Published in final edited form as:

Curr Microbiol. 2011 January ; 62(1): 198–208. doi:10.1007/s00284-010-9693-3.

Correlations Between Bacterial Ecology and Mobile DNA

Irene L. G. Newton and

Department of Biological Sciences, Wellesley College, Wellesley, MA 02481, USA

Seth R. Bordenstein

Department of Biological Sciences, Vanderbilt University, VU Station B, Box 35-1634, Nashville, TN 37235, USA

Seth R. Bordenstein: s.bordenstein@vanderbilt.edu

Abstract

Several factors can affect the density of mobile DNA in bacterial genomes including rates of exposure to novel gene pools, recombination, and reductive evolution. These traits are difficult to measure across a broad range of bacterial species, but the ecological niches occupied by an organism provide some indication of the relative magnitude of these forces. Here, by analyzing 384 bacterial genomes assigned to three ecological categories (obligate intracellular, facultative intracellular, and extracellular), we address two, related questions: How does the density of mobile DNA vary across the Bacteria? And is there a statistically supported relationship between ecological niche and mobile element gene density? We report three findings. First, the fraction of mobile element genes in bacterial genomes ranges from 0 to 21% and decreases significantly: facultative intracellular > extracellular > obligate intracellular bacteria. Results further show that the obligate intracellular bacteria that host switch have a higher mobile DNA gene density than the obligate intracellular bacteria that are vertically transmitted. Second, while bacteria from the three ecological niches differ in their average mobile DNA contents, the ranges of mobile DNA found in each category overlap a surprising extent, suggesting bacteria with different lifestyles can tolerate similar amounts of mobile DNA. Third, mobile DNA gene densities increase with genome size across the entire dataset, and the significance of this correlation is dependent on the obligate intracellular bacteria. Further, mobile DNA gene densities do not correlate with evolutionary relationships in a 16S rDNA phylogeny. These findings statistically support a compelling link between mobile element evolution and bacterial ecology.

Introduction

Any theory of bacterial genome evolution must account for the pervasive acquisition and erosion of genes. Most genome expansion occurs via the acquisition of genes from the environment or from other bacteria through horizontal gene transfer, while genome contraction is recognized to occur through the processes of pseudogenization and/or gene deletions. Some of the primary agents of horizontal gene transfer and deletions in bacterial genomes are mobile genetic elements including bacteriophages, transposons, and plasmids [1]. These elements, by definition, have the capacity to mobilize and transfer DNA from one portion of the genome to another or between bacterial cells and species.

Several forces can affect the accumulation and deletion of mobile genetic elements from bacterial genomes. First, rates of exposure to novel gene pools will determine whether a bacterium has an expanded or reduced opportunity to contact and acquire mobile elements [2,3]. Second, many species of vertically transmitted bacteria generally experience a population bottleneck during transmission from mother to developing eggs. The size of the bottleneck affects population structure and can reduce effective population size relative to free-living bacteria [4]. Third, loss of recombination genes and repeats typify several genomes of obligate intracellular bacteria [3,5–8], and the reliance of some mobile genetic elements on these features could alter their ability to invade or copy themselves in a genome. Fourth, periods of relaxed selection such as a sudden expansion of resources may be coupled with an accumulation of mobile elements [9,10]. Quantitative analyses of mobile element densities across the Bacteria will therefore help illuminate the constraints and similarities of mobile element accumulation. The dramatic expansion in the number of bacterial genomes sequenced makes it especially timely for a quantitative assessment of new and existing predictions from previous studies [2,3,9,11,12]. Here we ask two questions about bacterial mobile elements.

Question 1: How does mobile DNA density vary across bacteria in different ecological roles?

So far, our understanding of bacterial mobile elements is based principally on case studies of specific mobile elements, or comparative genomic datasets that represent a fraction of the currently available genomes. While these data highlight the overall reduction of mobile DNA in asexual obligate intracellular bacteria and the speed with which mobile DNA can be transferred from one species to another, it would be helpful to formally assess, by statistical comparisons, the patterns shaping the distribution of mobile DNA across many bacteria. These data would provide an important baseline for studying the mechanisms and relevance of mobile DNA.

Question 2: What are the evolutionary and ecological forces that shape variation in mobile element distributions in Bacteria?

Below we explicitly test the prediction that variation in mobile element gene density (i.e., the fraction of mobile DNA genes over total gene number) broadly correlates with ecological range. This hypothesis is predicated on the assumptions that first, bacteria vary in their rates of exposure to new gene pools based on their capacity to replicate as facultatively intracellular, extracellular, and obligately intracellular organisms, and second, these ecological constraints should be a general indicator of effective population sizes and the relative magnitude of reductive evolution [13]. While some species will be exceptions to the expected trends, a general prediction is that facultative, intracellular bacteria will, on average, have the highest mobile element gene densities because they replicate both inside and outside host cells. This lifestyle exhibits the most promiscuity in niches to replicate in and generally will experience a greater exposure to novel gene pools. Facultatively intracellular bacteria that enter the intracellular environment may only represent a small fraction of the total population, but they could escape from host cells back into the free-living world and then seed the newly acquired mobile elements into the free-living population. As selfish genetic elements often encode their own gene drive system, even rare instances of mobile element exchange could lead to the rapid propagation of the mobile DNA throughout the bacterial population. Similarly, bacteria with only an extracellular replicative stage will have higher mobile DNA gene densities on average than species subject to obligate intracellularity (i.e., the requirement of a host cell for replication). The differences in mobile element content between facultative and extracellular bacteria, to our knowledge, have never been evaluated. Finally, the “intracellular arena hypothesis” posits that obligate intracellular bacteria that transmit horizontally will more readily exchange

genetic material [2,14]. Thus, we test whether obligate intracellular species that host switch have more mobile DNA than those that do not and discuss explanations for the observation. Here we explore these predictions and others by statistically comparing the mobile element densities of 384 bacterial genomes against their phylogenetic relationships, genome sizes, and ecology.

