Genomic changes following host restriction in bacteria

Nancy A Moran* and Gordon R Plague

Many genomic sequences have been recently published for bacteria that can replicate only within eukaryotic hosts. Comparisons of genomic features with those of closely related bacteria retaining free-living stages indicate that rapid evolutionary change often occurs immediately after host restriction. Typical changes include a large increase in the frequency of mobile elements in the genome, chromosomal rearrangements mediated by recombination among these elements, pseudogene formation, and deletions of varying size. In anciently host-restricted lineages, the frequency of insertion sequence elements decreases as genomes become extremely small and strictly clonal. These changes represent a general syndrome of genome evolution, which is observed repeatedly in host-restricted lineages from numerous phylogenetic groups. Considerable variation also exists, however, in part reflecting unstudied aspects of the population structure and ecology of host-restricted bacterial lineages.

Addresses
Department of Ecology and Evolutionary Biology & Center for Insect Science, University of Arizona, Tucson, Arizona 85721, USA
*e-mail: nmoran@email.arizona.edu

Current Opinion in Genetics & Development 2004, **14**:627–633

This review comes from a themed issue on
Genomes and evolution
Edited by David Haig and Steve Henikoff

Available online 3rd October 2004

0959-437X/$ – see front matter

DOI 10.1016/j.gde.2004.09.003

Abbreviations
IS insertion sequence

Introduction

In the 'pre-genomics' era, it was already known that many bacteria had relatively large and labile genomes continuously undergoing a gain and loss of genes [1], whereas some bacterial groups had undertaken deviant evolutionary paths leading to small genomes, base compositions highly biased towards adenine and thymine (A+T), and rapid sequence evolution (e.g. see [2]). Now, the rapidly expanding set of complete bacterial genome sequences is providing ever-improving resolution to our picture of bacterial evolution in general and to the origins and nature of reduced genomes in particular.

Reduced genomes have evolved many times within several bacterial phyla in connection with their obligate dependence on a eukaryotic host as either a parasitic or mutualistic symbiont. Examples include the Mollicutes (mycoplasmas and spiroplasmas), the Firmicutes (e.g. see [3]), the Rickettsiae and Wolbachiae (α-Proteobacteria) [4,5••], the Chlamydiae [6], the Spirochetes [7••] and some insect endosymbionts such as *Buchnera aphidicola*, *Wigglesworthia glossinidia* and *Blochmannia camponotii* (all γ-Proteobacteria) [8–10,11••,12]. Many of these were little known until their genomes were sequenced in the past few years. Their distinguishing ecological attribute is that their replication is dependent on living, often intracellularly, in a host. Their transmission can be vertical (mother to progeny) and/or horizontal.

The antiquity of groups with very small genomes is supported by molecular clock estimates and, in the case of cospeciating hosts and symbionts, by host fossils. For example, *Buchnera* is over 150 million years old [13], whereas other groups, such as the Chlamydiae, Mollicutes and Rickettsiae, seem to be much older, approaching 1 billion years (e.g. see [6]). Thus, reduced genome bacteria have probably evolved for much of the period in which their eukaryotic hosts have existed.

In this review, we summarize recent findings regarding genomic changes that occur in host-restricted bacteria, particularly during the initial stages of this process.

Reconstructing genomic changes in host-dependent lineages of bacteria

The recognition that distinctive genomic features are consistently associated with host dependence raises questions of how and why these features have emerged repeatedly from 'normal' bacterial genomes. What new forces are imposed by a lifestyle confined to hosts? The answers are not likely to come from analyses of highly derived small genomes, however, because the current patterns of genome evolution in these small genomes are very different from those that characterize the early stages of host restriction.

For example, *Buchnera* genomes have been extraordinarily stable for the past 100 million years or more during their diversification within aphids, despite high levels of divergence in gene sequence [9,11••]. Similarly, mycoplasmas and Chlamydiae have unusually static genome arrangements relative to their levels of sequence divergence [14]. *Buchnera*, however, shows many chromosomal rearrangements and deletions in comparison to its free-living enteric relatives, implying that its earlier stages of

evolution were more dynamic [15]. For bacteria with highly reduced genomes, the gene order is too divergent from that of their free-living relatives to enable a reconstruction of the many rearrangements that must have occurred. Indeed, the insect endosymbionts *Buchnera* and *Blochmannia* are among the few very small genomes that have been sequenced that retain any appreciable gene order synteny with their relatives — in this case, enteric bacteria [15,16].

