

RESEARCH ARTICLE

# Selective advantages favour high genomic AT-contents in intracellular elements

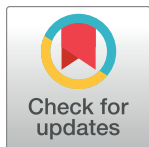
Anne-Kathrin Dietel<sup>1</sup>, Holger Merker<sup>1</sup>, Martin Kaltenpoth<sup>2#a\*</sup>, Christian Kost<sup>1#b\*</sup>

**1** Experimental Ecology and Evolution Research Group, Department of Bioorganic Chemistry, Max Planck Institute for Chemical Ecology, Jena, Germany, **2** Insect Symbiosis Research Group, Max Planck Institute for Chemical Ecology, Jena, Germany

#a Current address: Department of Evolutionary Ecology, Johannes Gutenberg-University, Mainz, Germany

#b Current address: Department of Ecology, University of Osnabrück, Osnabrück, Germany

\* [mkaltenpoth@uni-mainz.de](mailto:mkaltenpoth@uni-mainz.de) (MK); [christian.kost@biologie.uni-osnabrueck.de](mailto:christian.kost@biologie.uni-osnabrueck.de) (CK)



## Abstract

Extrachromosomal genetic elements such as bacterial endosymbionts and plasmids generally exhibit AT-contents that are increased relative to their hosts' DNA. The AT-bias of endosymbiotic genomes is commonly explained by neutral evolutionary processes such as a mutational bias towards increased A+T. Here we show experimentally that an increased AT-content of host-dependent elements can be selectively favoured. Manipulating the nucleotide composition of bacterial cells by introducing A+T-rich or G+C-rich plasmids, we demonstrate that cells containing GC-rich plasmids are less fit than cells containing AT-rich plasmids. Moreover, the cost of GC-rich elements could be compensated by providing precursors of G+C, but not of A+T, thus linking the observed fitness effects to the cytoplasmic availability of nucleotides. Accordingly, introducing AT-rich and GC-rich plasmids into other bacterial species with different genomic GC-contents revealed that the costs of G+C-rich plasmids decreased with an increasing GC-content of their host's genomic DNA. Taken together, our work identifies selection as a strong evolutionary force that drives the genomes of intracellular genetic elements toward higher A+T contents.

## OPEN ACCESS

**Citation:** Dietel A-K, Merker H, Kaltenpoth M, Kost C (2019) Selective advantages favour high genomic AT-contents in intracellular elements. *PLoS Genet* 15(4): e1007778. <https://doi.org/10.1371/journal.pgen.1007778>

**Editor:** Melanie Blokesch, Swiss Federal Institute of Technology Lausanne (EPFL), SWITZERLAND

**Received:** October 17, 2018

**Accepted:** March 27, 2019

**Published:** April 29, 2019

**Copyright:** © 2019 Dietel et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The doi for our data is [10.5061/dryad.p933f2c](https://doi.org/10.5061/dryad.p933f2c).

**Funding:** This work was supported by the International Max-Planck Research School "Exploration of Ecological Interactions with Molecular and Chemical Techniques" (AKD, <http://imprs.ice.mpg.de/ext/index.php?id=home>), the Volkswagen Foundation (CK, I/85 290, <https://www.volkswagenstiftung.de/en>), the German Research Foundation (CK: DFG SFB 944/ P19; MK: DFG KA2846/2-2, <http://www.dfg.de/en/index.jsp>), and the Max-Planck Society (AKD, HM, MK, CK,

## Author summary

Genomes of endosymbiotic bacteria are commonly more AT-rich than the ones of their free-living relatives. Interestingly, genomes of other intracellular elements like plasmids or bacteriophages also tend to be richer in AT than the genomes of their hosts. The AT-bias of endosymbiotic genomes is commonly explained by neutral evolutionary processes. However, since A+T nucleotides are both more abundant and energetically less expensive than G+C nucleotides, an alternative explanation is that selective advantages drive the nucleotide composition of intracellular elements. Here we provide strong experimental evidence that intracellular elements, whose genome is more AT-rich than the genome of the host, are selectively favoured on the host level. Thus, our results emphasize the importance of selection for shaping the DNA base composition of extrachromosomal genetic elements.

<https://www.mpg.de/en>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Bacterial genomes exhibit a considerable amount of variation in their nucleotide composition (G+C versus A+T), ranging from less than 13% to more than 75% GC [1, 2]. Despite intense efforts during the past decades, the selective pressures determining the evolution and maintenance of this variation remain elusive [3]. A general pattern that emerged from sequencing the genomes of numerous taxa is that bacteria, whose survival obligately depends on a eukaryotic host (i.e. endosymbionts), display genomic AT-contents that are significantly increased in comparison to the genomes of their free-living relatives as well as their hosts' genomes [4]. Interestingly, intracellular genetic elements that permanently or transiently exist outside the bacterial chromosome, such as plasmids, viruses, phages, and insertion sequence (IS) elements, are also usually characterized by a significantly higher AT-content than the genome of their host [5, 6]. Finding that the nucleotide composition of these very different elements is consistently biased in the same direction suggests similar evolutionary mechanisms operate to produce this pattern.

While less attention has been paid to extrachromosomal genetic elements such as plasmids and bacteriophages, two main hypotheses have been put forward to explain the biased nucleotide composition of obligate intracellular bacteria: First, high AT-contents can result from increased levels of genetic drift and mutational bias [4, 7, 8]. Genetic drift is particularly strong when the bacteria's effective population sizes are small, which is generally the case in vertically transmitted, intracellular symbionts [4]. Since the majority of DNA modifications caused by oxygen radicals (either from the environment or generated by endogenous cellular processes) lead to mispairing of DNA bases, which mostly results in GC→AT transitions and G/C→T/A transversions [9], in the long-run genetic drift is expected to increase the elements' overall AT-content.

On the other hand, it has been argued that the AT-bias of intracellular elements could be adaptive, and thus be favoured by natural selection [5]. The reasoning behind this idea is that both endosymbiotic bacteria and plasmids occupy the same ecological niche, i.e. the intracellular environment of a larger organism, and thus have access to metabolites in the host's cytoplasm [10]. This includes all nucleotides and their biochemical precursors. For the host cell, ATP and UTP nucleotides are energetically less expensive to produce than GTP and CTP nucleotides [5]. Moreover, ATP is the main energy currency used in cells and, thus, the most abundant nucleotide [11, 12]. Hence, a preferential uptake of A+T nucleotides by the respective intracellular element may impose a lower metabolic burden on its host than consumption of the more valuable G+C nucleotides. Strikingly, in both endosymbiotic bacteria and plasmids, selection tends to reduce the costs intracellular elements impose on their host (e.g. by reducing the size and transcriptional activity of the element [10]). The reason for this is that hosts harbouring metabolically 'costly' intracellular elements display a lower fitness than hosts with metabolically 'cheaper' elements [13]. As a consequence, selection acts against less fit host-symbiont/ bacteria-plasmid combinations, thereby favouring hosts that harbour metabolically cheaper intracellular elements (host-level selection). Accordingly, if AT-rich elements are metabolically 'cheaper' than GC-rich elements, hosts with more AT-rich elements should be selectively favoured.

Until now, GC-content variation in endosymbiotic bacteria has mainly been studied using comparative approaches [14, 15]. This is likely due to technical difficulties to experimentally disentangle and manipulate these complex and often obligate host-endosymbiont systems [16]. Although understandable from a methodological point of view, sequence comparisons can only reveal correlative relationships. To identify the underlying mechanistic causes,

however, manipulative experiments are required that allow to rigorously scrutinize the focal hypothesis under controlled conditions.

While the GC-content evolution in plasmids has also mainly been studied using comparative approaches ([6, 17, 18], but see [19]), plasmids and their bacterial hosts are highly tractable systems, in which a variety of different features can be experimentally manipulated [10]. Here we take advantage of the experimental tractability of plasmid-host interactions to unravel the evolutionary consequences resulting from a biased nucleotide composition of host-dependent, intracellular elements. Manipulating the plasmids' GC-content allowed us to experimentally test the hypothesis that a higher demand for G+C nucleotides due to the presence of an extrachromosomal genetic element (here: plasmid) limits host growth. Our results show indeed that bacteria that contain a GC-rich plasmid and thus have an increased demand for G+C nucleotides were less fit than bacteria containing an AT-rich plasmid. Supplying cells that contained a GC-rich plasmid with G+C nucleosides restored the fitness of host cells, while no such effect was observed for cells with AT-rich plasmids or when A+T nucleotides were supplemented to both types of plasmid-containing cells. These findings suggest that the cytoplasmic availability of G+C nucleotides limited the growth of cells with GC-rich plasmids. Moreover, introducing plasmids into different bacterial species with an increasing genomic GC-content revealed that fitness costs imposed by GC-rich plasmids decreased with increasing GC-content of the host genome. Taken together, our results provide strong experimental evidence that the commonly observed increased AT-content of host-dependent elements can be selectively favoured.

