements (MGEs) such as plasmids and phage-plasmids, and by domesticated secondary replicons such as chromids. Explaining this pattern, we argue, requires us to look beyond the extrachromosomal location of a core gene copy and ask whether a copy rides a mobile, self-interested element or a domesticated, host-aligned one. On domesticated, vertically inherited carriers, a second copy plausibly serves the host, through gene dosage, niche-specific benefit or functional specialization. On mobile, selfish elements, it may instead serve the element, by promoting host behaviors that favor onwards MGE transmission. Because they enter cells that already have a host-encoded copy, we propose that mobile variants can matter through dosage or through effects disproportionate to their abundance, for example by co-assembling with or competing with the chromosomal product. In this perspective, we reflect on the distribution of core genes duplicated onto mobile and domesticated carriers, their evolutionary drivers and consequences, and make the case that these overlooked copies deserve systematic study.

Glossary

- **Chromosome**: the largest replicon, carrying most core genes.

- **Chromid**: a secondary replicon carrying at least one core gene that originated from a plasmid and was subsequently domesticated, typically acquiring genomic signatures close to those of the chromosome and cell-cycle-coordinated replication.

- **Plasmid and megaplasmid**: secondary replicons that lack the genomic signatures and essential core functions of chromids, although they may carry core-gene homologues. Megaplasmid refers to the largest plasmids, but the boundary is not strict: plasmid size is continuous, fixed cutoffs (e.g. 350 kb) are arbitrary and size is better judged relative to the host genome.

- **Domestication**: the evolutionary process by which an initially acquired mobile genetic element becomes a stably maintained, functionally integrated component of the host genome.

- **Paralogue interference**: a phenomenon by which the binding of a protein to its paralogue alters its function, and thus the effects of mutations at its locus.

- **Genetic dominance**: a non-linear relation between the abundance of an altered gene product and the resulting phenotype, such that a product present in only a minority of the cellular pool, for example a mobile-encoded copy, can have a disproportionate effect on the outcome.

- **Neofunctionalization**: diversification of one member of a duplicate pair towards a new function.

- **Subfunctionalization**: division of the ancestral functions between the two members of a duplicate pair, so that each retains only part of the progenitor's functions.

- **Operational genes**: genes involved in housekeeping, such as amino acid, nucleotide and cofactor biosynthesis, central and energy metabolism, and cell envelope biogenesis.

- **Informational genes**: genes involved in transcription, translation and related processes.

- **Streamlining:** minimizing cell size and complexity to increase fitness in nutrient-poor environments. In the context of genome evolution describes a trajectory favoring gene deletions to reduce the metabolic burdens of maintenance and expression.

- **Replicon**: A region of DNA that is replicated independently from others in the cell.

Core genes and accessory genes

The pangenome concept describes how the genome of a bacterium can be partitioned into 'core' genes, which are present in all members of a species, and 'accessory' genes, which have variable patterns of presence and absence [1]. Accessory genes are frequently gained and lost from lineages, are associated with mobile genetic elements (MGEs) able to catalyze their own horizontal transmission, and provide niche-specific adaptations that allow organisms to colonize new habitats [2,3]. Core genes, on the other hand, are present across all sequenced isolates, and tend to perform functions necessary for fundamental cell activity. Examples include machinery for protein translation, DNA replication, and core catabolic metabolism [1,4]. Since their functions are consistently required, we imagine core genes and their products to be tightly co-evolved and integrated into the metabolic functioning of the cell. With bacterial genome evolution often playing out large populations competing for limited resources, and unnecessary DNA replication and expression consuming resources that could otherwise be directed towards increasing inclusive fitness [5], selection should impose a **streamlining** pressure (see Glossary), favoring reduced gene content and efficient architectures [6,7] while maintaining essential functions [8]. Cells with streamlined genomes are expected to have lower requirements for nitrogen and phosphorus, which are limiting in many niches, and simulations show that increases in either population size or mutation rate lead to genome size reduction [9].

The pangenome framework is aligned with the 'complexity hypothesis' of gene susceptibility to horizontal gene transfer. Under this model, **operational genes**, are on average, more prone to successful horizontal gene transfer (HGT) than **informational genes**, because the products of informational genes are larger and more densely connected to other cellular machineries, and thus their performance is more dependent on the molecular context which is unlikely to be the same in a new cell [10,11]. Although the two frameworks do not overlap exactly, the complexity hypothesis provides a compelling mechanism by which pangenome patterns might be understood: loss of core genes is likely to be lethal, gain of core genes unnecessary, and horizontal transfer likely to be inefficient, since transferred alleles would usually be less well integrated into cellular physiology. However, homologues of core genes are frequently found not on the **chromosome**, but divided across other **replicons**, including as cargo on MGEs. What are core genes doing on MGEs?

Streamlined vs. multipartite genomes

To understand the role of core gene duplications and translocations, we first consider the structure of multipartite bacterial genomes. In contrast to the reduced genomes expected from streamlining, multipartite genomes comprise a chromosome plus one or more additional **replicons** [12,13], and constitute approximately 10% of bacterial genomes [12,14]. The available evidence particularly supports a process of plasmid/megaplasmid acquisition followed by translocation or copying of core genes by interreplicon gene transfer as a route to secondary-replicon (**chromid**) evolution [12,13,15]. Secondary chromosomes, by contrast, can also arise through primary-chromosome fission and should not be assumed to share the same evolutionary route as chromids [12]. In either case, the result is an extrachromosomal core gene: a core gene, or a homologue of one, encoded on a secondary replicon, either as the only copy in the cell (translocation) or alongside a chromosomal copy (duplication) (**Figure 1**). We focus mainly on the latter.