Recently, numerous genome sequences and other genomic data for recently host-restricted lineages have been published, and some are readily comparable to genome sequences of closely related free-living bacteria. Together, these data suggest that various changes occur very quickly after the host-restricted niche is invaded. These comparisons repeatedly indicate that the transition to strict host dependence is followed by a sharp increase in frequencies of mobile elements, coupled with many genomic rearrangements and deletions. This picture contrasts markedly with the genomic stasis and lack of mobile elements seen in some long-established, host-dependent bacterial lineages.

IS element proliferation: a key feature in the initial stages of genome reduction

Strains and species that have recently evolved as specialized pathogens often show much higher numbers of insertion sequence (IS) elements as compared with their more ecologically generalized parental strains. A consequence of this is that IS elements have been especially useful as markers for the genetic typing of pathogen strains. Figure 1 shows the estimated loads of IS elements, adjusted for genome size, for the set of fully sequenced proteobacterial genomes. The species are divided into three general categories on the basis of their lifestyles: facultatively or completely free-living, recently host-dependent with no free-living stage, and anciently host-dependent. Many of the recently host-dependent bacteria have high loads of IS elements. By contrast, ancient symbionts, with the notable exception of *Wolbachia* (see below), generally have very few.

Famous examples of this high loading of IS elements exist among pathogenic enteric bacteria, including *Shigella* and *Salmonella enterica* Typhi, which are more host-restricted than are related enterics and which have many more mobile elements than do close relatives with more generalized lifestyles [17,18••]. Large numbers of IS elements are also found in other members of the Enterobacteriaceae that are host-restricted but have very different life cycles; examples include the arthropod symbionts *Photorhabdus luminescens* and *Yersinia pestis* [19,20••]. More examples exist among genomes that have not been fully sequenced, but for which we have substantial information. For example, the enterobacterial symbionts of *Sitophilus* grain weevils, which recently acquired their symbiotic lifestyle, seem to have an unusually high incidence of IS elements (GR Plague and NA Moran, unpublished), some of which disrupt rRNA operons [21•]. These examples involve various IS families, presumably reflecting the elements that seeded in the original symbiotic ancestor [22].

In several cases, IS elements have been implicated as the source of chromosomal rearrangements and gene inactivations in bacteria. Examples include two independently derived, recently host-restricted *Bordetella* species, which are compared to a species resembling the ancestor with capability for free-living replication [23••,24•]. The derived, host-restricted strains contain more IS elements, which flank rearranged segments. Similarly, a comparison of the two sequenced *Shigella* strains shows that abundant IS elements form the basis for both pseudogene formation and chromosomal deletions and rearrangements [18••]. Other examples of rapid genomic change following the acquisition of a host-restricted lifestyle include the *Mycobacterium bovis* genome, which shows deletions of some of the same genes that have been eliminated in *Mycobacterium leprae* [25•] and *Coxiella burnetii* [26•], the agent of Q-fever.

Pseudogene proliferation and gene deletion in the initial stages of genome reduction