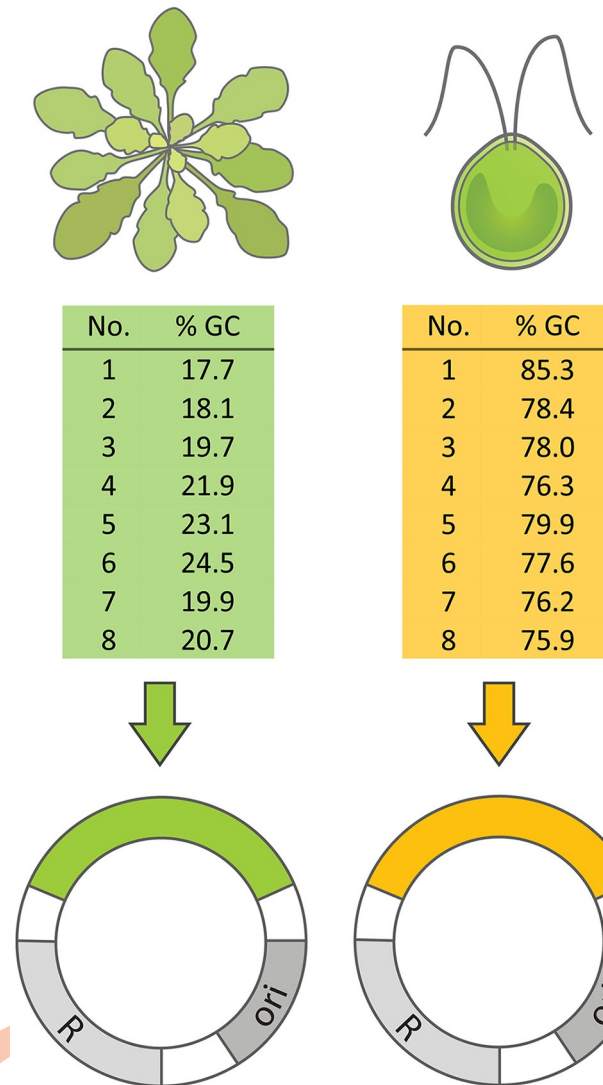
## Results

### Experimental model system

To determine if differences in the nucleotide composition of an extrachromosomal genetic element affect the fitness of the corresponding host cell, plasmids with high or low GC-contents were introduced into *Escherichia coli* cells. For this, two plasmids served as a backbone, into which eight non-coding AT- or GC-rich sequences of 1 kb in size were introduced to alter the cells' net GC-content (Fig 1). Sequences originated from eukaryotic DNA and were carefully selected, such that no genes or regulatory elements were present (see [Materials and methods](#)). In this way, chances of inadvertent gene expression were minimized, which could have resulted in additional metabolic costs. All AT-rich and GC-rich sequences were individually introduced into two different plasmid backbones that strongly differ in terms of their genetic architecture (i.e. origin of replication, copy number, and selectable marker). The resulting plasmid constructs (i.e. eight AT-rich and eight GC-rich plasmids for each of the two backbones) were used as replicates to rule out plasmid- or sequence-specific effects. This system allowed us to study the fitness consequences resulting from intracellular elements that differed in their genomic nucleotide composition in otherwise isogenic bacterial cells.

### *E. coli* cells containing AT-rich plasmids are fitter than cells containing GC-rich plasmids

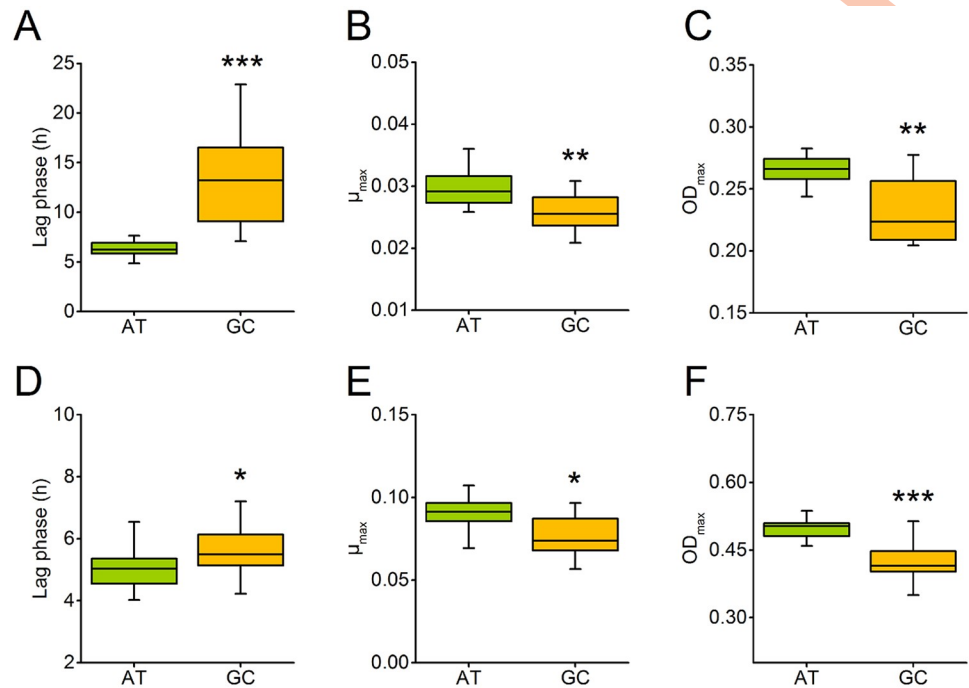
To determine whether the plasmids' nucleotide composition affected the host cell's fitness, AT-rich (i.e. cells harbouring AT-rich plasmids) and GC-rich (i.e. cells harbouring GC-rich plasmids) *E. coli* cells were grown for 24 h in minimal medium and the growth kinetics of these cultures were determined spectrophotometrically. The results of this experiment revealed that GC-rich cells harbouring the plasmid pJet1.2/blunt (hereafter: pJet) grew significantly less well than the corresponding AT-rich cells: GC-rich cells displayed a significantly extended lag-



**Fig 1. Construction of AT-rich and GC-rich plasmids.** Eight non-coding AT-rich and GC-rich sequences of 1 kb in size were amplified from the genomes of *Arabidopsis thaliana* and *Chlamydomonas reinhardtii*, respectively. Sequences were inserted into two different plasmid backbones (pJet and pBAV) to exclude plasmid-specific effects. Different plasmid-insert combinations were treated as independent replicates. All plasmids contained an origin of replication (ori), a resistance cassette (R), and one of the inserts (AT-rich DNA: green label, GC-rich DNA: orange label).

<https://doi.org/10.1371/journal.pgen.1007778.g001>

phase, a decreased maximum growth rate, and reached a lower maximum cell density as compared to AT-rich cells (Fig 2A–2C). Repeating the same experiments with *E. coli* strains that harboured the second plasmid backbone (i.e. pBAV1kT5gfp [20] lacking *T5gfp*; hereafter: pBAV), corroborated these results. Again, cell populations containing GC-rich plasmids displayed a significantly extended lag-phase, a decreased maximum growth rate, and reached a significantly reduced maximum cell density relative to the cognate AT-rich cells (Fig 2D–2F). Taken together, GC-rich cells grew significantly less well than AT-rich cells, indicating fitness costs resulted from the presence of GC-rich plasmids. Due to the design of the experiment, effects that might have been caused by the inserted sequences or the plasmid backbones used, could be ruled out as explanation.



**Fig 2. Cells containing GC-rich plasmids are less fit than those with AT-rich plasmids.** Growth experiments of *E. coli* harbouring AT-rich (green) or GC-rich (orange) pJet (A,B,C) or pBAV plasmids (D,E,F) were performed in minimal medium. Growth over 24 h measured as optical density at 600 nm was used to calculate fitness-relevant parameters. (A,D) Duration of lag phase, (B,E) maximum growth rate, and maximum optical density reached (C,F) after 24 h of cells harbouring either AT-rich or GC-rich plasmids. Asterisks indicate significant differences between cells containing AT-rich and GC-rich plasmids. Independent-samples t-test: \*\*\*  $P < 0.001$ , \*\*  $P < 0.01$ , and \*  $P < 0.05$ ,  $n = 16$ .

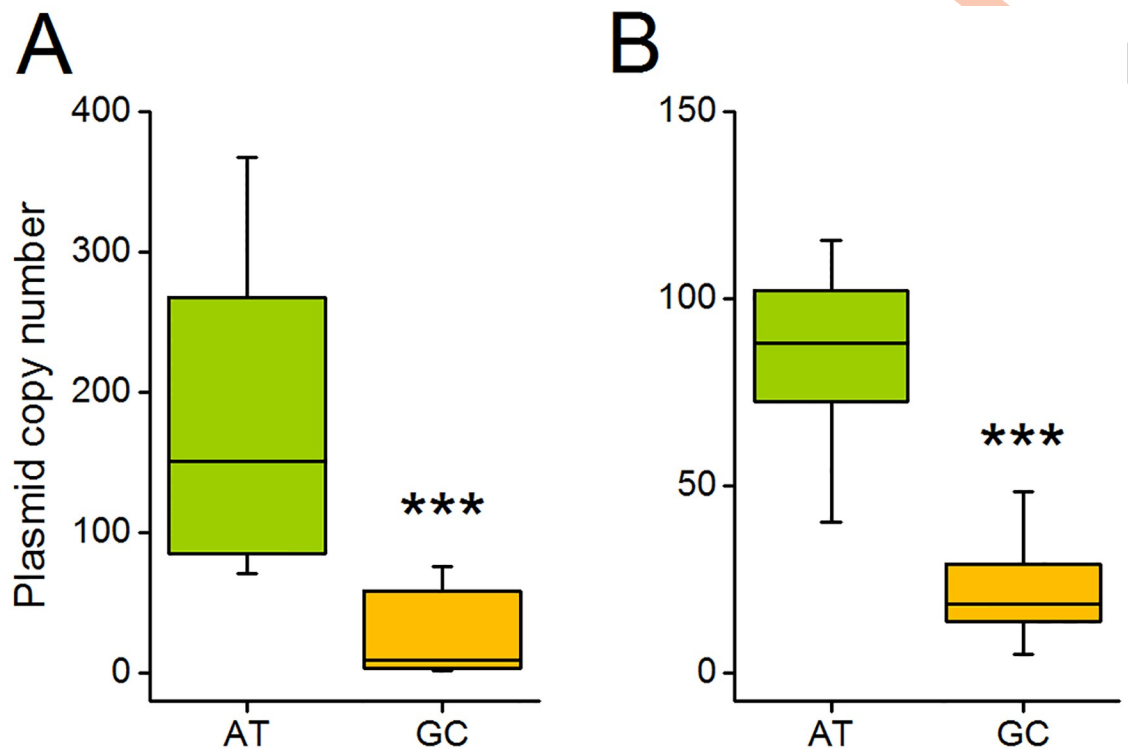
<https://doi.org/10.1371/journal.pgen.1007778.g002>

### GC-rich plasmids have a lower copy number than AT-rich plasmids

The copy number of plasmids is usually genetically determined by replication-control mechanisms that are encoded by the plasmid itself [21]. Nevertheless, the copy number of a single plasmid can vary, depending on the extent of metabolic costs it imposes on its host. For example, the copy number of a plasmid has been shown to decrease with increasing length or gene content of the accessory region [13]. Accordingly, if a GC-rich plasmid imposes a higher metabolic cost on its host than an AT-rich plasmid, it should also be present in a lower copy number than the less costly AT-rich plasmid. This hypothesis was tested by quantifying the copy numbers of both plasmids harbouring AT-rich or GC-rich inserts via quantitative PCR (qPCR). Indeed, qPCR analyses revealed that the copy number of both plasmid backbones analysed were drastically reduced when plasmids contained GC-rich inserts relative to plasmids with AT-rich inserts (Fig 3). This pattern was consistently observed over the entire growth cycle of the focal cells (S1 Fig). Together, these results further corroborate that GC-rich plasmids likely impose a higher metabolic burden on their host cells than AT-rich plasmids, thus resulting in a strongly reduced copy number.