Results

Several hundred bacterial genomes are now in the J. Craig Venter Institute's Comprehensive Microbial Resource (JCVI CMR) [15]. The JCVI CMR provides a unified framework for annotation by demarcating genes into role categories (e.g., phage, plasmid, and transposon) using sequence alignments, predictive homology tools, and similarities to curated hidden markov models (HMMs). Integrons and integrative conjugative elements are also included and annotated as transposons and plasmids, respectively. We note appropriate caution that mobile element gene functions are based principally on sequence annotations and are not experimentally proven in many of the specific genomes. For the purposes of our analysis, the total number of genes in each role category is divided by the total number of genes in the genome for each organism, yielding a normalized fraction.

Using the entire dataset from JCVI CMR release 22.0, we assigned 270 taxa as extracellular, 74 taxa as facultative intracellular, and 40 taxa as obligate intracellular (Fig. 1). The average genome sizes of bacteria within each lifestyle grouping were found to differ predictably (Kruskal–Wallis, $\chi^2 = 91.31$, $df = 2$, $P < 0.001$). The obligate intracellular bacteria (genome size $\pm$ SD = 1134 ± 520 kb) had expectedly smaller genomes than either facultatively intracellular (4022 ± 1897 kb) or extracellular bacteria (3440 ± 1486 kb) (Mann–Whitney U test, $P < 0.001$ for both). Further, facultatively intracellular species have significantly larger genomes than the extracellular species (MWU, $P = 0.021$). The AT content for these groups of bacteria was also different: the obligately intracellular bacteria (%G + C $\pm$ SD = 34.67 ± 8.18) had lower GC content than either facultatively intracellular (51.08 ± 13.54) or extracellular bacteria (47.95 ± 12.84) (MWU, $P < 0.001$ for both).

Is Mobile DNA Gene Density Independent of Phylogenetic Relationships?

Lifestyle, mobile element gene density, and bacterial taxonomy were color coded onto the 16S rDNA phylogeny of all the species used in this study, illustrating a broad phylogenetic sampling (Fig. 2). First, we tested the null hypothesis that phylogeny is an independent variable to mobile element gene densities. We used the descriptive statistic, K , in a computer program written by Liam Revell, 2006 [16,17]. The phylogeny-based statistical method fits continuous, phenotypic characters to a tree topology and compares the fit to a model in which the characters are randomly assigned throughout the tree. If the output statistic K is greater than one, then close relatives are more similar in the phenotypic character than expected by chance; if K is less than one, then relatives do not resemble each other. In our dataset, bacteria related by 16S rRNA gene phylogenies did not harbor a similar fraction of mobile DNA genes ($K = 0.0024$). Therefore, phylogenetic signal is not influencing our comparisons of mobile element densities. This conclusion is intuitive as mobile elements are well known to comprise a significant portion of the strain level differences found within species [18–21] and can transfer unrestrained across divergent taxa [22–24]. Regardless, we took extra cautionary measures to account for pseudoreplication bias in our analyses. We repeated each statistical analysis on ten randomly generated datasets comprising only one genomic representative per species each ($N = 268$ genomes). The single-species cumulative results are presented as the fraction ($X/10$) of those replicates that were consistent with the statistical inference based on the complete dataset, with a P -value cutoff of 0.05. In most cases, the results mirrored those of the complete dataset, as illustrated further below.

How Does Mobile DNA Gene Density Change with Genome Size?

Genome sizes vary more than an order of magnitude (545–9358 total genes), and this variation correlates with the range of the ecological niche: Obligate intracellular bacteria (mean $\pm$ SD: 1132 ± 519 genes) have smaller genomes than extracellular bacteria (3464 ± 1481 genes) (MWU, two-tailed test, $P < 0.0001$), which have smaller genomes than facultative, intracellular bacteria (4022 ± 1896 genes) (MWU, two-tailed test, $P = 0.014$). Based on a non-parametric regression analysis, the density of mobile DNA genes per genome significantly increases with gene number ($P < 0.001$, $\rho = 0.22$; reduced datasets 10/10), and the significance of this correlation is anchored by the strong correlation within the obligate intracellular bacteria ($P < 0.001$, $\rho = 0.51$; reduced datasets 10/10). Conversely, there is no correlation for total mobile DNA and gene number within the facultative intracellular ($P = 0.09$; reduced datasets 10/10), the extracellular bacteria alone ($P = 0.36$; reduced datasets 10/10), or when facultatives and extracellulars are pooled together ($P = 0.08$). The positive correlation between mobile DNA and genome size when all three categories are analyzed is also highly significant for bacteriophages ($P = 0.003$, $\rho = 0.15$), transposons ($P < 0.001$, $\rho = 0.3$), and plasmids ($P < 0.001$, $\rho = 0.21$), respectively.

As a negative control to demonstrate that different, functional role categories elicit different correlations with genome size, we analyzed two other JCVI CMR role categories: cellular processes and DNA metabolism. Genes in these categories are generally subject to purifying selection and because they tend to be functionally conserved, we expect them to be proportionally overrepresented in small, reduced genomes. As expected, findings reveal a strong, inverse correlation with genome size ($P < 0.0001$, $\rho = -0.46$ and -0.76).

Do Constraints in Ecological Lifestyle Vary with Mobile DNA Gene Density?