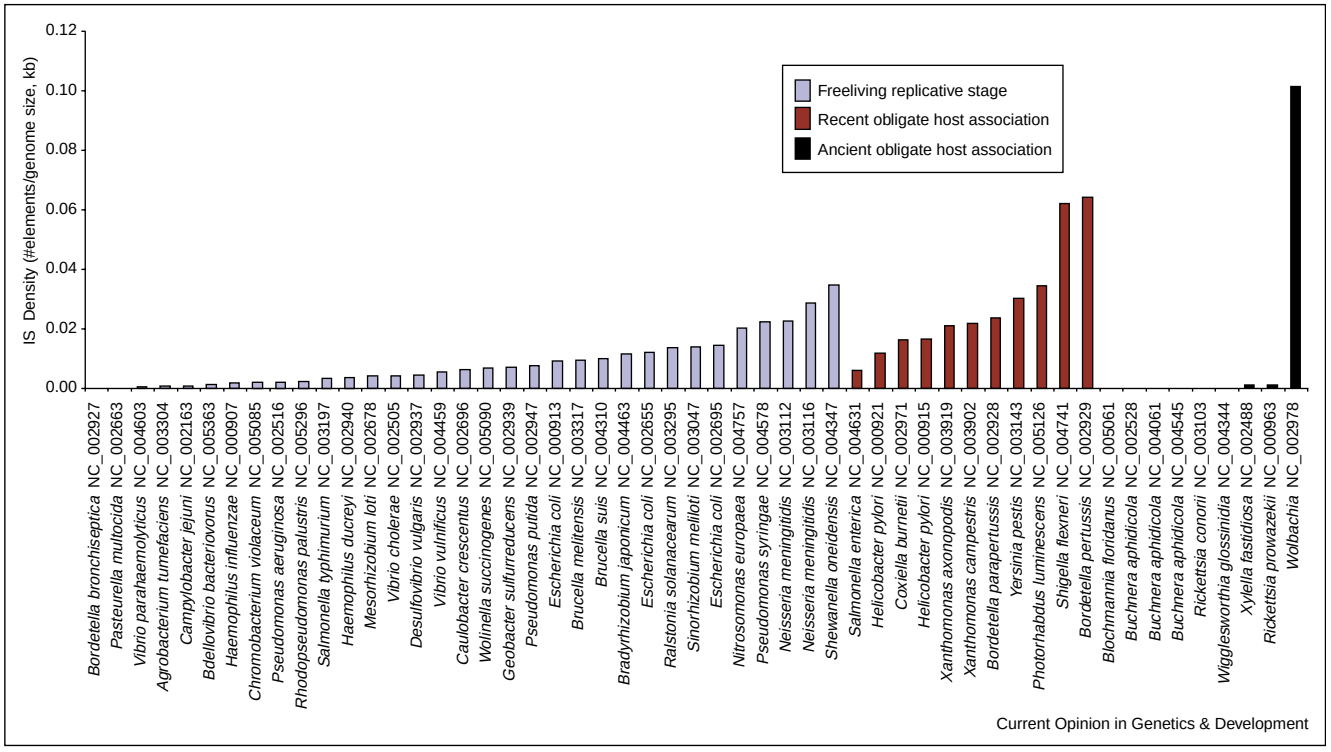
One of the most clear-cut signs that a bacterial genome is in transition is an abundance of pseudogenes. The genomes of *Rickettsia prowazekii* [4] and especially *M. leprae* [27] were the first to reveal striking numbers of pseudogenes — the remnants of a different evolutionary past. Pseudogenes are now recognized to be numerous in many host-restricted bacteria, and even extremely reduced genomes show evidence of ongoing gene inactivation (e.g. see [9,11••]).

The annotation of pseudogenes is erratic because their recognition depends on the availability of closely related genomes for comparison; thus, it is not feasible to compare frequencies within the genomes of free-living and host-dependent bacteria. However, much of the genome reduction of host-dependent bacteria can be attributed to the inactivation of individual genes, followed by their slow dwindling through deletions within the resulting pseudogene. As a result, ancestral genes are present in varying stages of decline in parasitic and symbiotic bacteria [9,23••,28,29,30•]. On the basis of observations in organisms that contain many young pseudogenes, such as the obligate pathogenic lineages of *Shigella* and *Bordetella*, the initial mutation that causes gene inactivation is most often a single base substitution, a single base insertion or deletion, or an IS transposition [18••,23••].

Why are there these changes after host restriction?

There are two reasons that selection to retain gene functions will be less effective in bacteria that lose free-living

Figure 1



Density of insertion sequences (number of IS elements per kilobase of genome) in completely sequenced proteobacterial genomes. The numbers of IS elements were taken from the original genome papers if specified; otherwise, they were obtained by searching the gene annotation database at the National Center for Biotechnology Information. These numbers will vary in accuracy and should be considered rough estimates. The species are divided into three categories according to information available on their lifestyle and the age of the clade, as specified in the main genome sequence paper. Species in the first category retain the ability to replicate in the environment independently of a host (although some stages may be host-associated); those in the second are dependent on a host for replication and are a recently derived pathogen or symbiont with very low divergence (>90% sequence identity of protein-encoding genes) from a species or strain with free-living stages; those in the third category are dependent on a host for replication and belong to a deeply branching clade of host-dependent pathogens or symbionts with a distant relationship (<80% sequence identity) to any lineages with free-living stages.

stages and that become restricted to living in hosts. First, losing the ability to live in the environment results in a much smaller effective population size, decreasing the efficiency of purifying selection and resulting in the inactivation of beneficial genes through genetic drift [31]. Second, living only in hosts reduces the intensity of purifying selection on many genes, which may be superfluous in the nutrient-rich environment provided by host tissues (e.g. see [8,32•]). This second explanation might provide the reason for loss of some genes, i.e. those with functions that might be replaced by uptake of host metabolites. However, among the genes that are eliminated from small genomes, many have functions involved in central informational processes of the cell, such as DNA repair. In these cases, gene inactivations are likely to represent slightly deleterious mutations. The overall consequence is that many more mutations that disrupt gene function can persist in host-restricted bacteria.