### The growth difference between AT-rich versus GC-rich cells is not due to a differential susceptibility to antibiotics

The decreased copy number of GC-rich plasmids relative to AT-rich plasmids could make the corresponding cells more susceptible to the antibiotic, which has been used in previous



**Fig 3. GC-rich plasmids have a lower copy number than AT-rich plasmids.** Plasmid copy number per bacterial chromosome equivalent was assessed by quantitative real-time PCR of *E. coli* cultures after 24 h of growth. Shown are copy numbers of AT-rich (green) and GC-rich (orange) (A) pJet plasmids and (B) pBAV plasmids. Independent-samples t-test: \*\*\*  $P < 0.001$ ,  $n = 16$ .

<https://doi.org/10.1371/journal.pgen.1007778.g003>

experiments to stabilize the focal plasmid. This effect could explain the reduced growth of GC-rich cells (Fig 2). To rule out this possibility, cells containing the AT-rich or GC-rich pJet or pBAV plasmids were grown with 100%, 50%, or 0% of the antibiotic concentration that had been used in the abovementioned growth experiment. Quantifying fitness-relevant growth parameters of the differentially treated cultures revealed that the previously observed growth difference between cells harbouring AT-rich and GC-rich plasmids was similarly detectable for the maximal growth rate (S2A and S2B Fig) and the duration of the lag phase (S2C and S2D Fig), yet less strongly pronounced for the maximal optical density achieved (S2E and S2F Fig). Only when cells of the antibiotic-free environment (0%) were compared, no significant difference between AT-rich or GC-rich cells was detectable for the maximal optical density reached (Independent-samples t-test,  $n = 8$ ).

Moreover, we tested whether cells lost their plasmids at the end of the 24 h growth period and whether this rate of plasmid loss depended on the antibiotic concentration in the growth medium and/or the GC-content of their plasmids. The resistance mechanisms that were used to stabilize both plasmids relied on the degradation of ampicillin and kanamycin, respectively. Hence, it is conceivable that towards the end of the growth period, antibiotic concentrations might have fallen below inhibitory levels, thus allowing cells to lose their plasmid. If GC-rich plasmids are costlier than AT-rich plasmids, the former should be lost more readily than the latter. To test this, cells from all treatments of the previous experiment were plated after 24 h of growth on agar plates that did or did not contain the respective antibiotic. Indeed, a lower proportion of cells that initially contained the GC-rich plasmids were able to grow on antibiotic-containing medium than cells with AT-rich plasmids (S2G and S2H Fig), which clearly shows

that GC-rich plasmids were lost much faster than AT-rich plasmids. This result can be explained by the higher costs of GC-rich plasmids, which lowered their copy number (Fig 2), thus increasing the risk for segregational loss. The preferential loss of GC-rich plasmids likely takes place during later stages of the exponential growth phase, since both lag phase and maximum growth rate, which are the first parameters measured over the time course of the experiment, revealed even in the absence of the antibiotic a significant difference between populations of cells containing AT-rich and GC-rich plasmids. In contrast, the maximum optical density reached under the same conditions did not differ significantly between AT-rich cells and GC-rich cells, which was most probably due to the loss of the GC-rich plasmids.

Taken together, these results clarify that the reduced growth of GC-rich cells can neither be explained by the presence of the antibiotic nor the concentration used.

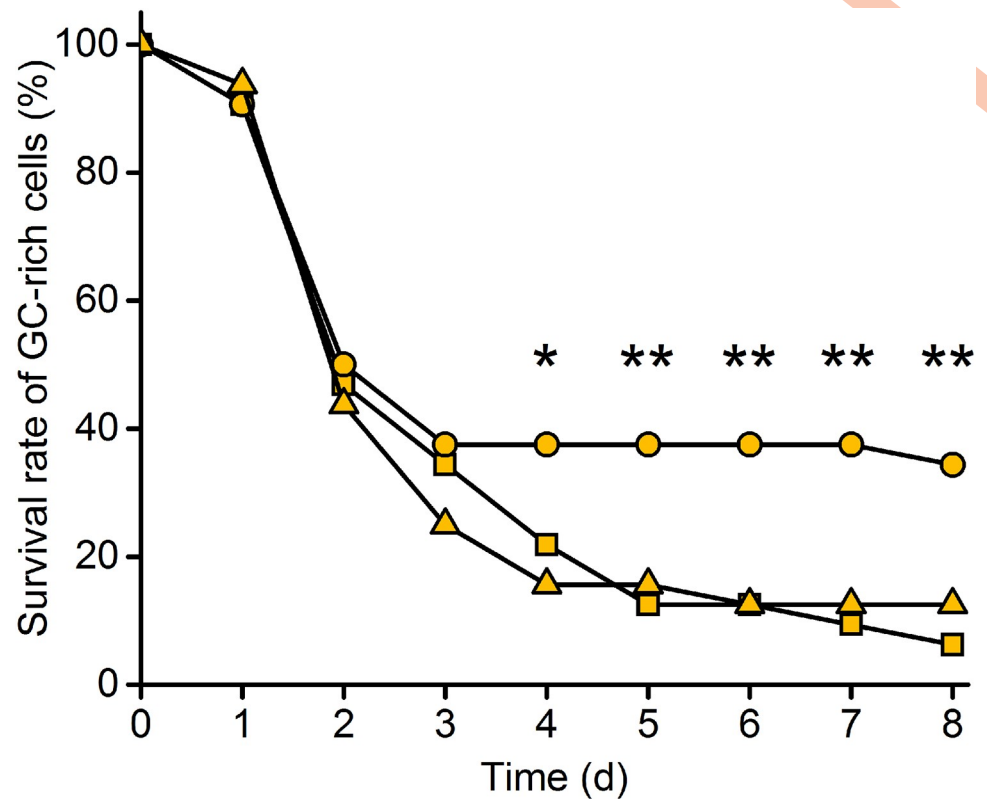
### Effect of H-NS on the growth difference between AT-rich versus GC-rich cells

Genetic information, which is acquired via horizontal gene transfer, can cause significant fitness defects by interfering with existing regulatory networks [22]. Given that horizontally acquired genes tend to be more AT-rich than the genome of the new host [23], bacteria use this deviation to recognize and silence newly acquired alien DNA. This is achieved via so-called xenogeneic silencing proteins that bind to DNA with a higher AT-content than the genome of the host and silence transcription emanating from these sequences [22]. The Histone-like Nucleoid Structuring (H-NS) protein is the best-characterised xenogeneic silencing protein, which is present in all  $\gamma$ -proteobacteria including *E. coli*. Theoretically, action of H-NS could have affected the observed fitness difference between cells containing AT-rich and GC-rich plasmids. To rule out this possibility, we deleted *hns* from the genome of *E. coli*, introduced all AT-rich and GC-rich pJet plasmids into the newly constructed mutant and repeated the previous growth experiment using the  $\Delta hns$  strain and *E. coli* wild type as host cells. The results of this experiment revealed no statistically significant effect of the presence of *hns* on the growth patterns caused by AT-rich and GC-rich plasmids (S3 Fig). However, it is well-known that deletion of *hns* results in a severe growth defect that is rapidly recovered by compensatory mutations [24]. The fact that the  $\Delta hns$  strain used in our study did not show this reduced growth phenotype, suggests additional mutations arose also during our experiments. Unfortunately, the high rate with which these mutations occur impedes a clean evaluation of the effect of H-NS. As a consequence, the experimental results involving the  $\Delta hns$  strain (S3 Fig) have to be interpreted with caution.

### Intracellular nucleotide availabilities limit cellular fitness

To determine whether the decreased growth of GC-rich cells also translates into a decreased competitive fitness relative to AT-rich cells, coculture experiments were performed with randomly chosen pairs of AT-rich and GC-rich strains. As expected, populations of GC-rich cells were readily outcompeted by AT-rich strains (Fig 4): In all 32 replicate populations, a strong decrease in the frequency of GC-rich cells relative to AT-rich cells was observed shortly after the onset of the experiment. Already after two days, GC-rich strains went extinct in 50% of all experimental populations. At the end of the experiment (i.e. after eight days), GC-rich strains were present in only two out of 32 populations (i.e. 6%), suggesting a strongly reduced fitness of GC-rich cells relative to AT-rich cells.

Two main explanations can account for this pattern. First, the growth of GC-rich cells could have been limited by the availability of G+C nucleotides within cells, while the cells containing AT-rich plasmids were less strongly affected. This hypothesis would be in line with the



**Fig 4. GC-supplementation partially rescues fitness of cells with GC-rich plasmids.** Eight pairwise cocultures of cells harbouring AT-rich and GC-rich pJet plasmids were co-inoculated (1:1 ratio) and serially propagated on a daily basis. Strains were supplied with AT or GC deoxyribonucleosides (100  $\mu$ M per nucleoside). Survival rates of GC-rich *E. coli* strains were calculated by scoring the number of populations, in which GC-rich strains were still present relative to the total number of populations within each treatment (n = 32). Each of the eight pairwise combination of strains was replicated four times. Treatment groups comprise the unsupplemented control (triangles), supply with AT-nucleosides (squares), and supply with GC-nucleosides (circles). Asterisks indicate survival rates that were significantly different from the unsupplemented control group. False discovery rate-corrected Wilcoxon signed ranks-test: \*\* P < 0.01, \* P < 0.05, n = 32 per group.

<https://doi.org/10.1371/journal.pgen.1007778.g004>

idea that the molecular composition of extrachromosomal genetic elements is strongly affected by natural selection. Second, other properties of GC-rich DNA in general could limit the growth of cells containing GC-rich plasmids. For example, an increased stability of GC-rich DNA could raise the chances for stable secondary structures that might hamper DNA replication [25] and thus slow down growth. These two effects can be distinguished in a competition experiment between AT-rich and GC-rich cells, in which cocultures are supplemented with either A+T or G+C nucleotides. If the observed decrease in fitness of GC-rich cells was truly due to a lack of G+C nucleotides within cells, then providing GC-rich cells with G+C nucleotides should enhance their growth more than supplementation with A+T. In contrast, if another mechanism such as e.g. the formation of secondary structures applies, nucleotide supplementation should not differentially affect the growth of AT-rich and GC-rich cells.