Variation in the range of mobile DNA content may reflect an intrinsic capacity for each bacterial genome to tolerate a certain density of mobile elements. For instance, there is little mobile DNA in the genomes of obligate symbionts of insects that are vertically transmitted [25–29]; yet, recent genome sequences suggest there can be high fractions of mobile DNA and lateral transfer in some intracellular symbiont lineages [30–33].

The total variation in genes dedicated to mobile DNA functions across the Bacteria ranges between 0 and 21%. Although genome size is nearly four-fold smaller in obligate species, the range in the mobile element gene density overlaps between all three ecological lifestyles (extracellular, facultative intracellular, and obligatory intracellular). As illustrated by a frequency histogram, the facultative, intracellular bacteria have a similarly broad range of mobile elements, ranging between 0 and 14% of the genes in the genome (Fig. 3a). The bacteria forming obligate intracellular relationships with eukaryotes also exhibited a broad range of mobile element genes, 0–12%. Interestingly, extracellular, environmental taxa had the narrowest range of mobile DNA content. With the exception of *Xanthomonas oryzae* KACC10331, which dedicates 21% of its genome to mobile DNA genes, these species had less than 8% mobile DNA genes in their genomes (Fig. 3a). The ranges of specific mobile element types reveal similar patterns (Fig. 3b–d). In sum, the three bacterial lifestyles have the genetic capacity to tolerate high fractions of mobile DNA. However, as we demonstrate below, there are general trends not reflected in the ranges of the three ecological categories that lead to significant differences in the distribution of mobile DNA contents.

There is a strong effect of ecological lifestyle on total mobile DNA gene content (Kruskal–Wallis, $\chi^2 = 28.2$, $df = 2$, $P < 0.001$; reduced datasets 9/10). Mean mobile DNA gene densities decrease from a broad to narrow ecological range: facultative > extracellular > obligate (Fig. 4). The facultative intracellular organisms had significantly higher mobile DNA gene densities in their genomes compared to both extracellular (MWU, $P = 0.003$;

reduced dataset 9/10) and obligate intracellular organisms (MWU, $P < 0.001$; reduced dataset 10/10). The obligate intracellular bacteria in turn had lower mobile element gene densities than that of the extracellular bacteria (MWU, $P < 0.001$; reduced dataset 10/10). These results statistically confirm with the largest dataset to date that genomic compositions of mobile DNA vary with constraints in bacterial ecology.

Does Mobile DNA Content Vary with Bacterial Transmission Mode?

Obligately intracellular bacteria that are strictly, vertically transmitted harbor few to no mobile elements in their genomes due to host restriction and reductive evolution [2,3,9], but the comparison to obligate intracellular species that are horizontally transmitted has not been scrutinized. Despite their highly reduced genomes, obligate intracellular bacteria that host switch could have a greater propensity for experiencing diverse gene pools because they can move between different hosts, often have the genetic repertoire to facilitate recombination of mobile elements, and tend not to be sequestered in bacteriocytes.

Therefore, we have posited that eukaryotic cells may act as arenas for mobile DNA exchange between co-infections of horizontally transferred intracellular bacteria [2,14]. This “intracellular arena hypothesis” predicts that obligate intracellular bacteria that host switch, and come into contact with other bacteria during infection, will have more mobile DNA genes, on average, than those that are strictly sequestered in one host lineage. For instance, *Wolbachia* wBm is a vertically transmitted mutualist of filarial nematodes and has only 3% of its genes dedicated to mobile DNA [34]. Conversely, the related *Wolbachia* wMel and wPip in arthropods are horizontally transmitted and have 11% and 21% of their genes dedicated to mobile DNA, respectively [33,35]. Some vertically transmitted obligate intracellular bacteria can co-infect host cells with host-switching species. Thus, an alternative explanation for why the former taxa have little to no mobile DNA is that they often have extremely reduced genomes that may be beyond the stable uptake of mobile elements. For instance, they could either lack the genetic machinery to incorporate new DNA or die once a mobile element invades the cell or genome.

To examine if variation in host range among the obligate intracellular bacteria globally associates with mobile DNA gene density, we tested whether the mobile DNA density of the horizontally transmitted species is greater than that of the vertically transmitted species. We report that while 5/8 genomes from the vertically transmitted organisms completely lack mobile DNA, only 1/32 (i.e., *Candidatus Ruthia magnifica*) genomes from the horizontally transmitted species lacked mobile DNA (Fisher’s exact test, $P = 0.0005$). Further, host-switching species have twice the average amount of mobile DNA genes than vertically transmitted ones (Fig. 5a, MWU, $P = 0.027$; reduced dataset 5/10). The variation in the reduced datasets appears to be influenced by the inclusion or exclusion of *Wolbachia* wMel, which has a high mobile DNA content that is typical of other, recent *Wolbachia* genomes not included in this dataset [30,36]. The horizontally transmitted species also have larger genome sizes (Fig. 5b, MWU, $P < 0.001$) and higher GC contents (Fig. 5b, MWU, $P = 0.005$). Two significant outliers in this analysis are worth discussing. They include *Wigglesworthia glossinidia brevipalpis*, an obligately intracellular, vertically transmitted symbiont with a large phage load (>3%) and *Ruthia magnifica*, an obligately intracellular, horizontally transmitted symbiont with no encoded mobile elements. *Wigglesworthia* are estimated to have begun their association with its tsetse fly host relatively recently [37], perhaps suggesting that population genetic processes have not had time to decrease their phage content. For the symbiont *Ruthia*, there is some debate as to the extent this bacteria is horizontally transmitted. It was for many years considered a strictly vertically transmitted symbiont but only recently has been suggested to experience horizontal transmission on rare occasions [38]. Therefore, although *Ruthia* may experience horizontal transmission, the vast

majority of the time it is vertically transmitted, limiting the persistence of mobile elements in this lineage.