The spread of mobile genetic elements is governed by opposing forces: their tendency to increase as 'selfish genes' versus selection against their spread, which acts either against genotypes with transpositions that destroy the function of particular genes or in favor of general genomic mechanisms that suppress transposition [33]. The explanation put forward for IS proliferation in host-restricted bacteria hinges on the proposal that the reduced effective population size lowers the efficiency by which purifying selection maintains genes [23••]. In other words, the transition to a small population size shifts the balance of forces that govern IS numbers, causing more transpositions to become fixed by genetic drift.

In addition, host restriction may limit opportunities for the horizontal transmission of elements. In eukaryotes that lose sexual reproduction, mobile elements are predicted to decrease in frequency and ultimately to be eliminated because they lack the potential to move into new genetic backgrounds [34]; similarly, in bacteria that never exchange DNA with other bacteria, no new IS elements enter the genome. In strictly asexual genomes, existing elements either will lead to extinction of the host

lineage or will undergo inactivation and dwindle over time, owing to harmful effects at the host level.

Thus, proliferation is expected in the short run, owing to diminished selection for gene maintenance, but inactivation and loss of elements are expected in the longer run, owing to the lack of opportunity for horizontal spread. Under this view, the anciently host-restricted genomes (Figure 1) have persisted through a stage of IS spread, but traces of these sequences are now deleted or mutated beyond recognition because of the lack of exposure to novel element transposition (although close inspection may reveal remnants even in the most reduced genomes such as *Buchnera* [22]).

These considerations may help to explain the exceptional case of *Wolbachia*, which shows typical features of anciently host-restricted bacteria, including small genome size, rapid sequence evolution, A+T bias, and apparent great age (>50 million years). The single published sequence of a *Wolbachia* genome, from an insect host, retains very large numbers of mobile elements [5••]. This may reflect the fact that, in contrast to the other anciently host-restricted lineages shown in Figure 1, several strains of arthropod-associated *Wolbachia* can coinfect individual insect hosts, where they undergo recombination and potentially spread elements between different strains [35–38]. Possibly as a result of its abundance of mobile elements, *Wolbachia* shows many lineage-specific gene rearrangements [5••]. In contrast to the arthropod-associated lineages, *Wolbachia* that are symbiotic in filarial nematodes seem to be transmitted strictly vertically [39], and it will be interesting to see whether these genomes possess fewer IS elements.

Consequences of mobile element proliferation

The proliferation of mobile elements has consequences. First, as evident in the examples cited above, IS transposition often inactivates genes, which are later degraded and thereby contribute to genome reduction. In some cases, gene deletions may represent adaptation to the host-restricted niche [40], and may be attributable to IS transposition (e.g. see [41•]). Second, as illustrated by comparisons of related genomes of host-restricted and free-living organisms, repeated transposition of an element results in regions of homology scattered along the chromosome and sets the stage for high rates of intrachromosomal homologous recombination [42•]. This yields higher spontaneous rates of rearrangements and deletions of fragments, as observed in the host-restricted *Bordetella*, in which IS elements flank inversions, deletions and other rearrangements [23••]. Third, many IS elements carry outwardly directed promoters that activate the expression of neighboring genes [22], including genes involved in catabolic pathways (e.g. see [43]), antibiotic resistance (e.g. see [44,45]) and pathogenicity (e.g. see [46•]). Although most such spontaneous IS-mediated changes will be deleterious, all three effects of IS transposition can be also beneficial.

Particularly interesting is the question of why some genes are deleted but others are retained in newly host-restricted lineages. The patterns observed reflect both mutational pressure and selection for retaining or deleting a region. Microarray assays of *Mycobacterium tuberculosis* isolates have shown that the clustering of deletions in some regions has occurred independently in evolutionarily distinct strains, suggesting that selection does drive some deletions [47••].

The exceptions

Long-term host dependence is clearly linked to genome reduction, loss of mobile elements, A+T bias and rapid sequence evolution, but these features are not uniform in all host-dependent groups. The loads of IS elements and the rearrangements that characterize newly host-dependent groups are highly variable; for example, different strains of *M. tuberculosis* show very different levels of IS elements [48]. Some of this variation may fit with general expectations for the behavior of selfish elements under different population conditions, as suggested above for *Wolbachia*. Usually, however, we know little about the genetic structure of pathogen and symbiont populations.