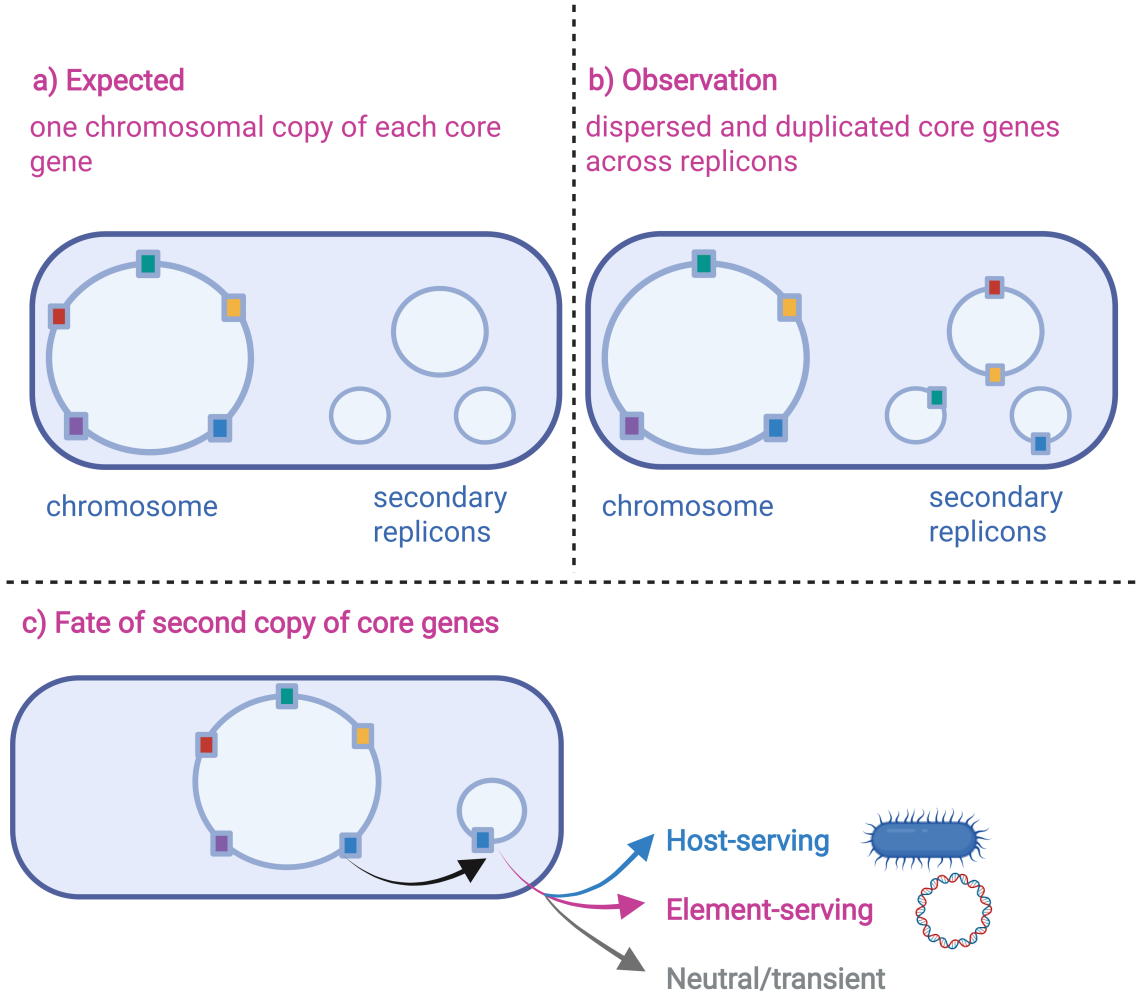


Figure 1. Core-gene homologues on secondary replicons and their potential fates. (A) Expectation: each core gene is present as a single copy on the chromosome, and secondary replicons (plasmids, megaplasmids, chromids) carry none. Colored squares are different core genes. (B) Observation: core genes are dispersed and duplicated across replicons. The gene shown in red has been translocated and is encoded only in a secondary replicon, whereas the genes shown in teal, orange and blue are duplicated, with a copy on the chromosome and a copy on a secondary replicon. The purple gene remains only on the chromosome. (C) A second copy of a core gene (here the blue gene, copied from the chromosome onto a secondary replicon) can be host-serving (blue), element-serving (magenta) or neutral/transient (grey). Created with BioRender.com.

Extrachromosomal core genes can be functionally integrated with the chromosome. In *Sinorhizobium meliloti*, over 10% of chromosomal genes appear to have a functionally redundant copy on two secondary replicons, pSymA or pSymB [16], which together make up 45% of the genome, so deleting a chromosomal gene often caused more severe growth

defects when either replicon was absent [17]. Similarly, the domesticated plasmids of plant-associated *Pantoea* carry core-gene families [18], and plasmid curing disrupts jointly encoded pathways and remodels chromosomal expression [19].

Secondary replicons vary widely in size, nucleotide composition, mobility, replication and partition systems, and contribution to core functions, making boundaries among plasmids, megaplasmids, and chromids ambiguous [15,20–24]. Large replicons often carry niche-adaptive, stress-response, host-association, or pathogenicity genes, but may also retain core genes [12,24]. Genes on secondary replicons can show lower average expression than chromosomal genes under some conditions [20], although this is not universal. Consequently, extrachromosomal core genes range from recently acquired, conditionally important functions to deeply integrated components of essential cellular pathways [18,25].