Horizontal transmission can allow obligatory intracellular taxa to escape reductive evolution by creating opportunities for genetic exchange with other bacteria. Many intracellular bacteria co-infect eukaryotic hosts, including *Wolbachia* co-infection with other *Wolbachia* [39–41] and insect symbionts [42–45], *Rickettsia* co-infections with other *Rickettsia* [46] and *Ehrlichia* [47], and *Phytoplasma* co-infections with each other [48].

Mobile Element Types

Transposons account for 66% of the total number of mobile DNA genes across all genomes, vastly more than phage genes (29%) or plasmid genes (5%). In fact, transposons make up the major mobile element in 190/384 genomes. In order to determine whether lifestyle affects the distribution of individual types of mobile elements, we analyzed the data with regard to three specific mobile element types: bacteriophages, transposons, and plasmids.

There is no significant difference in bacteriophage DNA content between the genomes of extracellular and facultative intracellular bacteria (Fig. 4, MWU, $P = 0.469$). Obligatory intracellular bacterial genomes have a significantly smaller proportion of phage genes than extracellular bacterial genomes (Fig. 4, MWU, $P < 0.001$) but this finding is not replicated in the single-species datasets; only two out of ten datasets recapitulate the results from the full dataset. Examples exist of obligatory intracellular bacteria that harbor substantial phage gene pools [30,32,33,49]. These phages can provide a source of novel genetic material for the bacterial symbiont and eukaryotic hosts.

Transposons are the most abundant type of mobile genetic element in 49% of the bacterial genomes. These pieces of DNA can excise or copy themselves from the resident chromosome to another site in the same chromosome or to an extrachromosomal element such as a plasmid or phage. Although these mobile elements might bring novel sequences into a bacterial lineage, their effects are presumed to be largely deleterious [50]. The variation in transposon content among the three lifestyles was found to be statistically significant (Kruskal–Wallis test, $\chi^2 = 39.75$, $df = 2$, $P < 0.001$; reduced datasets 10/10). Facultative intracellular bacteria have a greater proportion of transposable elements in their genomes than extracellular bacteria (Fig. 4, MWU, $P = 0.001$; reduced datasets 6/10), which in turn had a significantly greater fraction of transposons than the obligatory intracellular bacteria (Fig. 4, MWU, $P < 0.001$; reduced datasets 10/10).

Finally, DNA can be mobilized between bacterial cells via a transferable plasmid. The number of extrachromosomal plasmids harbored by bacteria varies according to lifestyle, with obligate intracellular bacteria having fewer plasmids (mean $\pm$ SD = 0.36 ± 0.72) than either facultative intracellular (0.80 ± 1.31) or extracellular bacteria (0.77 ± 1.73) (Fig. 4). Our analysis based on whole bacterial genomes also revealed that plasmid-related genes in bacterial genomes are not common and account for 5% of the mobile element genes. A Kruskal–Wallis test revealed significant variation among the three lifestyles ($\chi^2 = 8.417$, $df = 2$, $P < 0.015$; reduced datasets 6/10). While facultative intracellular and extracellular bacteria have similar plasmid-related gene contents, the obligate intracellular bacteria show reduced content (MWU, $P < 0.001$; reduced datasets 8/10).

In sum, the analysis of specific mobile elements densities shows: (i) transposons account for the majority of mobile DNA genes in nearly half of all the bacterial genomes and significantly vary between the three bacterial lifestyles, (ii) obligate intracellular bacteria have significantly lower mobile element gene densities compared to the other species, and (iii) the densities of those elements that frequently encode their own inter-cellular mobility,

phages, and plasmids, do not significantly differ between facultative and extracellular bacteria.

Concluding Remarks

Our results show that variation in the fraction of mobile DNA genes per genome correlates with variation in bacterial ecology. Previous studies have observed a compelling link between ecology and specific mobile element types (e.g., repeats [3] and insertion sequences [9,12]), but these studies were comprised of a fraction of the genomes now publicly available [2]. In this study, we observed statistically supported differences in total mobile DNA gene densities, as well in phage, transposon, and plasmid gene densities among the three ecological categories from 384 genomes. We also observed differences in mobile element contents between facultative intracellular and extracellular bacteria, and between horizontally transmitted and vertically transmitted obligate intracellular bacteria.

Several ecological and evolutionary processes that are not mutually exclusive could affect the fraction of mobile DNA in bacterial genomes. First, the frequency with which bacteria experience novel gene pools will determine whether the bacteria have an expanded or reduced opportunity to acquire new mobile elements [2]. Second, once the contact with mobile element gene pools is made, there could be genetic or physiological constraints for uptake and accommodating the presence of a new element in the cell's cytoplasm or genome. For example, loss of recombination genes and repeats typify some genomes of obligate intracellular bacteria [3,5–8], and some mobile elements rely on these genetic factors to invade or copy themselves in a bacterial genome. Third, effective population sizes vary significantly between the obligate intracellular bacteria and non-obligate intracellular bacteria, which affects the frequency of population bottlenecks and efficacy of selection. For instance, if mobile DNAs are harmful, clonal species with low effective population sizes that avoid the proliferation of these elements are likely to also avoid extinction. Conversely, species with a larger effective population size, increased exposure to novel gene pools, and proclivity for recombination are expected to harbor larger numbers of mobile genetic elements relative to clonal species. In support of these hypotheses, estimates of effective population size derived from genome-wide K_a/K_s ratios for a subset of bacteria [13] were compared across the three ecological categories investigated here. An increased K_a/K_s ratio across the genome can result from an increased level of slightly deleterious amino acid replacements due to genetic drift. The obligately intracellular bacteria were found to have significantly larger K_a/K_s ratios than either the facultatively intracellular ($P = 0.013$, MWU) or extracellular bacteria ($P = 0.003$, MWU). However, the facultatively intracellular and the extracellular bacteria did not have significantly different K_a/K_s ratios ($P = 0.541$, MWU). Also, K_a/K_s ratios were found to exhibit a significant negative correlation with total mobile element load ($P = 0.033$; $\rho = -0.251$); species with smaller estimated population sizes were found to harbor fewer mobile elements, as predicted above.