Clearly, the degree of clonality varies markedly among pathogenic and symbiotic bacteria [49•], and this is expected to affect the spread of mobile elements. In addition, some variability probably reflects idiosyncratic variation in the initial presence of IS elements in the genome of a newly host-dependent lineage, as well as the translocation modes of the resident IS elements, with conservative or 'cut-and-paste' elements presumably imposing a generally lower load than replicatively transposing elements.

Conclusions

The first reaction on discovering a distinctive genomic feature is to cite a possible adaptive basis. But genomic characteristics can also reflect a lack of adaptation, particularly when population structure or ecological conditions result in ineffective or weak selection on much of the genome. For example, *Shigella* strains show similar metabolic features because of the convergent loss of catabolic or biosynthetic functions that are not needed in the intracellular niche; how much of this convergence is adaptive and how much is neutral is not immediately clear [50]. Sorting out which changes are driven by selection and which reflect genetic drift is not simple but is fundamental to our understanding of what makes a particular bacterium pathogenic or symbiotic. Possibly, one or a few key changes underlie the change in lifestyle, with hundreds of others representing the subsequent repercussions of the altered population structure.

IS elements may be the most noticeable genomic features rather than the most biologically important ones. When large quantities of IS element were initially documented in pathogenic bacteria such as *M. tuberculosis* and *Shigella flexneri*, they were considered a possible means of generating plasticity for opportunistic organisms evolving to overcome host defenses, and such explanations are still common (e.g. see [51•,52•]). Because IS elements confer a high mutation rate for inactivation and for deletions of large chromosome fragments, they are expected occasionally to provide the mutational basis for adaptation (e.g. see [40]). But, similar to other sources of mutation, most transpositions are expected to be deleterious to fitness, and only under special circumstances can evolutionary models identify an advantage to a persistent increase in mutation rate. Furthermore, most adaptive explanations for IS proliferation, which usually cite a need for plasticity to counter host defense mechanisms, do not account for why similar changes are found in mutualistic symbionts (e.g. see [21•]). Hosts should be evolving to support, rather than to eliminate, mutualists.

The proliferation of mobile elements is expected to generate a high mutation rate for large deletions with little opposing selection, and massive reduction may be the consequence. This interpretation — that most changes can be attributed to genetic drift rather than to adaptation — is consistent with most information that is available on the population structure and ecology of host-restricted groups (e.g. see [53,54]). More ecological and population genetic information would be extremely useful in identifying the basis of genomic change in evolving pathogens and symbionts. As more genomic sequence data are produced for organisms with different ecological associations, we will gain ever better focus in our view of genomic changes that both cause and result from host restriction.

Acknowledgements

We thank Howard Ochman for discussion of this manuscript. Financial support during the writing of this review was from NSF 0313737 to NM and from an NIH PERT fellowship to GRP through the Center for Insect Science, University of Arizona.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Ochman H, Bergthorsson U: **Genome evolution in enteric bacteria**. *Curr Opin Genet Dev* 1995, **5**:734-738.

2. Woese CR: **Bacterial evolution**. *Microbiol Rev* 1987, **51**:221-271.

3. Oshima K, Kakizawa S, Nishigawa H, Jung HY, Wei W, Suzuki S, Arashida R, Nakata D, Miyata S, Ugaki M, Namba S: **Reductive evolution suggested from the complete genome sequence of a plant-pathogenic phytoplasma**. *Nat Genet* 2004, **36**:27-29.

4. Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Ponten T, Alsmark UC, Podowski RM, Naslund AK, Eriksson AS, Winkler HH, Kurland CG: **The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria**. *Nature* 1998, **396**:133-140.

5. •• 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.
This is an odd genome; it is overrun with apparent mobile elements, but otherwise has many features of the highly reduced genomes of long-term symbionts such as a small size and a base compositional bias favoring A+T.

6. Horn M, Collingro A, Schmitz-Esser S, Beier CL, Purkhold U, Fartmann B, Brandt P, Nyakatura GJ, Droege M, Frishman D *et al.*: **Illuminating the evolutionary history of chlamydiae**. *Science* 2004, **304**:728-730.

7. •• Seshadri R, Myers GS, Tettelin H, Eisen JA, Heidelberg JF, Dodson RJ, Davidsen TM, DeBoy RT, Fouts DE, Haft DH *et al.*: **Comparison of the genome of the oral pathogen *Treponema denticola* with other spirochete genomes**. *Proc Natl Acad Sci USA* 2004, **101**:5646-5651.
The comparison of *T. denticola* with *T. pallidum* indicates that gene loss and rearrangements can be important in the genome evolution of host-restricted pathogens and provides an example in which specialized pathogens continue to show dynamic genomic change.