The carrier, not the replicon: a domesticated-mobile axis

Core-gene homologues on transmissible MGEs challenge the formulation of the complexity hypothesis, and indeed the core/accessory dichotomy, described above. Plasmid-encoded informational genes appear to be structurally and functionally organized instead of randomly distributed [26]. This pattern admits several explanations, but we propose that a key factor is the carrier's degree of alignment with its host—particularly the balance of vertical inheritance versus horizontal transmission—rather than the replicon it occupies.

On one side are '**domesticated**' carriers, including chromids and other stably inherited secondary replicons. These elements are primarily vertically transmitted, are generally incapable of independent horizontal transmission, have replication and segregation coordinated with the host cell cycle [27,28], and their long-term persistence is therefore coupled to host fitness. Core-gene families in domesticated *Pantoea* plasmids have diverged deeply from their chromosomal counterparts, consistent with rare gene translocation and with redundant copies followed by differential gene loss [18]. Selection on an additional core-gene copy in such carriers is therefore predominantly mediated through vertical transmission, i.e. host fitness, although the element retains its own replication and inheritance dynamics.

On the other side are more horizontally transmitted carriers, including transmissible plasmids. For these elements, reproductive success can arise from horizontal transfer as well as vertical inheritance [29,30]. Because a mobile carrier can leave one host lineage and take

over another, selection may favor traits that promote transmission even when they impose costs on, or otherwise conflict with, host interests.

This distinction reframes the question this perspective sets out to answer by asking what a second copy is selected to do once it arrives on each kind of carrier. As the next two sections argue, the two carriers point to largely non-overlapping sets of mechanisms.

Host-serving copies

We start on the domesticated side of the axis, where host and carrier share a fate and a second core-gene copy has to earn its keep by increasing the host's own fitness. At least three, non-exclusive mechanisms could be at work:

Dosage: immediately after duplication, the two copies initially encode the same product, and total expression may increase relative to the ancestral single-copy state, with positive, neutral, or negative fitness consequences [31]. In longitudinal *Acinetobacter baumannii* isolates from ICU patients, horizontally acquired redundancy of alcohol dehydrogenase *frmA* was associated with urinary-tract invasion, increased biofilm formation, and higher virulence in experimental assays [32]. The conservation of positional biases across bacteria [33] suggests that replication-associated gene dosage effects are a primary driver of local chromosomal organization, so the replicon carrying the copy matters as well as the copy itself. Notably, duplication onto a distinct replicon can enable gene copy number to be adjusted dynamically by regulating the copy number of the replicon: such is one explanation for the plasmid-borne carriage of rRNA operons by *Aureimonas* and other clades [34,35], and, in the context of the accessory genome, resistance gene dosage [36].

Niche specificity: a copy may be kept because it is useful in certain situations, not because it replicates a function that is essential in all situations. Work on *Nitrososphaerales* indicates that early lateral gene acquisition followed by lineage-specific duplication drove genome expansion [37]. The duplicated families include genes implicated in carbohydrate transport and coenzyme metabolism, consistent with a conditional benefit in particular soil and sediment niches rather than maintenance of universally essential functions.

Functional specialization: adaptive conflict arises when a single ancestral gene product performs multiple functions that cannot be jointly optimized because of antagonistic pleiotropy [38]. Duplication can relax that constraint by allowing the copies to specialize through adaptive divergence. More generally, duplicate retention may also proceed by **subfunctionalization** or **neofunctionalization**. In the *Bacillus cereus* group, recurrent

plasmid–chromosome exchange has generated homologous copies with divergent upstream and downstream sequences, consistent with distinct promoter and terminator architectures [39,40]. Such regulatory divergence could allow the copies to respond to different cues, enabling specific responses to environmental challenges and favoring retention of both, although persistence was not directly demonstrated experimentally in that study.

A null hypothesis: redundancy is not always adaptive. Wein et al. asked what happens when a plasmid acquires an essential chromosomal gene [41]. Essential genes were rarely plasmid-encoded in *Escherichia* isolates, and the few that were (*ssb*, *groEL*/*groES*) appeared integrated into plasmid-related functions. A plasmid-borne *groE* operon complemented a silenced chromosomal copy but reduced host fitness and destabilized plasmid inheritance, consistent with a detrimental dosage effect of redundant GroE. A plasmid may therefore retain an essential gene only if the chromosomal copy is lost or the copies diverge functionally, so essential-gene acquisition can act as a barrier to plasmid-mediated transfer [41]. In other cases, an extrachromosomal core gene may be neutral or transient, recently acquired and not yet purged by selection.

Element-serving copies

If a mobile copy is costly to its host yet persists, it may primarily serve the element. Here, acquiring the copy induces traits by rewiring host metabolism to favor the carrier's horizontal transmission instead of the host's own fitness, for example by influencing biofilm production or motility to increase the chance of productive horizontal transmission events. To achieve these 'acquire and rewire' functions, the element-encoded copy often needs to subvert or reprogram the functions of the resident copies. Whether a copy mainly favors the host or the element can be judged from five features of the interaction: its functional relationship to the chromosomal product, its expression pattern, its regulation, its effect on transmission, and its net costs and benefits (**Figure 2**).