For this experiment, nucleosides instead of nucleotides were used, since *E. coli* can take up nucleosides, but not nucleotides [26]. When A+T nucleosides were externally provided, the survival rate of GC-rich cells was not significantly different from the one of the unsupplemented control group (Fig 4), indicating that these nucleosides do not generally limit cellular growth. When G+C nucleosides were added to the growing cultures, the decline of GC-rich



strains observed within the first three days of the experiment was as fast as in the untreated control group. However, from day three onwards, GC-rich cells continued to survive in ~35% of the cocultures, which represents a significant increase over both the A+T-supplemented and the untreated control group (Fig 4). This finding indicates that the low availability of G+C nucleotides limited the growth of GC-rich cells, thereby corroborating the hypothesis that the availability of nucleotides in the host cytoplasm plays a key role in shaping the GC-content of extrachromosomal genetic elements.

### The cost of extrachromosomal elements depends on the GC-content of the host chromosome

The finding that AT-rich plasmids impose a lower metabolic burden on host cells than GC-rich plasmids was obtained using *E. coli* as host, whose chromosome has a GC-content of ~50%. However, the intracellular availability of nucleotides likely depends on the base composition of the cell's chromosome, because the biosynthetic machinery of a cell is expected to have evolved in a way that it produces the required building block metabolites in optimal amounts. As a consequence, G+C nucleotides should be more abundant than A+T nucleotides in species with high genomic GC-contents, thus rendering GC-rich plasmids less costly than AT-rich plasmids.

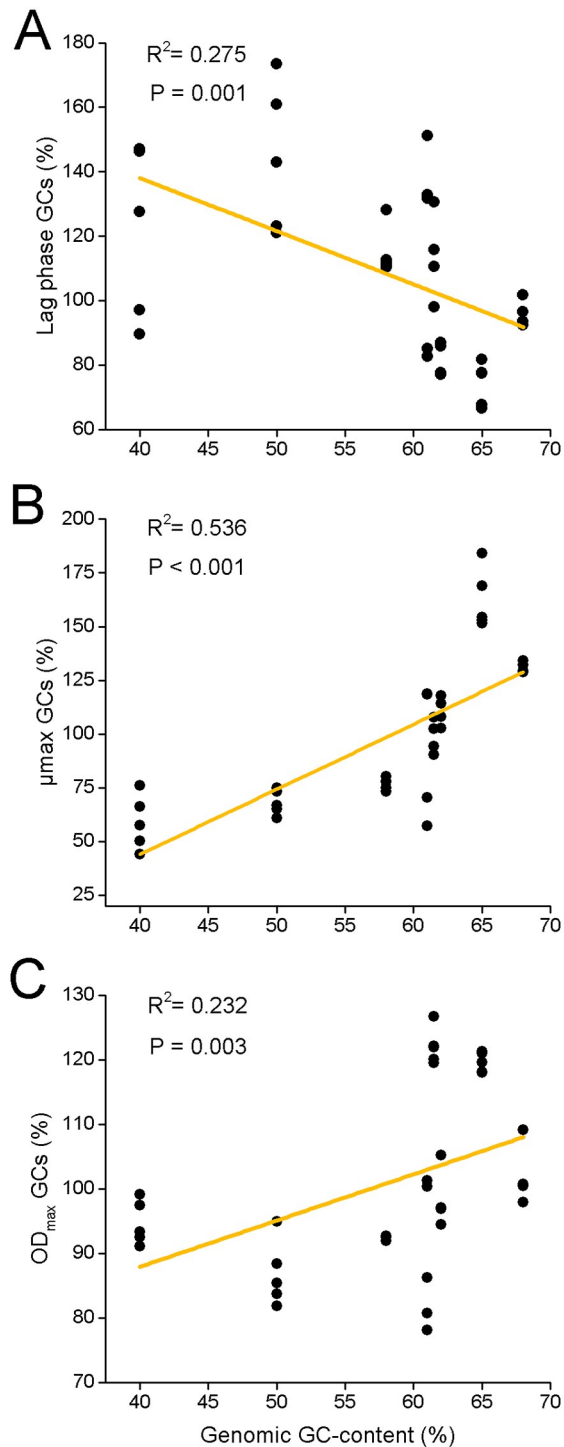
In order to test this hypothesis, eight bacterial species with a genomic GC-content ranging from 40% GC (*Acinetobacter baylyi* ADP1) to 68% GC (*Azospirillum brasilense* Tarrand) were transformed with one randomly chosen pair of AT-rich and GC-rich pBAV plasmids. In this context, it should be noted that the GC-rich plasmid was characterized by a net GC-content of 51% (AT-plasmid: 33% GC) and was thus enriched in GC-nucleotides when compared to the chromosomes of species with low to intermediate GC-contents. However, the GC-rich plasmid displayed a slightly lower GC-content when compared to the genome of GC-rich species such as *Azospirillum brasilense* and *Xanthomonas campestris*. Subsequently, fitness of different host cells carrying the GC-rich versus AT-rich plasmids was quantified spectrophotometrically as before. Indeed, the results of these experiments revealed that the burden imposed by these two plasmids depended on the genomic GC-content of the bacterial host: when the host chromosome was more AT-rich, the fitness of cells containing the AT-rich plasmid was higher than the one of cells containing the GC-rich plasmid (Fig 5). In contrast, when host cells with a more GC-rich chromosome were considered, the fitness of cells containing the GC-rich plasmid was increased relative to cells containing the AT-rich plasmid (Fig 5). Thus, the molecular composition of the host's genome strongly affected the fitness cost imposed by extrachromosomal genetic elements.

In addition, plasmid copy number measurements revealed that AT-rich plasmids were generally more abundant in species with low and intermediate genomic GC-contents (S4 Fig). However, in the two species with the highest GC-contents (i.e. *Azospirillum brasilense* and *Xanthomonas campestris*), GC-rich plasmids were present in higher copy numbers than the AT-rich plasmids.

Taken together, GC-rich plasmids imposed higher fitness costs than AT-rich plasmids on species with low to intermediate genomic GC-contents, while AT-rich plasmids were more costly for bacterial species with higher genomic GC-contents.

### Discussion

The DNA of intracellular, host-dependent elements such as bacterial endosymbionts and plasmids is generally more AT-rich than the DNA of their host's genome. In the case of bacterial endosymbionts, this pattern is commonly thought to be due to neutral evolutionary processes,



**Fig 5. The fitness cost of plasmids depends on the GC-content of both plasmid and host genome.** Growth experiments of eight bacterial species that differ in their genomic GC-content and which harboured one AT-rich or GC-rich pBAV plasmid were performed in minimal medium. Fitness parameters of cells with GC-rich plasmids were calculated relative to the respective strain harbouring the AT-rich plasmid. (A) Lag-phase, (B) maximum growth rate, and (C) maximum optical density reached after 24 h. Linear Regression analysis,  $n \geq 4$  in all cases.

<https://doi.org/10.1371/journal.pgen.1007778.g005>

such as genetic drift or a mutational bias. However, here we provide strong experimental evidence that selective advantages can contribute to this pattern. By experimentally manipulating the GC-content of plasmids and quantifying the resulting fitness consequences for the corresponding bacterial host, we show that the fitness cost of plasmids strongly depended on the nucleotide composition of both the plasmid and the host's genome. Specifically, when the genome of the host cell was characterized by intermediate to high A+T contents, GC-rich plasmids were more costly (Figs 2 and 5) and present in a lower copy number than AT-rich plasmids (Fig 3 and S1 Fig). In contrast, when the host chromosome was enriched in G+C, GC-rich plasmids were less costly (Fig 5) and present in a higher copy number than AT-rich plasmids (S1 Fig). Supplementation experiments confirmed that the observed fitness effects were indeed due to limiting pools of the corresponding nucleotides and not resulting from the GC-content of the introduced sequences *per-se* (Fig 4).

The continuous synthesis of nucleotides is crucial for DNA replication in all dividing cells. Under optimal conditions, *E. coli* cells can divide every 20 minutes, while the replication of the chromosome takes about 40 minutes [27]. To overcome this problem, multiple chromosomal copies are simultaneously generated, such that replication can keep up with the speed of cell division [28]. Interestingly, not the activity of the polymerase, but the nucleotide biosynthesis required for DNA replication seems to be limiting growth, as evidenced by the observation that the shortage of a single nucleotide drastically decreases growth [29]. Moreover, up-regulation of the ribonucleotide reductase (RNR) in yeast, which is responsible for the synthesis of dNTPs, increases the speed of the replication fork [30]. Both studies show that nucleotide synthesis is the rate-limiting step for DNA synthesis and hence also growth. By linking the availability of nucleotides to cellular fitness, these studies support the main findings reported here.

Our results are consistent with an evolutionary scenario, in which the intracellular availability of nucleotides (A+T versus G+C) depends on the genomic nucleotide composition of the bacterial cell. Over evolutionary time, a bacterial cell should establish an equilibrium, in which the biochemical machinery that produces all four nucleotides is tailored to meet the cell's requirements. This includes not only the nucleotides that are required for DNA replication, but also those that are needed to produce RNA, signalling molecules (e.g. ppGpp, cAMP), or coenzymes (e.g. ATP). In addition, energetic and stoichiometric parameters are likely to affect nucleotide production rates. For example, the biosynthetic cost to produce A+T nucleotides is less than the energy that is required to biosynthesize the same amounts of G+C [5]. An extra-chromosomal genetic element that enters such a cellular system disturbs this equilibrium by withdrawing nucleotides from intracellular pools to enable its own replication. By doing so, plasmids (and likely also intracellular bacterial endosymbionts) incur a cost to the hosting cell that depends on both the nucleotide availability in the host's cytoplasm and the amount and identity of nucleotides it consumes.