The fraction of mobile DNA genes ranges in this analysis from 0 to 21% across the Bacterial domain, and this variation is independent of phylogeny, as would be expected for elements that evolve separately from the core genome. Mobile element gene distribution decreases from facultative > extracellular > obligate intracellular bacteria and is most polarized in the obligate intracellular taxa. Among the obligate, intracellular species, there are further differences. Few of the strictly vertically transmitted bacterial species contain any mobile element genes (3/8), while nearly all horizontally transmitted ones did harbor mobile DNA (31/32).

These findings demonstrate that even species confined to obligate intracellularity can experience variation in mobile gene content that can be correlated to variation in ecological

traits. The obligate lifestyle is associated with genome reduction, a narrow ecological niche, dependence on host nutrients, and the pervasive effects of Muller's ratchet. However, changes in ecology such as increased rates of horizontal transmission and recombination will counter the effects of a narrow niche and increase exposure to other bacteria that may promote mobile element acquisition. In addition, specialized bacteria that assist host fitness may have evolved beyond the stable uptake of mobile elements. For instance, *Wolbachia* endosymbionts depend on host cells for replication, and all major lineages have small genomes characteristic of obligate intracellular species. However, strain *wBm* is a vertically transmitted mutualist of filarial nematodes and has only 3% of its genes dedicated to mobile DNA [34]. Conversely, the related *Wolbachia* *wMel* and *wPip* in arthropods are horizontally transmitted and have 11% and 21% of their genes dedicated to mobile DNA, respectively [33,35]. In addition, the horizontally transmitted *Wolbachia* harbor active mobile elements and show evidence of rampant chromosomal recombination and lateral gene transfer [14,30,51,52]. It is noteworthy that *Wolbachia* are part of the 500 My old *Rickettsiales* clade of obligatory intracellular species and yet many of the specific mobile element genes are unique to this genus. Thus, these differences in mobile DNA gene densities exemplify how changes in ecological range in related, intracellular bacteria can predict genomic content differences.

Our analysis of the specific element types also revealed that while obligately intracellular bacteria had significantly fewer phage, plasmid, and transposon genes than the other ecological categories, the facultatively intracellular and extracellular bacteria only varied in their transposon content. One explanation for this difference is that the endogenous capacity of many phages and plasmids to encode their own recombination allows them to freely move between bacteria that have an extracellular, replicative stage. Alternatively, facultative intracellular bacteria that are subject to shifting environmental boundaries may require beneficial, mobile DNAs that can inactivate certain functional genes or expression phenotypes in different environments [53–55]. Further, deleterious insertion sequences may preferentially accumulate in the genomes of symbiotic bacteria at the onset of a host association [9] because the infection experiences a period of decreased selection owing to the new expanded set of resources of the host cell. All these reasons support the findings that facultative intracellular bacteria will have the highest mobile element densities, followed by extracellular and then the obligate intracellular bacteria.

Our study shows that comparisons of mobile element abundance between ecologically different groups can be used as a proxy to test the links between effective population size and mobile elements. In summary, the statistical patterns corroborate the association of mobile element density with bacterial ecology across a large dataset. Particularly through the use of genome sequencing, we have seen substantial progress in understanding how bacterial genomes evolve. Since Barbara McClintock's ideas were presented 70 years ago, mobile elements have been recognized as central to the evolution of nearly all genomes. The comparative genomics data presented here has provided a quantitative framework for generalizations of the patterns or rules that shape mobile genetic elements in bacteria.

Materials and Methods

Genomic Data Acquisition and Handling

The role category content for all genomes was downloaded from the JCVI CMR release 22.0 [15]. For the genomes assigned to different ecological categories, primary and automatic genome annotation were not significantly different from each other (i.e., the obligately intracellular category for both annotations: $P = 0.969$, MWU). All analyses are therefore based on primary annotation. Each role category bin of interest was summed, and these data were tabulated using perl scripts (written by I.L.G. Newton). The raw number of genes in

each category was normalized by total gene number to give a percentage of mobile element genes (Supplementary Table 1). Statistical analyses (Kruskal–Wallis and Mann–Whitney *U* tests) were performed in SPSS.

Classification of Lifestyle and Transmission Strategy

Bacteria in this analysis included organisms from a broad phylogenetic swath—from Firmicutes to Gamma-proteo-bacteria (Fig. 1). These organisms were classified as extracellular, facultative intracellular, and obligatory intracellular based on searches in the primary literature, *The Prokaryotes* [56], and the web-based resource Island-Path [57] (Supplementary Table 1). Extracellular bacteria are those that have no documented intracellular component to their life cycle. Facultative intracellular bacteria are those that can replicate within host cells and outside host cells in an extracellular state. Obligate intracellular bacteria are those that have not yet been cultured outside of their host cells and depend on intracellular growth for their replication (e.g., *Buchnera*, *Baumannia*, and *Chlamydia*). We note a caveat to these ecological designations in that they depend on field-specific methodological advances. For example, intracellular bacteria for whom appropriate media has been developed will be classified as facultatively intracellular (e.g., *Coxiella burnetii* has recently been cultured in cell-free extract) while those for whom appropriate growth conditions have not been devised will remain classified as obligately intracellular.