8. 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**:81-86.

9. 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.

10. 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.

11. •• van Ham RC, Kamerbeek J, Palacios C, Rausell C, Abascal F, Bastolla U, Fernandez JM, Jimenez L, Postigo M, Silva FJ *et al.*: **Reductive genome evolution in *Buchnera aphidicola***. *Proc Natl Acad Sci USA* 2003, **100**:581-586.
This comparison of the three completed *Buchnera* genomes demonstrates that continued, slow rates of gene loss can be ongoing even in the extremely reduced genomes of organisms that are anciently host-dependent.

12. 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.

13. Moran NA, Munson MA, Baumann B, Ishikawa H: **A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts**. *Proc R Soc Lond B Biol Sci* 1993, **253**:167-171.

14. Suyama M, Bork P: **Evolution of prokaryotic gene order: genome rearrangements in closely related species**. *Trends Genet* 2001, **17**:10-13.

15. Moran NA, Mira A: **The process of genome shrinkage in the obligate symbiont, *Buchnera aphidicola***. *Genome Biol* 2001, **2**:research0054.1-research0054.12.

16. Silva FJ, Latorre A, Moya A: **Genome size reduction through multiple events of gene disintegration in *Buchnera* APS**. *Trends Genet* 2001, **17**:615-618.

17. Jin Q, Yuan Z, Xu J, Wang Y, Shen Y, Lu W, Wang J, Liu H, Yang J, Yang F *et al.*: **Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity through comparison with genomes of *Escherichia coli* K12 and O157**. *Nucleic Acids Res* 2002, **30**:4432-4441.

18. •• Wei J, Goldberg MB, Burland V, Venkatesan MM, Deng W, Fournier G, Mayhew GF, Plunkett G III, Rose DJ, Darling A *et al.*: **Complete genome sequence and comparative genomics of *Shigella flexneri* serotype 2a strain 2457T**. *Infect Immun* 2003, **71**:2775-2786.
The *Shigella* strains represent host-restricted strains derived from *E. coli*. They are characterized by an abundance of insertion sequence (IS)

elements, some mediating chromosomal rearrangements and some causing gene inactivations. In this strain, most of the recent pseudogenes are attributed to base substitutions.

19. Parkhill J, Wren BW, Thomson NR, Titball RW, Holden MTG, Prentice MB, Sebaihia M, James KD, Churcher C, Mungall KL *et al.*: **Genome sequence of *Yersinia pestis*, the causative agent of plague**. *Nature* 2001, **413**:523-527.

20. •• Duchaud E, Rusniok C, Frangeul L, Buchrieser C, Taourit S, Bocs S, Boursaux-Eude C, Chandler M, Dassa E, Derose R *et al.*: **The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens***. *Nat Biotechnol* 2003, **21**:1307-1313.

This genome, like that of *Yersinia pestis*, shows that symbiotic bacteria of invertebrates can share many of the features of host-restricted mammalian pathogens, including an increase in insertion sequence element load and chromosomal rearrangements.

21. • Dale C, Wang B, Moran N, Ochman H: **Loss of DNA recombinational repair enzymes in the initial stages of genome degeneration in mutualistic bacterial endosymbionts**. *Mol Biol Evol* 2003, **20**:1188-1194.

This obligate symbiont of weevils, a member of the enteric bacteria, shows remarkable changes owing to the transposition of insertion sequence elements, resulting in the inactivation of some rRNA operons. Among the inactivated genes is *recA*, and loss of homologous recombination is likely to underlie some of the observed changes, such as the unusual divergence of rRNA operons in the genome. This illustrates a common feature of reduced genomes: namely, the elimination of repair gene function can lead to an increase in mutations and possibly to a faster accumulation of genomic changes.

22. Chandler M, Mahillon J: **Insertion sequences revisited**. In *Mobile DNA II.* Edited by Craig NL, Craigie R, Gellert M, Lambowitz AM. Washington DC: ASM Press; 2002:305-366.

23. •• Parkhill J, Sebaihia M, Preston A, Murphy LD, Thomson N, Harris DE, Holden MTG, Churcher CM, Bentley SD, Mungall KL *et al.*: **Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica***. *Nat Genet* 2003, **35**:32-40.