In cells of *Escherichia coli*, whose genome is characterized by a mean AT-content of ~ 50%, ATP is the most abundant nucleotide (3.5 mM ATP, 2.0 mM UTP, 1.9 mM GTP, and 1.2 mM CTP under exponential growth [11]). This is likely because of the dual function of ATP, which is not only used for RNA and DNA synthesis, but also plays a key role for transferring energy within cells. Unfortunately, to the best of our knowledge, no other study exists to date that quantified cytoplasmic nucleotide concentrations in other bacterial species, especially those that feature higher genomic GC-contents. Nevertheless, it appears reasonable to assume that cells with a higher genomic GC-content should also have an increased demand for G+C nucleotides including both ribo- and deoxyribonucleotides for RNA- and DNA-biosynthesis, respectively. This would imply higher cytoplasmic G+C levels and thus render the consumption of G+C nucleotides by GC-rich plasmids potentially less detrimental than in host cells with high genomic AT-content and thus, low cytoplasmic G+C. The observation that plasmid

copy numbers of AT-rich plasmids were higher in species with low to intermediate genomic GC-contents, but decreased in species with more GC-rich genomes (S4 Fig), is in line with this hypothesis.

Previous work that compared the GC-content of plasmids with the genomes of their corresponding bacterial and archaeal hosts revealed that on average, the GC-content of plasmids was lower than the GC-content of their host's genome (i.e. between 3% and 10%; see [5] and [6]). Moreover, also genes that have been acquired by horizontal gene transfer seem to be generally characterized by a lower G+C-content relative to the resident genome's nucleotide composition [23]. Two fundamentally different processes could cause the observed increased AT-content of plasmids and horizontally acquired genes relative to the genome of their bacterial host. First, plasmids and genes with an increased relative AT-content might be more successful in establishing in new host cells via routes of horizontal gene transfer (i.e. conjugation, transduction, transformation). In this case, the compositional difference between donor and recipient cell would act as a barrier that limits the horizontal transfer of genetic material [31]. Second, plasmids or horizontally transferred genes could evolve towards increased AT-contents relative to their host's chromosome once being present in the new cell [32]. In either way, the results presented in this study can help to explain the AT-bias of horizontally acquired genes or plasmids. If A+T nucleotides are more abundant in the cytoplasm of bacterial hosts and/or cheaper to produce than G+C nucleotides, a plasmid or a bacteriophage that is more AT-rich is more likely to successfully establish in the recipient cell, because of the reduced metabolic burden it imposes [5].

However, how would selection operate to favour extrachromosomal genetic elements with an increased AT-content? In principle, selection can operate on two different levels. First, several intracellular elements that differ in their genomic composition can compete against each other. If selective advantages are sufficiently strong, elements with an increased AT-content should outcompete elements with a lower AT-content, thus resulting in a globally increased AT-content of the entire population of intracellular elements (i.e. selection acting on the level of the extrachromosomal genetic element). Alternatively, selection can act on the level of the host (i.e. host-level selection). In this case, host individuals that contain more AT-rich elements are evolutionarily fitter than hosts containing more GC-rich intracellular elements. As a consequence of competition, hosts that contain more AT-rich elements will survive and reproduce with a higher chance, thus favouring AT-rich elements on the level of the host population in the long-run. For the given experimental set-up, our results demonstrate that host-level selection can be strong and result in an almost complete elimination of GC-rich plasmids within a few days (Fig 4). Unfortunately, it was not possible to test whether AT-rich plasmids could outcompete GC-rich plasmids within a given host cell, as plasmids using the same mode of replication cannot coexist within the same cell (i.e. when they belong to the same plasmid incompatibility group, see [33]).

Our results do not only help to understand the GC-content variation in extrachromosomal genetic elements such as plasmids and viruses, but have also significant ramifications for endosymbiotic bacteria. Similar to the interaction between bacteria and their plasmids, host-dependent bacterial cells regularly feature genomes with drastically increased AT-contents relative to the DNA of their host cell. In addition, many bacterial endosymbionts have lost the genes for an autonomous biosynthesis of all four nucleotides [34, 35]. Hence, to maintain a sufficient nucleotide-supply, cells require uptake mechanisms that allow them to import nucleotides from the host's cytoplasm. Indeed, uptake systems for nucleotide triphosphates in intracellular bacteria have been previously identified for *Rickettsia* and *Chlamydia* [36, 37], which are also known to lack specific genes essential for nucleotide biosynthesis pathways.

Bacterial endosymbionts [38], and in fact the majority of prokaryotic and eukaryotic organisms [39, 40], display a characteristic mutational bias that generally increases the genomic AT-content. In addition, newly established endosymbionts sometimes show an unexpectedly large number of polymerase slippage events that preferentially eliminate G+C-rich repetitive sequences, thus also biasing the endosymbiont's genome towards an increased A+T-content [41]. Finally, population bottlenecks that occur frequently when populations of bacterial endosymbionts are vertically transmitted from parent to offspring host, result in random assortment of bacterial genotypes that can lead to the fixation of AT-rich symbiont populations within hosts. In the early onset of an endosymbiotic interaction, all of the abovementioned processes are likely selectively neutral. However, at some point, host individuals that harbour symbionts with increased AT-contents will display a higher fitness than hosts that contain more GC-rich symbionts, particularly given the large number of endosymbiont cells in an individual host that amplify the costs associated with nucleotide requirements of the symbiont population. Due to this fitness difference, host-level selection should favour hosts with metabolically 'cheap' AT-rich symbionts. We thus believe that the evolution of AT-rich endosymbionts is likely a combination of both neutral processes such as mutational bias/ genetic drift and natural selection.

Taken together, our results provide strong experimental support for the hypothesis that the availability of nucleotides represents a significant evolutionary force that shapes the base composition of host-dependent, extra-chromosomal elements such as plasmids and likely also endosymbiotic bacteria. This interpretation is at odds with the widely-held view of drift as being the sole explanation for the AT-bias observed in the genomes of host-restricted bacteria. While our study adds an important new facet to this on-going discussion, it is most likely a combination of multiple factors that determines the nucleobase composition of bacterial genomes.

## Materials and methods

### Identification of AT-rich and GC-rich sequences

Eight AT-rich and GC-rich stretches of 1 kb in size each were identified from the AT-rich genome of *Arabidopsis thaliana* Col-0 (genome version TAIR9 v171 obtained from the Plant Genomic Database) and GC-rich genome of *Chlamydomonas reinhardtii* wild type 137 C (assembly and annotation v4 obtained from DOE Joint Genome Institute). Both annotated genomes were imported into Geneious (version 6.1.8, Biomatters, New Zealand) [42] that was used to identify AT-rich and GC-rich DNA stretches, respectively. Sequences containing simple sequence repeats in a total length of more than 30 bp were excluded to avoid the possible formation of stem-loop structures. Importantly, sequences were selected such that the chances for promoters, start codons, ribosome binding sites, or other regulatory elements were minimized.

Putative prokaryotic promoters were predicted using Softberry BPROM (softberry.com, [43]) which revealed on average two to three promoter sites within each of the AT-rich sequences, whereas none was identified for the GC-rich sequences. The chances for putative promoter regions are more likely for AT-rich DNA, since promoter regions are generally enriched in AT-content. Putative open reading frames (ORFs) were predicted using the Geneious Ver. 6.1.8. ORF finder. No or few ATG-ORFs were predicted for both AT-rich and GC-rich sequences (i.e. 0–4 per sequence), a slightly higher number of alternative ORFs starting with GTG and TGG were found for the GC-rich sequences, which is due to the higher GC-content.

Finally, sequences were tested for the presence of ribosome binding sites (RBS, i.e. Shine Dalgarno Sequences). No RBS sequences were present in the AT-rich sequences and only

three out of the eight GC-rich sequences contained one or two putative RBS (i.e. GC04, 06, and 08). The RBS identified, however, were not in close proximity of any of the ATG start codons, thus rendering translational activities unlikely. In addition, neighbouring regions of the insert position on both plasmid backbones used (i.e. a modified pJet and pBAV plasmid, see below) were analysed for the same features. For pJet, no putative promoters, ATG, GTG, or TGG ORFs or RBS were found in close proximity of the AT-rich/ GC-rich inserts (i.e. 600 bp up- and downstream). For pBAV, a single ~200 bp GTG ORF was detected ~350 bp upstream of the insert position (within the aminoglycoside-3'-phosphotransferase resistance cassette). However, due to its small size, it did not overlap with the insert sequences.

Taken together, only a small number of insert sequences contain DNA-elements required for transcription or translation. However, the chances for gene expression are minimal, since i) not all required elements are present, ii) they are in the wrong order, and iii) eight completely different sequences were used per treatment group (AT/GC).

PCR primers (S1 Table) were designed using the software Primer 3 [44] and synthesized by Metabion International AG (Martinsried, Germany).

### Amplification of AT-rich and GC-rich sequences

Genomic DNA of *A. thaliana* Col-0 was extracted following the method of Allen *et al.* [45] and of *C. reinhardtii* wild type 137 C using the protocol described by [46]. AT-rich DNA was amplified by PCR using Phusion HiFi Polymerase (Fermentas/ Thermo Fisher Scientific, Waltham, Massachusetts, US) following the manufacturer's protocol. PCR program: 98 °C 1 min, 30x: 98 °C 15 s,  $T_m$  primer 15 s, 68 °C 40 s. Elongation temperatures were decreased to 68 °C according to Su *et al.* [47], since no PCR product was observed at 72 °C. GC-rich DNA was amplified from *C. reinhardtii* using Kapa2G Robust Polymerase (Peqlab; Erlangen, Germany) following the manufacturer's recommendations for GC-rich DNA. PCR program: 95 °C 5 min, 30x: 95 °C 15 s,  $T_m$  primer 5 s, 72 °C 40 s. PCR products were purified by gel electrophoresis (1% agarose) using the NucleoSpin Extract II gel and PCR clean-up Kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany).

### Construction of AT-rich and GC-rich plasmids

Two plasmid backbones were used for the insertion of AT-rich and GC-rich DNA. The first backbone was pJet1.2/blunt (Thermo Fisher Scientific), a commercially available, high copy number plasmid of small size. The plasmid carries a pMBI\* origin of replication and encodes a beta-lactamase (*bla*) that confers resistance to ampicillin, which was used to select for plasmid-containing cells. AT-rich and GC-rich sequences were inserted into the blunt-end multiple cloning site of pJet using the pJet1.2/blunt Cloning Kit (Thermo Fisher Scientific) lacking the  $P_{lacUV5}$  promoter. Plasmids were transformed into chemically competent *E. coli* TOP10 cells (Invitrogen, Thermo Fisher Scientific) using the heat shock method [48]. Transformed colonies were screened for the respective insert using the Colony Fast-Screen Kit (Epicentre; Madison, Wisconsin, USA) following the manufacturer's instructions. Plasmids of selected transformants were sequenced at MWG Eurofins (Ebersberg, Germany).