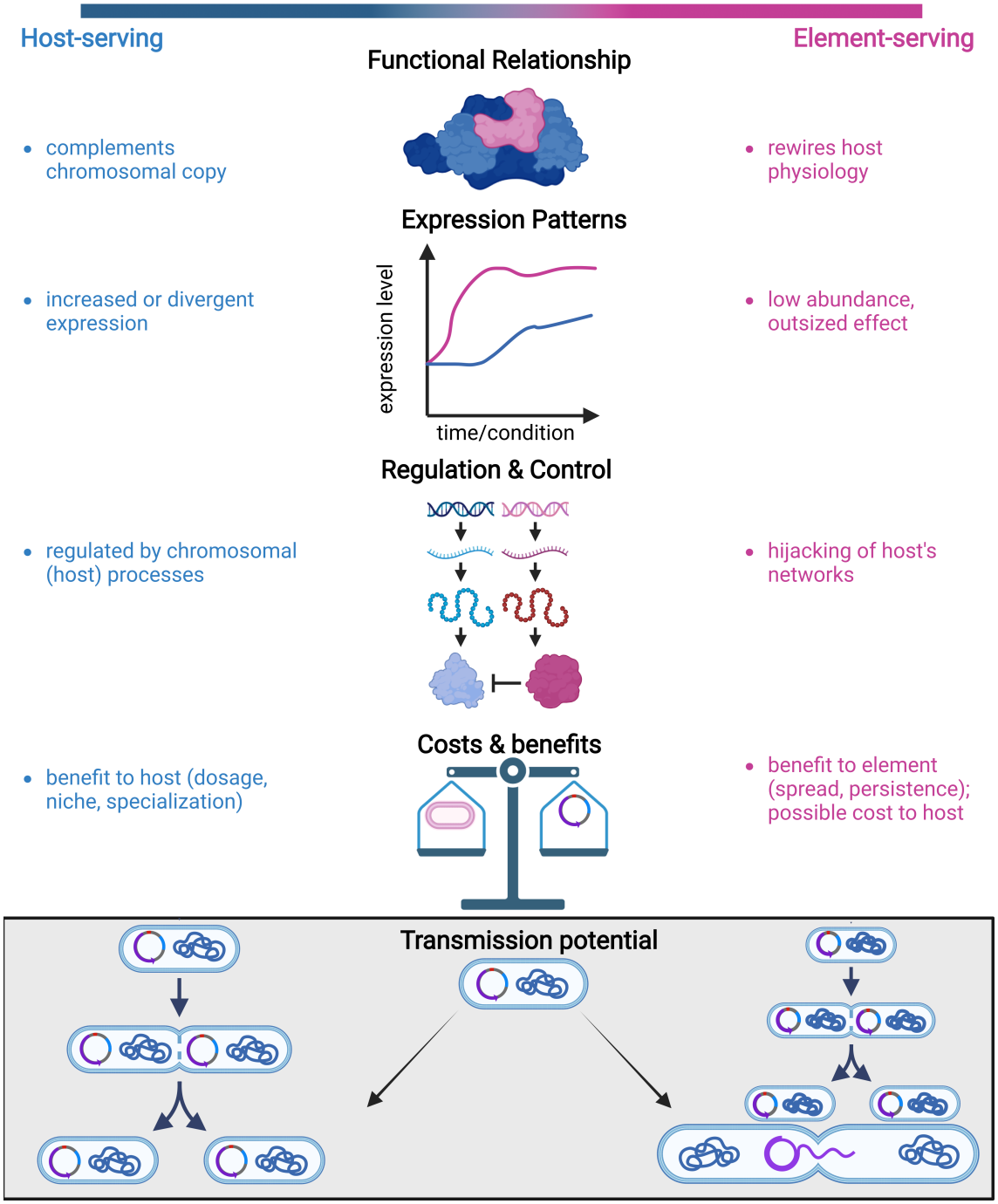


Figure 2. Features indicating whether a mobile-encoded copy of a core gene favors the host or the mobile element. Five features of the interaction between a mobile-encoded copy and its chromosomal counterpart, arranged along a spectrum from host-serving (blue, left) to element-serving (magenta, right). *Functional relationship:* the mobile copy may complement the chromosomal copy or rewire host physiology, for example as a minority subunit (pink) in a complex made mostly of host subunits (blue). *Expression patterns:* the mobile copy may be expressed at an increased level or with a divergent profile, or may be present at low abundance yet have an outsized effect. The curves are illustrative expression

profiles of the mobile (magenta) and chromosomal (blue) copies over time or across conditions. *Regulation and control:* the mobile copy may be regulated by chromosomal (host) processes or may hijack host regulatory networks, for example by its product inhibiting the chromosomal product. *Costs and benefits:* the balance of host (bacterium) and element (plasmid) benefits and costs, with benefits to the host through dosage, niche-specific benefit or specialization, or benefits to the element through spread and persistence, possibly at a cost to the host. *Transmission potential (shaded box):* a copy on a mainly vertically inherited carrier (left) is passed on to daughter cells, whereas a copy on a carrier that is also horizontally transmitted (right) can additionally spread to other cells. Transmission is the most indicative feature because copies that promote horizontal transfer at the expense of the host suggest element-serving. The features are not mutually exclusive, and individual copies may fall between the two extremes. Created with BioRender.com.

How can auxiliary copies have an effect?