This paper, based on three complete *Bordetella* genome sequences, provides the best comparison so far between genomes that differ in host restriction. *B. pertussis* and *B. parapertussis* are independently derived, host-restricted species that evolved from *B. bronchiseptica*, which can live independently of hosts in the environment. The derived strains have undergone a considerable increase in insertion sequence (IS) elements, the inactivation of many genes resulting in several pseudogenes (358 and 220, respectively, or 5–10% of open reading frames) and numerous deletions of up to 120 kb. The deletions have caused an overall decrease in genome size, from 5.3 Mb in *B. bronchiseptica* to 4.1 Mb and 4.8 Mb in the two host-restricted lineages. Many chromosomal rearrangements have occurred in the two host-restricted strains, particularly in *B. pertussis*. IS elements positioned at the boundaries of nearly all rearranged and deleted regions provide strong evidence that homologous recombination between these sequences is the mechanistic basis for the inversions, translocations and deletions. Mobile element transposition has also apparently caused substantial gene inactivation in the host-dependent *Bordetella*, with most identifiable pseudogenes containing IS elements. These extensive genomic changes have occurred over an estimated 0.3 to 4 million years on the basis of the sequence divergence of homologous genes.

24. • Cummings CA, Brinig MM, Lepp PW, van de Pas S, Relman DA: ***Bordetella* species are distinguished by patterns of substantial gene loss and host adaptation**. *J Bacteriol* 2004, **186**:1484-1492.

This analysis shows that *B. bronchiseptica* is a paraphyletic taxon with some diversity and that *B. pertussis* and *B. parapertussis* represent two independently derived clades of host-restricted pathogens. The authors propose that gene loss, mediated by transposable elements, has been important in adaptation of the strains.

25. • Garnier T, Eiglmeier K, Camus JC, Medina N, Mansoor H, Pryor M, Duthoy S, Grondin S, Lacroix C, Monsempe C *et al.*: **The complete genome sequence of *Mycobacterium bovis***. *Proc Natl Acad Sci USA* 2003, **100**:7877-7882.

M. bovis seems to have evolved from an ancestor very close to *M. tuberculosis*, from which it has diverged very little at homologous genes (>99.9% sequence identity). Owing to gene deletion, it has lost metabolic self-sufficiency, and it represents a very recent stage of genome deterioration in a pathogen, in contrast to the highly degraded genome of *M. leprae* described in [27].

26. • Seshadri R, Paulsen IT, Eisen JA, Read TD, Nelson KE, Nelson WC, Ward NL, Tettelin H, Davidsen TM, Beanan MJ *et al.*: **Complete genome sequence of the Q-fever pathogen *Coxiella burnetii***. *Proc Natl Acad Sci USA* 2003, **100**:5455-5460.

This is another apparent example of genomic degradation following host restriction; the 2 Mb genome contains relatively high loads of both insertion sequence elements (32 in total) and pseudogenes (83 in total).

27. Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, Honore N, Garnier T, Churcher C, Harris D *et al.*: **Massive gene decay in the leprosy bacillus**. *Nature* 2001, **409**:1007-1011.

28. Mira A, Ochman H, Moran NA: **Deletional bias and the evolution of bacterial genomes**. *Trends Genet* 2001, **17**:589-596.

29. Amiri H, Alsmark CM, Andersson SGE: **Proliferation and deterioration of *Rickettsia* palindromic elements**. *Mol Biol Evol* 2002, **19**:1234-1243.

30. • Amiri H, Davids W, Andersson SGE: **Birth and death of orphan genes in *Rickettsia***. *Mol Biol Evol* 2003, **20**:1575-1587.

Comparing the two completed *Rickettsia* genomes shows that sequence repeats are important in mediating deletions of chromosomal fragments. Gene loss is ongoing in these small genomes.

31. Moran NA: **Accelerated evolution and Muller's ratchet in endosymbiotic bacteria**. *Proc Natl Acad Sci USA* 1996, **93**:2873-2878.

32. • Rio RV, Lefevre C, Heddi A, Aksoy S: **Comparative genomics of insect-symbiotic bacteria: influence of host environment on microbial genome composition**. *Appl Environ Microbiol* 2003, **69**:6825-6832.

The use of microarrays to examine gene inventories reveals extensive gene loss in two insect symbionts: *Sodalis glossinidius* of tsetse flies; and *Sitophilus oryzae*, the obligate symbiont of the rice grant weevil.