To validate the experimental results and exclude plasmid-specific effects, all AT-rich and GC-rich sequences were additionally inserted into a second, high copy number plasmid, a modified pBAV1kT5-*gfp* [20] (ordered from Addgene <https://www.addgene.org/>; Cambridge, Massachusetts, US). This plasmid uses a different replication system (i.e. repA-mediated replication) and encodes a different selectable marker, aminoglycoside-3'-phosphotransferase (*aph* (3')), which confers resistance to the antibiotic kanamycin. The gene encoding the green fluorescent protein present on the plasmid was not needed for this study and hence removed by

digesting the plasmid with *NotI* (Thermo Fisher Scientific). Blunt ends were generated and clean-up was carried out as described above. AT-rich and GC-rich sequences were inserted into the same position. The resulting plasmids were transformed into chemically competent *E. coli* TOP10 cells. Transformants were sequenced in order to validate loss of the T5-*gfp* cassette and successful insertion of the AT-rich and GC-rich sequences. In the main text, the modified plasmid lacking T5-*gfp* is denoted as pBAV instead of pBAV1Kt5-*gfp*.

### Bacterial strains

All AT-rich plasmids were transformed into *E. coli* BW25113 Ara- [49], whereas GC-rich plasmids were transformed into *E. coli* BW25113 Ara+ [50], respectively that were made chemically competent using the rubidium chloride method [48]. The  $\Delta$ *ara* mutation renders the strain unable to catabolize arabinose. Both strains can be phenotypically distinguished when plated on tetrazolium arabinose indicator plates, on which *E. coli* BW25113 (Ara+) forms white and BW25113  $\Delta$ *ara* (Ara-) red colonies [51, 52]. The arabinose marker is selectively neutral under the cultivation conditions used in this study (independent-samples t-test:  $P > 0.05$ ,  $n = 8$ ). This phenotypic marker was used to distinguish both strains when grown in coculture.

*E. coli* BW25113  $\Delta$ *hns* was derived from the Keio collection (strain JW1225, [49]) and cured from the kanamycin-resistance cassette as described [53]. Subsequently, cells were made chemically competent as above and used to transform all AT- and GC-rich pJet plasmids.

### Culture conditions and growth kinetics

All experiments were performed in M9 minimal medium [54], which was complemented with 2 mM MgSO<sub>4</sub>, 0.1 mM CaCl<sub>2</sub>, and 5 g l<sup>-1</sup> Glucose (Sigma, St. Louis, Missouri, USA). For pBAV-containing strains, 0.25% Casamino acids (Sigma) were added to promote growth. Precultures were prepared by streaking genotypes from glycerol stocks on Lysogeny Broth (LB, Sigma) agar plates (Thermo Fisher Scientific), which were incubated overnight (16 h) at 37 °C. Subsequently, single colonies were picked and inoculated into 0.8 ml of M9 medium in a 96-deepwell plate (Eppendorf, Hamburg, Germany), which was then incubated overnight at 37 °C under shaking conditions. To ensure plasmid maintenance, 100 µg ml<sup>-1</sup> ampicillin or 50 µg ml<sup>-1</sup> kanamycin (Sigma) were always added to the culture media for pJet- and pBAV-harbouring strains, respectively. Next, optical densities (OD determined at a wavelength of 600 nm) of all precultures were measured in a microwell platereader (Spectramax 250, Molecular Devices; Sunnyvale, USA) using a 96-well plate (Nunc, Fisher Scientific GmbH; Schwerte, Germany) with a culture volume of 200 µl. The OD<sub>600nm</sub> of each culture was adjusted to 0.001. Growth kinetic assays were performed in the same instrument. Using a culture volume of 200 µl, growth was measured as absorbance at 600 nm every 5 min at 37 °C for 24 h. Cultures were shaken for 3 min after each and 15 s prior to each measurement. Fitness-related growth parameters (i.e. lag phase, maximum growth rate, and maximal OD<sub>600nm</sub>) were calculated using Magellan 7.1 SP 1 software (Magellan Software GmbH; Dortmund, Germany). Growth experiments using *E. coli* BW25113  $\Delta$ *hns* harbouring AT-rich (i.e. pJet AT01-08) or GC-rich (i.e. pJet GC01-08) plasmids were performed and analysed as described above.

For 24 h-copy number experiments, four *E. coli* strains harbouring AT-rich (i.e. AT01-04) and GC-rich plasmids (i.e. GC01-04) were chosen. Cells were precultured as stated above and then inoculated in 200 ml Erlenmeyer flasks containing 20 ml M9 medium supplemented with 100 µg ml<sup>-1</sup> ampicillin (initial inoculum: OD<sub>600nm</sub> = 0.001). Subsequently, the OD<sub>600nm</sub> was monitored in regular intervals. Samples for plasmid copy number determination were taken every 6 h and stored by adding the same volume of 40% w/v glycerol at -80 °C until real time PCR measurements were performed.

## Plasmid copy number determination

Plasmid copy numbers were determined using quantitative real-time PCR (qPCR) following a previously described method [55]. For this, monocultures of cells were harvested after 24 h of growth. Plasmid copy numbers were determined by calculating both the total number of chromosomal and plasmid copies in each sample. Chromosome copy numbers were determined using a primer pair targeting the single copy gene *dxs* (1-deoxy-D-xylulose-5-phosphate synthase). For total plasmid numbers per sample, primer pairs targeting the respective antibiotic resistance gene were used (i.e. either *bla* on pJet or *aph(3')* on pBAV; genes and primer details see S2 Table). Bacterial cultures from both monocultures were diluted ~1:100 and used for qPCR. QPCR was performed using the Brilliant III Ultra-Fast SYBR Green QPCR Master Mix (Agilent Technologies; Santa Clara, US) in a BioRad CFX96 thermocycler (Hercules, California, USA) according to the manufacturer's instructions. PCR program: 10 min 95 °C, 40x: 30 s 95 °C, 20 s 61 °C, 30 s 72 °C. Standard curves were prepared by 10-fold dilutions of both isolated plasmids and bacterial cells ( $R^2$  of all standard curves: > 0.99). Plasmid numbers per ng plasmid DNA template were calculated using an online calculator (<http://cels.uri.edu/gsc/cndna.html>, Andrew Staroscik, Genomics & Sequencing Center, University of Rhode Island, Kingston, Rhode Island, USA). Cell numbers of each standard curve sample were measured using a CyFlow Space flow cytometer (Partec, Görlitz, Germany), for which cells were stained with SYBR Green (Sigma) following the manufacturer's protocol. Average plasmid copy numbers per cell were calculated from the respective standard curves ( $R^2$  of all standard curves > 0.99) by dividing total plasmid numbers by the total number of cells.

## Growth experiments and plasmid loss in the absence of selection

In order to test whether the reduced growth of cells harbouring the GC-rich plasmids was due a higher susceptibility to the supplemented antibiotic, additional growth experiments were performed in the absence of the antibiotic. For this, all AT-rich and GC-rich plasmid-containing strains (i.e. both pJet and pBAV) were precultured in M9 medium containing the respective antibiotic as described above. After that, cultures were centrifuged at 9,000 rpm for 1 min. Supernatants were discarded and cells were washed twice with fresh M9 medium in order to ensure that cultures are free from any residual antibiotic. Next, cultures were diluted as described above and subjected to one of three treatments: M9 medium containing 100%, 50%, or 0% of the antibiotic (i.e. 100  $\mu\text{g ml}^{-1}$ , 50  $\mu\text{g ml}^{-1}$ , and 0  $\mu\text{g ml}^{-1}$  ampicillin, and 50  $\mu\text{g ml}^{-1}$ , 25  $\mu\text{g ml}^{-1}$ , and 0  $\mu\text{g ml}^{-1}$  kanamycin, respectively). The standard concentration of antibiotics used (i.e. 100%) was chosen according to the provider's recommendation (i.e. pJet: Thermo Fisher Scientific, pBAV: [20]). Growth kinetics were performed and measured as described above. In order to test whether plasmids were still present in the antibiotic-free cultures after the experiment has been terminated, cells were plated on agar plates that did (i.e. 50% of the standard concentration) or did not contain antibiotics. By comparing the number of CFUs (colony forming units) on the antibiotic-containing plates relative to antibiotic-free plates, the rate of plasmid loss was determined.

## Competitive fitness and nucleoside feeding experiments

For coculture experiments, one AT-rich strain was paired with a GC-rich strain, respectively (all harbouring pJet plasmids). Eight combinations were randomly chosen (i.e. Fig 1: AT1-GC1, AT2-GC2, etc.) and each combination replicated 4 times ( $n = 32$ ). To test if the decrease in growth of GC-rich cells can be explained by their increased demand for G+C nucleotides, cocultures were grown in one of three different media: (1) pure M9 minimal medium, (2) M9 medium supplemented with A+T-nucleosides, or (3) M9 medium supplemented with



G+C-nucleosides. In this experiment, deoxyribonucleosides have been used instead of deoxyribonucleotides, because no nucleotide transport systems are known for *E. coli*, while two nucleoside transport systems (i.e. NupG and NupC), have been described [26]. Either 2'-deoxyadenosine and thymidine or 2'-deoxyguanosine and 2'-deoxycytidine (Sigma) were added to the growth medium at a final concentration of 100  $\mu\text{M}$  per nucleoside. The  $\text{OD}_{600\text{nm}}$  of all pre-cultures was adjusted to 0.0005, resulting in a final  $\text{OD}_{600\text{nm}}$  of 0.001 after mixing of cocultures. Fitness experiments were performed in a 96-deepwell plate (Nunc) with a culture volume of 0.8 ml. Cocultures were incubated at 37 °C under shaking conditions (220 rpm). 0.8  $\mu\text{l}$  (1:1,000 dilution) of all cocultures were transferred daily into fresh medium for a total of eight days. Every day, serial dilutions of all cultures were plated on TA agar plates to distinguish Ara+ (AT-rich) and Ara- strains (GC-rich) of *E. coli* BW25113 using the arabinose utilization marker as described above.