A mobile-encoded copy of a core gene usually enters a cell that already contains an established chromosomal pool of the same product. Whether the incoming copy has an effect then depends first on how much of it is made: more for multicopy carriers or strongly expressed copies, less for single-copy megaplasmids or weakly expressed ones. If abundance is high, the mobile copy can contribute substantially through dosage alone. If it is low, its contribution should be only proportional to its share of the total. However, **genetic dominance** means that a minority variant can disproportionately shape the overall outcome. If the chromosomal and mobile products co-assemble into oligomers or compete for shared molecular partners, substrates, or regulatory targets, a lower-dose mobile product can exert an outsized effect by impairing or redirecting mixed oligomers, or by competitively sequestering a limiting partner [42]. Whether it does so depends on its share of the total pool, the stoichiometry and assembly rules of the complex, and whether co-assembly occurs. Cotranslational or allele-specific assembly can limit mixed-complex formation and thereby buffer dominant-negative effects [43].

Paralogue interference is the best-understood way in which dominance can arise between duplicated genes. Paralogues of homomeric proteins can form mixed heteromers, so a deleterious mutation in one paralogue can sequester wild-type subunits of the other into inactive or partially inactive complexes, reducing the activity of the unmutated copy. This resembles a dominant-negative effect, but occurs between paralogous loci rather than alleles of one locus [31,44]. For MGEs, this mechanism provides a plausible route by which a

foreign copy could modulate host proteins: not by replacing them, but by participating in, and potentially altering, their functional complexes. A divergent mobile-encoded subunit would play the part of the mutant paralogue. Després *et al.* showed experimentally that duplicated homodimeric proteins can acquire complementary mutations that abolish the function of each homodimer but preserve a functional heterodimer, demonstrating how interface divergence can generate dependence between duplicate products through constructive neutral evolution [45]. Applied to MGEs, the same physical logic suggests that a mobile-encoded paralogue could influence host physiology without replacing or silencing its chromosomal counterpart if it co-assembles with that product and changes the activity, abundance, or target specificity of the mixed complex. Alternatively, penetrance of the trait encoded by the mobile copy may require suppression or silencing of the chromosomally-encoded copy, which may come about by the activity of a different mobile gene such as a specific repressor.

H-NS homologues: Sfh and Sfx

H-NS is an abundant nucleoid-associated protein that oligomerizes along AT-rich DNA and silences horizontally acquired genes in *Enterobacteria*, which helps hosts tolerate incoming foreign DNA [46]. The H-NS family also includes paralogues, such as the chromosomal StpA and several plasmid-encoded homologues. Plasmid-encoded H-NS-like proteins (Sfh) can interact with chromosomal H-NS, and Sfh encoded by pSf-R27 can heterodimerize with H-NS and StpA in vivo [46–48]. The AT-rich pSf-R27 plasmid can titrate chromosomal H-NS away from chromosomal binding sites, causing a fitness cost. The Sfh protein mitigates this disruption, apparently by acting as a backup at a subset of H-NS-regulated chromosomal loci [47,48]. This is cooperative with respect to the host H-NS system, but element-serving with respect to plasmid establishment and transmission.

A second mobile-encoded H-NS paralogue, Sfx, is encoded by the IncX2 plasmid R6K and preferentially associates with plasmid rather than chromosomal DNA [49]. It represses the R6K *vir* conjugation operon, acting with Rho to silence conjugation genes, and deletion of *sfx* increases R6K conjugation in biofilm [50]. This suggests that Sfx serves the element by regulating the timing or cost of transfer, potentially reducing expression of conjugation functions, and not by complementing chromosomal H-NS. Whereas Sfh has been interpreted as mitigating the disruption caused by an AT-rich plasmid entering a new host, Sfx selectively targets plasmid DNA and represses a plasmid-encoded transfer module. In both cases, the mobile paralogue is expected to benefit the carrier.

RsmQ in Pseudomonas

A cell may already encode many traits that could be manipulated to favor MGE transmission, such as motility and biofilm formation [51]. Homologues of secondary metabolism regulators therefore offer a powerful lever for MGE interference, potentially allowing an MGE to rewire a suite of complex cellular traits with a small number of cargo genes. The *Pseudomonas fluorescens* GacA/S two-component system is a global regulator of secondary metabolism, controlling the expression of over 700 genes, including secretion systems, biosynthesis, twitching motility and the oxidative stress response [52]. The signaling pathway operates through RsmA-like proteins: post-translational repressors that bind to mRNA to prevent translation. GacA/S activation causes transcription of small non-coding RNAs (RsmXYZ) that titrate out RsmA to enable transcript translation. Intriguingly, almost a quarter of *Pseudomonas* plasmids were found to harbor a homologue of RsmA, and experimental studies of the variant found on the environmental megaplasmid pQBR103 showed that the plasmid gene could abolish swarming motility, upregulate biofilm, and modulate cellular metabolism in *Pseudomonas fluorescens* SBW25 [53]. RsmA-like proteins dimerize for function, and the plasmid variant, named RsmQ, could both bind the sRNAs involved in GacA/S signal transduction, and heterodimerize with the chromosomally-encoded variants, suggesting at least two routes for hijacking the signaling network.