Organisms with host-associated lifestyles (facultative or obligate) were also characterized with regard to their transmission strategy, based either on phylogenetic or experimental data. Traditionally, bacteria are thought of as occupying two distinct transmission strategies: vertical (bacteria are passed directly from mother to offspring) or horizontal (bacteria are passed from one individual or species to another, often via the environment). It must be emphasized, however, that these mechanisms are not distinct categories but instead represent endpoints in a range of transmission strategies. The parasite *Wolbachia*, for example, shows a mixed pattern of transmission among arthropods, with primarily vertical transmission within and horizontal transmission between host species. For the purposes of this analysis, we characterize any bacteria with at least some host switching as horizontally transmitted. As a result, bacteria were sorted into those species that are horizontally transmitted (encompassing both host switching and environmental transmission) versus those that are strictly, vertically transmitted to the next generation.

Phylogeny

16S rRNA gene sequences for each organism in this analysis were downloaded from the ARB-SILVA rRNA database [58]. In this database, the 16S rRNA genes are pre-aligned based on secondary structure. The program RAxML [59] was used to infer a maximum likelihood phylogeny (GTR + Γ) for the full and reduced dataset taxa. The resulting trees combined with the descriptive and numeric characteristics described above were used as input to an implementation of the *K* statistic (written by Liam Revell [17]).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Patrick Abbot, Robert Brucker, and Antonis Rokas for their comments and suggestions to improve the manuscript. We also thank Liam Revell for access to his implementation of the *K* statistic ahead of publication. This study was supported by grants NSF IOS-0852344 and NIH R01 GM085163-01 to SRB and an NSF Postdoctoral Fellowship to ILG Newton.

References

1. Frost LS, Leplae R, Summers AO, Toussaint A. Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol* 2005;3:722–732. [PubMed: 16138100]
2. Bordenstein SR, Reznikoff WS. Mobile DNA in obligate intracellular bacteria. *Nat Rev Microbiol* 2005;3:688–699. [PubMed: 16138097]
3. Frank AC, Amiri H, Andersson SG. Genome deterioration: loss of repeated sequences and accumulation of junk DNA. *Genetica* 2002;115:1–12. [PubMed: 12188042]
4. Mira A, Moran NA. Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. *Microb Ecol* 2002;44:137–143. [PubMed: 12087426]
5. Andersson SG, Alsmark C, Canback B, Davids W, Frank C, Karlberg O, Klasson L, Antoine-Legault B, Mira A, Tamas I. Comparative genomics of microbial pathogens and symbionts. *Bioinformatics* 2002;18 Suppl 2:S17. [PubMed: 12385978]
6. Dale C, Wang B, Moran N, Ochman H. Loss of DNA recombinational repair enzymes in the initial stages of genome degeneration. *Mol Biol Evol* 2003;20:1188–1194. [PubMed: 12777518]
7. Moran NA, Wernegreen JJ. Lifestyle evolution in symbiotic bacteria: insights from genomics. *Trends Ecol Evol* 2000;15:321–326. [PubMed: 10884696]
8. Silva FJ, Latorre A, Moya A. Why are the genomes of endosymbiotic bacteria so stable? *Trends Genet* 2003;19:176–180. [PubMed: 12683967]
9. Moran NA, Plague GR. Genomic changes following host restriction in bacteria. *Curr Opin Genet Dev* 2004;14:627–633. [PubMed: 15531157]
10. Plague GR, Dunbar HE, Tran PL, Moran NA. Extensive proliferation of transposable elements in heritable bacterial symbionts. *J Bacteriol* 2008;190:777–779. [PubMed: 17981967]
11. Casjens S. Prophages and bacterial genomics: what have we learned so far? *Mol Microbiol* 2003;49:277–300. [PubMed: 12886937]
12. Touchon M, Rocha EP. Causes of insertion sequences abundance in prokaryotic genomes. *Mol Biol Evol* 2007;24:969–981. [PubMed: 17251179]
13. Kuo CH, Moran NA, Ochman H. The consequences of genetic drift for bacterial genome complexity. *Genome Res* 2009;19:1450–1454. [PubMed: 19502381]
14. Bordenstein SR, Wernegreen JJ. Bacteriophage flux in endosymbionts (*Wolbachia*): infection frequency, lateral transfer, and recombination rates. *Mol Biol Evol* 2004;21:1981–1991. [PubMed: 15254259]
15. Peterson JD, Umayam LA, Dickinson T, Hickey EK, White O. The comprehensive microbial resource. *Nucleic Acids Res* 2001;29:123–125. [PubMed: 11125067]
16. Blomberg SP, Garland T Jr, Ives AR. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 2003;57:717–745. [PubMed: 12778543]
17. Revell LJ, Harrison AS. PCCA: a program for phylogenetic canonical correlation analysis. *Bioinformatics* 2008;24:1018–1020. [PubMed: 18292115]
18. Banks DJ, Beres SB, Musser JM. The fundamental contribution of phages to GAS evolution, genome diversification and strain emergence. *Trends Microbiol* 2002;10:515–521. [PubMed: 12419616]
19. Ogura Y, Kurokawa K, Ooka T, Tashiro K, Tobe T, Ohnishi M, Nakayama K, Morimoto T, Terajima J, Watanabe H, et al. Complexity of the genomic diversity in enterohemorrhagic *Escherichia coli* O157 revealed by the combinational use of the O157 Sakai OligoDNA microarray and the whole genome PCR scanning. *DNA Res* 2006;13:3–14. [PubMed: 16766508]
20. Ohnishi M, Kurokawa K, Hayashi T. Diversification of *Escherichia coli* genomes: are bacteriophages the major contributors? *Trends Microbiol* 2001;9:481–485. [PubMed: 11597449]
21. Van Sluys MA, de Oliveira MC, Monteiro-Vitorello CB, Miyaki CY, Furlan LR, Camargo LE, da Silva AC, Moon DH, Takita MA, Lemos EG, et al. Comparative analyses of the complete genome sequences of Pierce's disease and citrus variegated chlorosis strains of *Xylella fastidiosa*. *J Bacteriol* 2003;185:1018–1026. [PubMed: 12533478]
22. Brussow H, Canchaya C, Hardt WD. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev* 2004;68:560–602. [PubMed: 15353570]