### Transformation of bacterial species with different GC-contents

One AT-rich and one GC-rich pBAV plasmid, i.e. pBAV-AT01 and pBAV-GC02, were randomly chosen to be introduced into other bacterial species differing in their genomic GC-contents. PJet could not be used for this purpose as it does not replicate in species other than *E. coli*. In contrast, pBAV has been shown to be a broad host range plasmid replicating in many other bacterial species [20]. The following species were used: *Acinetobacter baylyi* ADP1 (40% GC), *Serratia entomophila* (DSM 12358) (58% GC), *Pseudomonas protegens* (61% GC), *Pseudomonas putida* (62% GC), *Arthrobacter aurescens* (DSM 20116) (61.5% GC), *Xanthomonas campestris* (DSM 3586) (65% GC), and *Azospirillum brasilense* (DSM 1690) (68% GC). All strains were tested to be Kanamycin (50  $\mu\text{g}/\text{ml}$ ) sensitive. Both plasmids were introduced in electrocompetent cells of the above-listed species using a MicroPulser Electroporator (Bio-Rad, Hercules, California, US) with the following settings: 25  $\mu\text{F}$ , 200 mA, and 2.5 kV using 70  $\mu\text{l}$  of electrocompetent cells and 100–150 ng plasmid DNA. Colonies obtained were grown in LB medium supplemented with 50  $\mu\text{g ml}^{-1}$  kanamycin. Plasmid isolation was performed as described previously and plasmids were sequenced using plasmid-specific primers targeting the AT/GC-rich insert.

### Growth experiments and plasmid copy number determination using bacterial species with different GC-contents

All strains were grown in M9 minimal medium containing glucose, sucrose, and malate (glucose and sucrose: 5  $\text{g l}^{-1}$  each, malate: 2  $\text{g l}^{-1}$ ) as carbon source, as well as 2 mM  $\text{MgSO}_4$ , 0.1 mM  $\text{CaCl}_2$ , 45  $\mu\text{M}$   $\text{FeSO}_4$ , 0.5  $\text{mg ml}^{-1}$   $\text{NaMO}_4$ , and 0.01  $\text{mg ml}^{-1}$  Biotin (Sigma) at 28 °C and 220 rpm for 24 h. Growth experiments were performed as mentioned above using a Spectra-max plate reader. Plasmid copy numbers of different bacterial species were determined by measuring plasmid numbers via quantitative Real-Time PCR as described earlier. However, all cell numbers were quantified by Flow Cytometry instead of qPCR, since using *dxs*-specific primers did not result in DNA amplification in most of the species, either due to an altered sequence or absence of the corresponding gene. Thus, plasmid copy numbers were calculated as plasmid number per cell count (see above).

### Data analysis

Experiments were performed using plasmids that contained one of eight different AT- or GC-rich inserts. To reduce the impact of sequence-specific effects, data were analysed by treating AT-rich and GC-rich strains as replicates. In monoculture experiments, fitness-relevant parameters of AT-rich and GC-rich cells were statistically compared by two-sample

independent t-tests. In coculture experiments, survival rates of strains harbouring GC-rich plasmids were calculated by scoring the number of populations, in which the strains were present relative to the total number of populations within each treatment ( $n = 32$ ). Strains were considered to be extinct, if the number of colony forming units decreased below 2% of the total CFU counts. Survival rates between treatments were compared by performing Wilcoxon signed ranks tests. False discovery rate (FDR) was applied to P-values to correct for multiple testing [56]. Linear regression analyses were performed to correlate growth parameters of species harbouring AT/GC-rich plasmids with the species' GC-content. All statistical analyses were performed using SPSS Software (version 17.0, SPSS Inc., Chicago, IL, USA) and R Studio (Boston, USA) [57].

## Supporting information

**S1 Fig. Plasmid copy number of GC-rich plasmids is decreased over the entire growth cycle of *E. coli*.** Plasmid copy number per bacterial chromosome equivalent was assessed by quantitative real-time PCR of *E. coli* cultures at four time points during a 24 h growth period. Shown are copy numbers of pJet plasmids harbouring four randomly chosen AT-rich or GC-rich inserts, respectively. The first two sampling points (i.e. 6 h, 12 h) correspond to the beginning and mid-exponential growth phase, the third sampling point (i.e. 18 h) to the end of exponential growth phase, and the fourth sampling point (i.e. 24 h) to the beginning of stationary growth. Asterisks indicate significant differences between *E. coli* cells containing AT-rich and GC-rich plasmids within the same time point. Independent-samples t-test: \*\*\*  $P < 0.001$ , \*  $P < 0.05$ ,  $n = 4$ .

(TIF)

**S2 Fig. In the absence of selection, cells harbouring GC-rich plasmids are less fit and lose their plasmids faster than cells harbouring AT-rich plasmids.** Growth experiments of *E. coli* harbouring AT-rich (green) or GC-rich (orange) pJet (A, C, E, G) or pBAV plasmids (B, D, F, H) were performed in minimal medium supplemented with either 100%, 50%, or 0% of the respective antibiotic. (A, B) Duration of lag phase, (C, D) maximum growth rate achieved, and maximum optical density reached (E, F) during 24 h of growth. (G, H) The proportion of antibiotic-resistant cells (i.e. a measure for plasmid loss) was determined by dividing the number of colonies formed on antibiotic-containing plates with the number of colonies formed on antibiotic-free agar plates after 24 h of growth under the respective conditions. Asterisks indicate significant differences between cells containing AT-rich and GC-rich plasmids within the same treatment. Independent-samples t-test: \*\*\*  $P < 0.001$ , \*\*  $P < 0.01$ , \*  $P < 0.05$ , #  $P = 0.071$ , §  $P = 0.054$ ,  $n = 8$ .

(TIF)

**S3 Fig. The fitness of AT-rich and GC-rich cells is independent of H-NS.** Growth experiments of *E. coli*  $\Delta hns$  harbouring AT-rich (green) or GC-rich (orange) pJet plasmids were performed in minimal medium. Growth over 24 h measured as optical density at 600 nm was used to calculate fitness-relevant parameters. (A) Duration of lag phase, (B) maximum growth rate. Asterisks indicate significant differences between cells containing AT-rich and GC-rich plasmids. Independent-samples t-test: \*\*  $P < 0.01$ ,  $n = 16$ .

(TIF)

**S4 Fig. Plasmid copy number of GC-rich plasmids is increased in bacterial species with GC-rich genomes relative to AT-rich plasmids.** Copy number of GC-rich plasmids relative to AT-rich plasmids in the same bacterial species are displayed. Plasmid copy number per cell equivalent was assessed by quantitative real-time PCR and flow cytometry of all

bacterial species after 24 h of growth. Asterisks denote significant deviations from equal copy numbers of AT-rich and GC-rich plasmids in the same host environment. One sample t-test: \*\*\*  $P < 0.001$ , \*\*  $P < 0.01$ , \*  $P < 0.05$ ,  $n = 5$ .

(TIF)

**S1 Table. Non-coding AT-rich and GC-rich sequences used in this study.** AT-rich sequences (AT01-08) were amplified from the genome of *Arabidopsis thaliana* (chromosome 4), whereas GC-rich sequences (GC01-GC08) were amplified from *Chlamydomonas reinhardtii* (+chromosome 1, \*chromosome 2).

(DOCX)

**S2 Table. Primer pairs used for quantitative real-time PCR to determine the copy number of pJet and pBAV plasmids relative to the copy number of the chromosome.**

(DOCX)

## Acknowledgments

We thank the Experimental Ecology and Evolution Group, the Insect Symbiosis Group, the Department of Bioorganic Chemistry as well as Günter Theißen for helpful discussion, Colin Dale and Rahul Raghavan for critically reading an earlier version of the manuscript, as well as three anonymous reviewers for constructive criticism. Moreover, help by Eva Limpinsel and Piyali Pal Chowdhury for generating some of the strains is gratefully acknowledged. The authors would also like to thank Wilhelm Boland for his generous support.

## Author Contributions

**Conceptualization:** Anne-Kathrin Dietel, Holger Merker, Martin Kaltenpoth, Christian Kost.

**Data curation:** Anne-Kathrin Dietel.

**Formal analysis:** Anne-Kathrin Dietel.

**Funding acquisition:** Martin Kaltenpoth, Christian Kost.

**Investigation:** Anne-Kathrin Dietel, Christian Kost.

**Methodology:** Anne-Kathrin Dietel, Christian Kost.

**Project administration:** Christian Kost.

**Resources:** Christian Kost.

**Software:** Anne-Kathrin Dietel.

**Supervision:** Holger Merker, Martin Kaltenpoth, Christian Kost.

**Validation:** Anne-Kathrin Dietel.

**Visualization:** Anne-Kathrin Dietel.

**Writing – original draft:** Anne-Kathrin Dietel.

**Writing – review & editing:** Anne-Kathrin Dietel, Martin Kaltenpoth, Christian Kost.

## References

1. McCutcheon JP, Moran NA. Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. *Genome biology and evolution*. 2010; 2:708–18. <https://doi.org/10.1093/gbe/evq055> PMID: 20829280