CysS in Salmonella

Translation is a core function of cellular life, and a fundamental process in translation is the 'charging' of tRNAs with their cognate amino acids. Yet tRNA synthetases are not exclusively core chromosomal proteins, with copies also found on diverse MGEs [54]. An interesting example is found in the case of the invasive nontyphoidal strain of *Salmonella*, D23580, which possesses two copies of the *cysS* cysteinyl-tRNA synthetase responsible for charging tRNAs with cysteine. One is on the chromosome, and another is on the ~85 kb plasmid pBT1, with the two alleles diverging by more than 20% nucleotide identity. Compared with the well-characterized isolate 4/74, that lacks pBT1, D23580 has markedly lower expression of chromosomal *cysS*, to the extent that D23580 experiences growth defects in minimal media if the plasmid-borne copy is disrupted [55]. Biochemical analyses suggested that the pBT1-encoded copy of *cysS* was less efficient and stable than that encoded on the chromosome [56], features that would make the cell more prone to an accumulation of uncharged tRNAs and consequent triggering of the stringent response, shutting down translation and potentially enabling the cell to deal better with stress [57]. Curiously, while

deletion of *cysS* from the plasmid resulted in reduced fitness, cells in which the entire pBT1 plasmid had been removed appeared to grow as well as the wild-type, and RNAseq supported the hypothesis that the plasmid was responsible for downregulating expression of the chromosomal *cysS* variant. Whether this reprogramming serves the plasmid directly by enhancing horizontal transmission, or indirectly by promoting host stress tolerance, remains unresolved.

metG and pWPMR2

A genome-wide association analysis of 3,745 *Shigella sonnei* isolates identified a divergent copy of *metG*, encoding methionyl-tRNA synthetase (MetRS), on pWPMR2, an SSU5-like phage-plasmid found in 650 isolates and enriched in lineages circulating in sexual transmission networks, a distribution comparable to that of resistance loci such as *mphA* and *ermB* [54]. pWPMR2 carries no known resistance determinants, but it temporally appeared before, and increased alongside, major resistance genotypes. Related plasmids have also been found in other enteric pathogens and extrachromosomal *metG* was found on plasmids and chromids in *Ralstonia* [54], suggesting that mobile *metG* is not restricted to *S. sonnei*.

The functional phenotype is not consistent with a simple dosage increase, which could implicate paralogue interference or dominance, although this has not been tested. Induction of an additional wild-type *metG* copy increased lag time, whereas the pWPMR2 variant produced diauxic growth and a persister-like subpopulation [54]. These phenotypes are more consistent with antibiotic tolerance than classical resistance [58] and echo prior work linking altered MetRS activity to persistence [59,60]. Whether the variant interacts with or modulates the chromosomal MetRS, or complements its loss, remains unresolved, as does whether it serves the phage-plasmid or the cell. If element-serving, a slower-growing, persister-prone host could survive antibiotic exposure and remain available for onward transmission, resembling other conditionally beneficial MGE-encoded traits.

Why can antagonism persist?

If mobile copies interfere with host core functions, the host is expected to counter-adapt, for example through chromosomal alleles that reduce susceptibility to interference, or by eliminating the element. Antagonism may nonetheless persist for three reasons. First, recurrent HGT reintroduces the copy into naïve hosts. Second, selection is asymmetric: the host bears the recurring cost, whereas the element benefits whenever the phenotype

improves survival or transmission. Third, conditional effects, such as the antibiotic-dependent persistence suggested for *metG*, keep routine costs modest.

Future perspectives

No single explanation is likely to apply to all core-gene homologues on secondary replicons. We distinguish three outcomes, each with corresponding evidence and predictions (**Figure 1C**).

Host-serving. Removing the copy should lower host fitness under ecologically relevant conditions, and the copies should show dosage compensation or functional specialization and compensatory evolution that reduces the cost of carriage, as in plant-associated *Pantoea* [18,19]. The expected trajectory is increasing domestication and, eventually, chromosomal capture and loss of the element [61].

Element-serving. Removing the copy should reduce element persistence or transmission with little beneficial effect on the host, and the phenotype should track the presence of the mobile allele rather than its copy number, often involving interaction with the chromosomal homologue. RsmQ on pQBR103 is a strong candidate, since it is divergent from the chromosomal copies, interacts with them, and reprograms motility and biofilm in ways that could favor transmission. The benefit to the element should exceed that to the host, and this is most plausible for mobile carriers.

Neutral or transient. Signatures are recent acquisition, patchy distribution across related lineages, little divergence from the source gene and no consistent deletion phenotype. Given deletional bias, these states should be rare over long timescales.

Distinguishing these hypotheses most directly requires deleting the mobile copy in an isogenic background with the chromosomal homologue intact and, where possible, testing whether the mobile copy complements chromosomal loss. The experiment should separately measure host growth or competition, element stability, conjugative transfer and survival under ecologically relevant stress. A copy may benefit its carrier in one host or environment and impose a cost in another, or begin as neutral cargo and become element-serving after divergence. Constructive neutral evolution can explain how a mobile copy becomes embedded in a host complex, but not necessarily why it is retained, so fitness effects must be assessed across relevant hosts and transmission conditions. Where heteromer formation is proposed, complex composition and activity should be measured with tagged chromosomal and mobile paralogues at controlled expression.

As more multipartite genomes and MGEs are sequenced, the domesticated–mobile axis can guide prioritization of candidates before functional experiments. Replicon type, copy number, mobility machinery, phyletic distribution, expression context and signs of recent acquisition can identify the most informative cases. Resolving these cases will clarify how bacteria and their MGEs continue to shape the core genome.

Acknowledgements

Thanks to Cat Thompson and Jay Hinton for positive comments on an earlier version of this manuscript. J.B. acknowledges financial support from national funds through FCT – Fundação para a Ciência e a Tecnologia under contract no. 2023.15056.TENURE.054. J.B. also acknowledges financial support from national funds through the FCT - Foundation for Science and Technology projects UID/50016/2025 (https://doi.org/10.54499/UID/50016/2025) and LA/P/0076/2020 (https://doi.org/10.54499/LA/P/0076/2020). JPJH is supported by a Medical Research Council Career Development Award (MR/W02666X/1), and work in this area in JPJH's lab is also supported by the Biotechnology and Biological Sciences Research Council (APP37189).