23. Hendrix RW, Hatfull GF, Smith MC. Bacteriophages with tails: chasing their origins and evolution. *Res Microbiol* 2003;154:253–257. [PubMed: 12798229]
24. Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. *Nature* 2000;405:299–304. [PubMed: 10830951]
25. Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T, Hattori M, Aksoy S. Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. *Nat Genet* 2002;32:402–407. [PubMed: 12219091]
26. Degnan PH, Lazarus AB, Wernegreen JJ. Genome sequence of *Blochmannia pennsylvanicus* indicates parallel evolutionary trends among bacterial mutualists of insects. *Genome Res* 2005;15:1023–1033. [PubMed: 16077009]
27. Gil R, Sabater-Munoz B, Latorre A, Silva FJ, Moya A. Extreme genome reduction in *Buchnera* spp.: toward the minimal genome needed for symbiotic life. *Proc Natl Acad Sci USA* 2002;99:4454–4458. [PubMed: 11904373]
28. Gil R, Silva FJ, Zientz E, Delmotte F, Gonzalez-Candelas F, Latorre A, Rausell C, Kamerbeek J, Gadau J, Holldobler B, et al. The genome sequence of *Blochmannia floridanus*: comparative analysis of reduced genomes. *Proc Natl Acad Sci USA* 2003;100:9388–9393. [PubMed: 12886019]
29. Tamas I, Klasson L, Canback B, Naslund AK, Eriksson AS, Wernegreen JJ, Sandstrom JP, Moran NA, Andersson SG. 50 million years of genomic stasis in endosymbiotic bacteria. *Science* 2002;296:2376–2379. [PubMed: 12089438]
30. Klasson L, Westberg J, Sapountzis P, Naslund K, Lutnaes Y, Darby AC, Veneti Z, Chen L, Braig HR, Garrett R, et al. The mosaic genome structure of the *Wolbachia* wRi strain infecting *Drosophila simulans*. *Proc Natl Acad Sci USA* 2009;106:5725–5730. [PubMed: 19307581]
31. Ogata H, Renesto P, Audic S, Robert C, Blanc G, Fournier PE, Parinello H, Claverie JM, Raoult D. The genome sequence of *Rickettsia felis* identifies the first putative conjugative plasmid in an obligate intracellular parasite. *PLoS Biol* 2005;3:e248. [PubMed: 15984913]
32. Wei W, Davis RE, Jomantiene R, Zhao Y. Ancient, recurrent phage attacks and recombination shaped dynamic sequence-variable mosaics at the root of phytoplasma genome evolution. *Proc Natl Acad Sci USA* 2008;105:11827–11832. [PubMed: 18701718]
33. Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC, McGraw EA, Martin W, Esser C, Ahmadinejad N, et al. Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol* 2004;2:E69. [PubMed: 15024419]
34. Foster J, Ganatra M, Kamal I, Ware J, Makarova K, Ivanova N, Bhattacharyya A, Kapatral V, Kumar S, Posfai J, et al. The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. *PLoS Biol* 2005;3:e121. [PubMed: 15780005]
35. Klasson L, Walker T, Sebaihia M, Sanders MJ, Quail MA, Lord A, Sanders S, Earl J, O'Neill SL, Thomson N, et al. Genome evolution of *Wolbachia* strain wPip from the *Culex pipiens* group. *Mol Biol Evol* 2008;25:1877–1887. [PubMed: 18550617]
36. Salzberg SL, Puiu D, Sommer DD, Nene V, Lee NH. Genome sequence of the *Wolbachia* endosymbiont of *Culex quinquefasciatus* JHB. *J Bacteriol* 2009;191:1725. [PubMed: 19114486]
37. Chen XA, Li S, Aksoy S. Concordant evolution of a symbiont with its host insect species: molecular phylogeny of genus *Glossina* and its bacteriome-associated endosymbiont, *Wigglesworthia glossinidia*. *J Mol Evol* 1999;48:49–58. [PubMed: 9873076]
38. Stewart FJ, Young CR, Cavanaugh CM. Evidence for homologous recombination in intracellular chemosynthetic clam symbionts. *Mol Biol Evol* 2009;26:1391–1404. [PubMed: 19289597]
39. Werren JH, Windsor DM. *Wolbachia* infection frequencies in insects: evidence of a global equilibrium? *Proc Biol Sci* 2000;267:1277–1285. [PubMed: 10972121]
40. Kikuchi Y, Fukatsu T. Diversity of *Wolbachia* endosymbionts in heteropteran bugs. *Appl Environ Microbiol* 2003;69:6082–6090. [PubMed: 14532065]
41. Kondo N, Shimada M, Fukatsu T. Infection density of *Wolbachia* endosymbiont affected by co-infection and host genotype. *Biol Lett* 2005;1:488–491. [PubMed: 17148240]