2. Sueoka N. Directional mutation pressure and neutral molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America*. 1988; 85(8):2653–7. PMID: [3357886](#)
3. Agashe D, Shankar N. The Evolution of Bacterial DNA Base Composition. *J Exp Zool Part B*. 2014; 322(7):517–28.
4. Moran NA. Accelerated evolution and Muller’s ratchet in endosymbiotic bacteria. *Proceedings of the National Academy of Sciences of the United States of America*. 1996; 93:2873–8. PMID: [8610134](#)
5. Rocha EPC, Danchin A. Base composition bias might result from competition for metabolic resources. *Trends Genet*. 2002; 18(6):291–4. [https://doi.org/10.1016/S0168-9525\(02\)02690-2](https://doi.org/10.1016/S0168-9525(02)02690-2) PMID: [12044357](#)
6. Nishida H. Comparative analyses of base compositions, DNA sizes, and dinucleotide frequency profiles in archaeal and bacterial chromosomes and plasmids. *International journal of evolutionary biology*. 2012;2012:-.
7. Wernegreen JJ, Moran NA. Evidence for genetic drift in endosymbionts (*Buchnera*): Analyses of protein-coding genes. *Molecular biology and evolution*. 1999; 16(1):83–97. <https://doi.org/10.1093/oxfordjournals.molbev.a026040> PMID: [10331254](#)
8. Wernegreen JJ, Funk DJ. Mutation exposed: a neutral explanation for extreme base composition of an endosymbiont genome. *Journal of molecular evolution*. 2004; 59(6):849–58. <https://doi.org/10.1007/s00239-003-0192-z> PMID: [15599516](#)
9. Wang D, Kreutzer DA, Essigmann JM. Mutagenicity and repair of oxidative DNA damage: insights from studies using defined lesions. *Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis*. 1998; 400(1–2):99–115. PMID: [9685598](#)
10. Dietel A-K, Kaltentpoth M, Kost C. Convergent Evolution in Intracellular Elements: Plasmids as Model Endosymbionts. *Trends in Microbiology*.
11. Danchin A, Dondon L, Daniel J. Metabolic alterations mediated by 2-ketobutyrate in *Escherichia coli* K12. *Mol Gen Genet*. 1984; 193(3):473–8. PMID: [6369074](#)
12. Traut TW. Physiological concentrations of purines and pyrimidines. *Molecular and cellular biochemistry*. 1994; 140(1):1–22. PMID: [7877593](#)
13. Smith MA, Bidochka MJ. Bacterial fitness and plasmid loss: the importance of culture conditions and plasmid size. *Can J Microbiol*. 1998; 44(4):351–5. PMID: [9674107](#)
14. Moran NA. Microbial minimalism: Genome reduction in bacterial pathogens. *Cell*. 2002; 108(5):583–6. PMID: [11893328](#)
15. Wernegreen JJ. Endosymbiont evolution: predictions from theory and surprises from genomes. In: Mousseau TA, Fox CW, editors. *Year in Evolutionary Biology*. *Annals of the New York Academy of Sciences*. p. 16–35.
16. Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera sp* APS. *Nature*. 2000; 407(6800):81–6. <https://doi.org/10.1038/35024074> PMID: [10993077](#)
17. Shintani M, Sanchez ZK, Kimbara K. Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy. *Frontiers in microbiology*. 2015;6.
18. Almpanis A, Swain M, Gatherer D, McEwan N. Correlation between bacterial G plus C content, genome size and the G plus C content of associated plasmids and bacteriophages. *Microb Genomics*. 2018; 4(4):8.
19. Raghavan R, Kelkar YD, Ochman H. A selective force favoring increased G plus C content in bacterial genes. *Proceedings of the National Academy of Sciences of the United States of America*. 2012; 109(36):14504–7. <https://doi.org/10.1073/pnas.1205683109> PMID: [22908296](#)
20. Bryksin AV, Matsumura I. Rational design of a plasmid origin that replicates efficiently in both gram-positive and gram-negative bacteria. *PloS one*. 2010; 5(10).
21. Nordstrom K, Austin SJ. Mechanisms that contribute to the stable segregation of plasmids. *Annu Rev Genet*. 1989; 23:37–69. <https://doi.org/10.1146/annurev.ge.23.120189.000345> PMID: [2694936](#)
22. Navarre WW, Porwollik S, Wang YP, McClelland M, Rosen H, Libby SJ, et al. Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*. *Science*. 2006; 313(5784):236–8. <https://doi.org/10.1126/science.1128794> PMID: [16763111](#)
23. Daubin V, Lerat E, Perriere G. The source of laterally transferred genes in bacterial genomes. *Genome Biology*. 2003; 4(9):12.
24. Ali SS, Soo J, Rao CT, Leung AS, Ngai DHM, Ensminger AW, et al. Silencing by H-NS Potentiated the Evolution of *Salmonella*. *PLoS Pathog*. 2014; 10(11).
25. Yakovchuk P, Protozanova E, Frank-Kamenetskii MD. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Research*. 2006; 34(2):564–74. <https://doi.org/10.1093/nar/gkj454> PMID: [16449200](#)

26. Munchpetersen A, Mygind B. Nucleoside transport systems in *Escherichia coli* K12 -Specificity and regulation. *Journal of Cellular Physiology*. 1976; 89(4):551–9. <https://doi.org/10.1002/jcp.1040890410> PMID: 827549
27. Lark KG. Regulation of chromosome replication and segregation in bacteria. *Bacteriological Reviews*. 1966; 30(1):3–32. PMID: 5324649
28. Fossum S, Croke E, Skarstad K. Organization of sister origins and replisomes during multifork DNA replication in *Escherichia coli*. *Embo J*. 2007; 26(21):4514–22. <https://doi.org/10.1038/sj.emboj.7601871> PMID: 17914458
29. Barnes MH, R R. Effect of thymine limitation on chromosomal deoxyribonucleic acid synthesis in *Proteus mirabilis*. *Journal of bacteriology*. 1972; 111(3):750–7. PMID: 4559826
30. Poli J, Tsaponina O, Crabbe L, Keszthelyi A, Pantesco V, Chabes A, et al. dNTP pools determine fork progression and origin usage under replication stress. *Embo J*. 2012; 31(4):883–94. <https://doi.org/10.1038/emboj.2011.470> PMID: 22234185
31. Popa O, Dagan T. Trends and barriers to lateral gene transfer in prokaryotes. *Current Opinion in Microbiology*. 2011; 14(5):615–23. <https://doi.org/10.1016/j.mib.2011.07.027> PMID: 21856213
32. Lawrence JG, Ochman H. Amelioration of bacterial genomes: Rates of change and exchange. *Journal of molecular evolution*. 1997; 44(4):383–97. PMID: 9089078
33. Novick RP. Plasmid Incompatibility. *Microbiol Rev*. 1987; 51(4):381–95. PMID: 3325793
34. Moran NA, Wernegreen JJ. Lifestyle evolution in symbiotic bacteria: insights from genomics. *Trends Ecol Evol*. 2000; 15(8):321–6. PMID: 10884696
35. Klasson L. Evolution of minimal-gene-sets in host-dependent bacteria. *Trends in Microbiology*. 2004; 12(1):37–43. PMID: 14700550
36. Tjaden J, Winkler HH, Schwoppe C, Van der Laan M, Mohlmann T, Neuhaus HE. Two nucleotide transport proteins in *Chlamydia trachomatis*, one for net nucleoside triphosphate uptake and the other for transport of energy. *Journal of bacteriology*. 1999; 181(4):1196–202. PMID: 9973346
37. Haferkamp I, Schmitz-Esser S, Wagner M, Neigel N, Horn M, Neuhaus HE. Tapping the nucleotide pool of the host: novel nucleotide carrier proteins of *Protochlamydia amoebophila*. *Molecular microbiology*. 2006; 60(6):1534–45. <https://doi.org/10.1111/j.1365-2958.2006.05193.x> PMID: 16796686
38. Van Leuven JT, McCutcheon JP. An AT mutational bias in the tiny GC-rich endosymbiont genome of *Hodgkinia*. *Genome biology and evolution*. 2012; 4(1):24–7. <https://doi.org/10.1093/gbe/evr125> PMID: 22113795
39. Hershberg R, Petrov DA. Evidence that mutation is universally biased towards AT in bacteria. *PLoS Genet*. 2010; 6(9).
40. Hildebrand F, Meyer A, Eyre-Walker A. Evidence of Selection upon Genomic GC-Content in Bacteria. *PLoS Genet*. 2010; 6(9).
41. Clayton AL, Jackson DG, Weiss RB, Dale C. Adaptation by Deletogenic Replication Slippage in a Nascent Symbiont. *Molecular biology and evolution*. 2016; 33(8):1957–66. <https://doi.org/10.1093/molbev/msw071> PMID: 27189544
42. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012; 28(12):1647–9. <https://doi.org/10.1093/bioinformatics/bts199> PMID: 22543367
43. Solovyev V, Salamov A. *Automatic Annotation of Microbial Genomes and Metagenomic Sequences. Metagenomics and its Applications in Agriculture* ed: Nova Science Publishers; 2011.
44. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. *Methods in molecular biology* (Clifton, NJ). 2000; 132:365–86.
45. Allen GC, Flores-Vergara MA, Krasnyanski S, Kumar S, Thompson WF. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nature Protocols*. 2006; 1(5):2320–5. <https://doi.org/10.1038/nprot.2006.384> PMID: 17406474
46. Harris EH. *The Chlamydomonas Sourcebook: Introduction to Chlamydomonas and its laboratory use*. 2008.
47. Su XZ, Wu YM, Sifri CD, Wellems TE. Reduced extension temperatures required for PCR amplification of extremely A+T-rich DNA. *Nucleic Acids Research*. 1996; 24(8):1574–5. PMID: 8628694
48. Hanahan. Studies on transformation of *Escherichia coli* with plasmids. *Journal of molecular biology*. 1983; 166:557–80. PMID: 6345791
49. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*. 2006;2.

50. D'Souza G, Waschina S, Pande S, Bohl K, Kaleta C, Kost C. Less is more: Selective advantages can explain the prevalent loss of biosynthetic genes in bacteria. *Evolution*. 2014; 68(9):2559–70. <https://doi.org/10.1111/evo.12468> PMID: 24910088
51. Lenski RE. Long-term experimental evolution in *Escherichia coli*. 1. Adaptation and divergence during 2,000 generations. *The American Naturalist*. 1991; 138(6):1315–41.
52. Levin BR, Stewart FM, Chao L. Resource-limited growth, competition, and predation—A model and experimental studies with bacteria and bacteriophage. *Am Nat*. 1977; 111(977):3–24.
53. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America*. 2000; 97(12):6640–5. <https://doi.org/10.1073/pnas.120163297> PMID: 10829079
54. Miller JH. *Experiments in Molecular Genetics*: Cold Spring Harbor; 1972.
55. Lee C, Kim J, Shin SG, Hwang S. Absolute and relative QPCR quantification of plasmid copy number in *Escherichia coli*. *Journal of biotechnology*. 2006; 123(3):273–80. <https://doi.org/10.1016/j.jbiotec.2005.11.014> PMID: 16388869
56. Benjamini Y, Hochberg Y. On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of Educational and Behavioral Statistics*. 2000; 25(1):60–83.
57. RStudio-Team. *RStudio: Integrated Development for R*, Inc., Boston, MA. 2015.