References

1. McInerney, J.O. *et al.* (2017) Why prokaryotes have pangenomes. *Nature Microbiology* 2, 17040
2. LeGault, K.N. *et al.* (2021) Temporal shifts in antibiotic resistance elements govern phage-pathogen conflicts. *Science* 373, eabg2166
3. Botelho, J. (2023) Defense systems are pervasive across chromosomally integrated mobile genetic elements and are inversely correlated to virulence and antimicrobial resistance. *Nucleic Acids Research* 51, 4385–4397
4. Brockhurst, M.A. *et al.* (2019) The Ecology and Evolution of Pangenomes. *Current Biology* 29, R1094–R1103
5. Lynch, M. and Marinov, G.K. (2015) The bioenergetic costs of a gene. *Proceedings of the National Academy of Sciences* 112, 15690–15695
6. Moger-Reischer, R.Z. *et al.* (2023) Evolution of a minimal cell. *Nature* 620, 122-127
7. Giovannoni, S.J. *et al.* (2014) Implications of streamlining theory for microbial ecology. *The ISME Journal* 8, 1553–1565
8. Karcagi, I. *et al.* (2016) Indispensability of Horizontally Transferred Genes and Its Impact on Bacterial Genome Streamlining. *Mol Biol Evol* 33, 1257–1269
9. Luiselli, J. *et al.* (2024) Genome Streamlining: Effect of Mutation Rate and Population Size on Genome Size Reduction. *Genome Biol Evol* 16, evae250
10. Jain, R. *et al.* (1999) Horizontal gene transfer among genomes: The complexity hypothesis. *Proceedings of the National Academy of Sciences* 96, 3801–3806
11. Cohen, O. *et al.* (2011) The Complexity Hypothesis Revisited: Connectivity Rather Than Function Constitutes a Barrier to Horizontal Gene Transfer. *Molecular Biology and Evolution* 28, 1481–1489

12. diCenzo, G.C. and Finan, T.M. (2017) The Divided Bacterial Genome: Structure, Function, and Evolution. *Microbiology and molecular biology reviews : MMBR* 81, e00019-17
13. Harrison, P.W. *et al.* (2010) Introducing the bacterial 'chromid': not a chromosome, not a plasmid. *Trends in Microbiology* 18, 141–148
14. Liu, H. *et al.* (2025) Unexplored diversity and potential functions of extra-chromosomal elements. *mSystems* 10, e0017525
15. diCenzo, G.C. *et al.* (2019) Chromids Aid Genome Expansion and Functional Diversification in the Family *Burkholderiaceae*. *Molecular Biology and Evolution* 36, 562–574
16. diCenzo, G.C. and Finan, T.M. (2015) Genetic redundancy is prevalent within the 6.7 Mb *Sinorhizobium meliloti* genome. *Molecular Genetics and Genomics* 290, 1345–1356
17. diCenzo, G.C. *et al.* (2018) Robustness encoded across essential and accessory replicons of the ecologically versatile bacterium *Sinorhizobium meliloti*. *PLOS Genetics* 14, e1007357
18. Romero Picazo, D. *et al.* (2025) Evolution of the Plant-Associated *Pantoea* was Accompanied by Plasmid Domestication Events. *Molecular Biology and Evolution* 42, msaf273
19. Vichare, S. *et al.* (2026) Regulatory and metabolic integration accompany plasmid domestication during adaptation to a plant-associated lifestyle. bioRxiv, 2026.09.30.755635
20. Sonnenberg, C.B. and Haugen, P. (2021) The *Pseudoalteromonas* multipartite genome: distribution and expression of pangene categories, and a hypothesis for the origin and evolution of the chromid. *G3 Genes|Genomes|Genetics* 11, jkab256
21. Ren, Z. *et al.* (2022) Conformation and dynamic interactions of the multipartite genome in *Agrobacterium tumefaciens*. *Proceedings of the National Academy of Sciences* 119, e2115854119
22. Sonnenberg, C.B. and Haugen, P. (2023) Bipartite Genomes in *Enterobacterales*: Independent Origins of Chromids, Elevated Openness and Donors of Horizontally Transferred Genes. *Int J Mol Sci* 24, 4292
23. Riccardi, C. *et al.* (2023) Independent origins and evolution of the secondary replicons of the class Gammaproteobacteria. *Microbial Genomics* 9, 001025
24. Ostermayer, J. *et al.* (2025) Taming wild replicons: evolution and domestication of large extrachromosomal replicons. *Current Opinion in Microbiology* 88, 102657
25. diCenzo, G.C. *et al.* (2014) Examination of Prokaryotic Multipartite Genome Evolution through Experimental Genome Reduction. *PLoS Genet* 10, e1004742
26. Zhong, Y. and Wang, T. (2026) Informational architecture organizes plasmid genomes. bioRxiv, 2026.09.01.748738
27. Czarnecki, J. *et al.* (2026) Large extrachromosomal replicons are widespread across bacterial lineages and show coordinated replication termination and spatial coupling with the chromosome. *Nature Communications* 17, 5962
28. Fournes, F. *et al.* (2018) Replicate Once Per Cell Cycle: Replication Control of Secondary Chromosomes. *Frontiers in microbiology* 9, 1833
29. Horne, T. *et al.* (2023) How do interactions between mobile genetic elements affect horizontal gene transfer? *Current Opinion in Microbiology* 73, 102282
30. Ares-Arroyo, M. *et al.* (2024) Hijackers, hitchhikers, or co-drivers? The mysteries of mobilizable genetic elements. *PLOS Biology* 22, e3002796
31. Cisneros, A.F. *et al.* (2026) Evolutionary causes and consequences of gene duplication. *Nature Reviews Genetics* 27, 512–529
32. Wang, P. *et al.* (2025) Deciphering gene redundancy in prokaryotic genomes provides evolutionary insights for pathogenicity and its roles in clinical infections. *Nature Communications* 16, 10797
33. Hu, X.-P. *et al.* (2025) Most bacterial gene families are biased toward specific chromosomal positions. *Science* 388, 186–191