42. Heddi A, Grenier AM, Khatchadourian C, Charles H, Nardon P. Four intracellular genomes direct weevil biology: nuclear, mitochondrial, principal endosymbiont, and *Wolbachia*. *Proc Natl Acad Sci USA* 1999;96:6814–6819. [PubMed: 10359795]
43. Gomez-Valero L, Soriano-Navarro M, Perez-Brocal V, Heddi A, Moya A, Garcia-Verdugo JM, Latorre A. Coexistence of *Wolbachia* with *Buchnera aphidicola* and a secondary symbiont in the aphid *Cinara cedri*. *J Bacteriol* 2004;186:6626–6633. [PubMed: 15375144]
44. Weeks AR, Velten R, Stouthamer R. Incidence of a new sex-ratio-distorting endosymbiotic bacterium among arthropods. *Proc Biol Sci* 2003;270:1857–1865. [PubMed: 12964989]
45. Zchori-Fein E, Perlman SJ. Distribution of the bacterial symbiont *Cardinium* in arthropods. *Mol Ecol* 2004;13:2009–2016. [PubMed: 15189221]
46. Weinert LA, Werren JH, Aebi A, Stone GN, Jiggins FM. Evolution and diversity of *Rickettsia* bacteria. *BMC Biol* 2009;7:6. [PubMed: 19187530]
47. Mixson TR, Campbell SR, Gill JS, Ginsberg HS, Reichard MV, Schulze TL, Dasch GA. Prevalence of *Ehrlichia*, *Borrelia*, and *Rickettsial* agents in *Amblyomma americanum* (Acari: Ixodidae) collected from nine states. *J Med Entomol* 2006;43:1261–1268. [PubMed: 17162962]
48. Leyva-Lopez NE, Ochoa-Sanchez JC, Leal-Klevezas DS, Martinez-Soriano JP. Multiple phytoplasmas associated with potato diseases in Mexico. *Can J Microbiol* 2002;48:1062–1068. [PubMed: 12619818]
49. Everson JS, Garner SA, Fane B, Liu BL, Lambden PR, Clarke IN. Biological properties and cell tropism of Chp2, a bacteriophage of the obligate intracellular bacterium *Chlamydophila abortus*. *J Bacteriol* 2002;184:2748–2754. [PubMed: 11976304]
50. Wagner A. Periodic extinctions of transposable elements in bacterial lineages: evidence from intragenomic variation in multiple genomes. *Mol Biol Evol* 2006;23:723–733. [PubMed: 16373392]
51. Cordaux R, Pichon S, Ling A, Perez P, Delaunay C, Vavre F, Bouchon D, Greve P. Intense transpositional activity of insertion sequences in an ancient obligate endosymbiont. *Mol Biol Evol* 2008;25:1889–1896. [PubMed: 18562339]
52. Baldo L, Bordenstein S, Wernegreen JJ, Werren JH. Widespread recombination throughout *Wolbachia* genomes. *Mol Biol Evol* 2006;23:437–449. [PubMed: 16267140]
53. Edwards RJ, Brookfield JF. Transiently beneficial insertions could maintain mobile DNA sequences in variable environments. *Mol Biol Evol* 2003;20:30–37. [PubMed: 12519903]
54. Schneider D, Lenski RE. Dynamics of insertion sequence elements during experimental evolution of bacteria. *Res Microbiol* 2004;155:319–327. [PubMed: 15207863]
55. Zinser ER, Kolter R. Prolonged stationary-phase incubation selects for *lrp* mutations in *Escherichia coli* K-12. *J Bacteriol* 2000;182:4361–4365. [PubMed: 10894750]
56. Dworkin, M., editor. *The prokaryotes*. New York: Springer; 2007.
57. Hsiao W, Wan I, Jones SJ, Brinkman FS. IslandPath: aiding detection of genomic islands in prokaryotes. *Bioinformatics* 2003;19:418–420. [PubMed: 12584130]
58. Yarza P, Richter M, Peplies J, Euzéby J, Amann R, Schleifer KH, Ludwig W, Glockner FO, Rossello-Mora R. The All-Species Living Tree project: a 16S rRNA-based phylogenetic tree of all sequenced type strains. *Syst Appl Microbiol* 2008;31:241–250. [PubMed: 18692976]
59. Stamatakis A, Ludwig T, Meier H. RAXML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* 2005;21:456–463. [PubMed: 15608047]

Fig. 1.
Schematic flowchart of the three distinct ecological categories in this analysis, along with relevant genomic statistics and example genera

Fig. 2. Circular representation of an unrooted 16S rRNA gene phylogeny. Alignments of 16S rRNA genes were downloaded from arb-silva (<http://www.arb-silva.de/>), and a curated set was aligned according to the secondary structure of the functional RNA. These alignments, of at least 1,555 nucleotides in length, were used as input to RAxML V 7.0.4, a phylogeny-building program that can handle relatively large datasets [59]. The phylogeny was generated under a maximum likelihood framework implementing a GTR model with a gamma distribution of rate variation among sites

Fig. 3. Frequency histogram of mobile DNA content. The percentage of mobile DNA genes per genome is binned in intervals of 2% against the frequency of bacterial genomes having that percentage of mobile DNA genes. Bacterial lifestyles are plotted separately

Fig. 4. Mobile DNA contents and ecological range. The average normalized percentage of mobile gene content $\pm$ standard error for genomes across all three lifestyles (facultative intracellular, extracellular, or obligatory intracellular). Total content is first displayed, followed by transposons, bacteriophages, and plasmids

Fig. 5. Mobile DNA gene content, G + C content, and genome size correlates with transmission strategy. **a** Mean mobile DNA content $\pm$ standard error is shown as a percentage of the total number of genes in a genome for obligatory intracellular bacteria with different transmission strategies. **b** Mean G + C content $\pm$ standard error (*left*) and mean gene number per genome $\pm$ standard error (*right*) for horizontally and vertically transmitted obligatory intracellular bacteria