34. Anda, M. *et al.* (2015) Bacterial clade with the ribosomal RNA operon on a small plasmid rather than the chromosome. *Proceedings of the National Academy of Sciences* 112, 14343–14347
35. Anda, M. *et al.* (2023) Bacteria can maintain rRNA operons solely on plasmids for hundreds of millions of years. *Nature Communications* 14, 7232
36. Yao, Y. *et al.* (2022) Intra- and interpopulation transposition of mobile genetic elements driven by antibiotic selection. *Nature Ecology and Evolution* 6, 555–564
37. Sheridan, P.O. *et al.* (2020) Gene duplication drives genome expansion in a major lineage of Thaumarchaeota. *Nature Communications* 11, 5494
38. Storz, J. (2009) Gene duplication and the resolution of adaptive conflict. *Heredity (Edinb)* 102, 99–100
39. Castellanos, M. del P. *et al.* (2024) The roles of gene duplications in the dynamics of evolutionary conflicts. *Proceedings of the Royal Society B: Biological Sciences* 291, 20240555
40. Zheng, J. *et al.* (2015) Plasmids are vectors for redundant chromosomal genes in the *Bacillus cereus* group. *BMC Genomics* 16, 6
41. Wein, T. *et al.* (2021) Essential gene acquisition destabilizes plasmid inheritance. *PLOS Genetics* 17, e1009656
42. Gerasimavicius, L. *et al.* (2022) Loss-of-function, gain-of-function and dominant-negative mutations have profoundly different effects on protein structure. *Nature Communications* 13, 3895
43. Badonyi, M. and Marsh, J.A. (2023) Buffering of genetic dominance by allele-specific protein complex assembly. *Science Advances* 9, eadf9845
44. Cisneros, A.F. *et al.* (2026) Paralog interference contributes to the preservation of genetic redundancy. *Current Biology* 36, 1373-1384.e10
45. Després, P.C. *et al.* (2024) Compensatory mutations potentiate constructive neutral evolution by gene duplication. *Science* 385, 770–775
46. Fitzgerald, S. *et al.* (2020) Redefining the H-NS protein family: a diversity of specialized core and accessory forms exhibit hierarchical transcriptional network integration. *Nucleic Acids Res* 48, 10184–10198
47. Doyle, M. *et al.* (2007) An H-NS-like stealth protein aids horizontal DNA transmission in bacteria. *Science* 315, 251–252
48. Dillon, S.C. *et al.* (2010) Genome-wide analysis of the H-NS and Sfh regulatory networks in *Salmonella Typhimurium* identifies a plasmid-encoded transcription silencing mechanism. *Molecular Microbiology* 76, 1250–1265
49. Wang, A. *et al.* (2024) Evolutionary and functional divergence of Sfx, a plasmid-encoded H-NS homolog, underlies the regulation of IncX plasmid conjugation. *mBio* 0, e02089-24
50. Wang, B. *et al.* (2026) Selective targeting of a histone-like silencer Sfx to the R6K conjugal transfer operon. *Nucleic Acids Res* 54, gkag583
51. Madsen, J.S. *et al.* (2012) The interconnection between biofilm formation and horizontal gene transfer. *FEMS Immunol Med Microbiol* 65, 183–195
52. Cheng, X. *et al.* (2013) The Gac regulon of *Pseudomonas fluorescens* SBW25. *Environmental Microbiology Reports* 5, 608–619
53. Thompson, C.M.A. *et al.* (2023) Plasmids manipulate bacterial behaviour through translational regulatory crosstalk. *PLOS Biology* 21, e3001988
54. Tam, Y.L. *et al.* (2024) Phage-plasmid borne methionine tRNA ligase mediates epidemiologically relevant antimicrobial persistence. medRxiv, 2024.10.27.24316207
55. Canals, R. *et al.* (2019) Adding function to the genome of African *Salmonella Typhimurium* ST313 strain D23580. *PLOS Biology* 17, e3000059
56. Canals, R. *et al.* (2019) The fitness landscape of the African *Salmonella Typhimurium* ST313 strain D23580 reveals unique properties of the pBT1 plasmid. *PLoS Pathog* 15, e1007948

57. Urwin, L. *et al.* (2024) Microbial Primer: What is the stringent response and how does it allow bacteria to survive stress? *Microbiology* 170, 001483
58. Fridman, O. *et al.* (2014) Optimization of lag time underlies antibiotic tolerance in evolved bacterial populations. *Nature* 513, 418–421
59. Wood, W.N. *et al.* (2024) Methionyl-tRNA synthetase synthetic and proofreading activities are determinants of antibiotic persistence. *Front. Microbiol.* 15
60. Khare, A. and Tavazoie, S. (2020) Extreme Antibiotic Persistence via Heterogeneity-Generating Mutations Targeting Translation. *mSystems* 5
61. MacLean, R.C. *et al.* (2025) Chromosomal capture of beneficial genes drives plasmids toward ecological redundancy. *The ISME Journal* 19, wraf091