33. Campbell A: **Eubacterial genomes**. In *Mobile DNA II.* Edited by Craig NL, Craigie R, Gellert M, Lambowitz AM. Washington DC: ASM Press; 2002:1024-1038.

34. Nuzhdin SV, Petrov DA: **Transposable elements in clonal lineages: lethal hangover from sex**. *Biol J Linn Soc* 2003, **79**:33-41.

35. Jiggins FM: **The rate of recombination in *Wolbachia* bacteria**. *Mol Biol Evol* 2002, **19**:1640-1643.

36. Werren JH, Bartos JD: **Recombination in *Wolbachia***. *Curr Biol* 2001, **11**:431-435.

37. Jamnongluk W, Kittayapong P, Baimai V, O'Neill SL: ***Wolbachia* infections of tephritid fruit flies: molecular evidence for five distinct strains in a single host species**. *Curr Microbiol* 2002, **45**:255-260.

38. Reuter M, Keller L: **High levels of multiple *Wolbachia* infection and recombination in the ant *Formica exsecta***. *Mol Biol Evol* 2003, **20**:748-753.

39. Bandi C, Anderson CG, Blaxter ML: **Phylogeny of *Wolbachia* in filarial nematodes**. *Proc R Soc Lond B Biol Sci* 1998, **265**:2407-2413.

40. Maurelli AT, Fernández RE, Bloch CA, Rode CK, Fasano A: **'Black holes' and bacterial pathogenicity: a large genomic deletion that enhances the virulence of *Shigella* spp. and enteroinvasive *Escherichia coli***. *Proc Natl Acad Sci USA* 1998, **95**:3943-3948.

41. • Kresse AU, Dinesh SD, Larbig K, Romling U: **Impact of large chromosomal inversions on the adaptation and evolution of *Pseudomonas aeruginosa* chronically colonizing cystic fibrosis lungs**. *Mol Microbiol* 2003, **47**:145-158.

Strains that cause chronic infections show accelerated chromosomal evolution, mediated by transposition of IS*6100*. These elements form the basis for both chromosomal rearrangements and the inactivation of numerous genes.

42. • Hinchliffe SJ, Isherwood KE, Stabler RA, Prentice MB, Rakin A, Nichols RA, Oyston PC, Hinds J, Titball RW, Wren BW: **Application of DNA microarrays to study the evolutionary genomics of *Yersinia pestis* and *Yersinia pseudotuberculosis***. *Genome Res* 2003, **13**:2018-2029.

The authors propose that *Y. pestis* evolved quickly from *Y. pseudotuberculosis*, a mammalian enteropathogen that occurs widely in the environment, to become a more host-restricted lineage that is symbiotic in insects and blood-borne in mammals, and has limited ability to live outside these hosts. The authors identify numerous genomic differences between *Y. pestis* and *Y. pseudotuberculosis* and suggest that ''genome rearrangements, particularly as a result of the recombination of insertion sequence elements and the accumulation of pseudogenes, may also have played a significant role in the rapid evolution of *Y. pestis*'', which is estimated to have diversified over a period of only 20 000 years.

43. Kallastu A, Hõrak R, Kivisaar M: **Identification and characterization of IS*1411*, a new insertion sequence which causes transcriptional activation of the phenol degradation genes in *Pseudomonas putida***. *J Bacteriol* 1998, **180**:5306-5312.

44. Podglajen I, Breuil J, Rohaut A, Monsempes C, Collatz E: **Multiple mobile promoter regions for the rare carbapenem resistance gene of *Bacteroides fragilis***. *J Bacteriol* 2001, **183**:3531-3535.

45. Katayama Y, Takeuchi F, Ito T, Ma XX, Ui-Mizutani Y, Kobayashi I, Hiramatsu K: **Identification in methicillin-susceptible *Staphylococcus hominis* of an active primordial mobile genetic element for the staphylococcal cassette chromosome *mec* of methicillin-resistant *Staphylococcus aureus***. *J Bacteriol* 2003, **185**:2711-2722.

46. • Safi H, Barnes PF, Lakey DL, Shams H, Samten B, Vankayalapati R, Howard ST: **IS*6110* functions as a mobile, monocyte-activated promoter in *Mycobacterium tuberculosis***. *Mol Microbiol* 2004, **52**:999-1012.

This study provides an example of a mobile element that mediates apparent adaptive change. In some *M. tuberculosis* strains, IS*6110* upregulates the *Rv1468c* gene, which plays a presumptive role in evading host immune recognition.

47. •• Tsolaki AG, Hirsh AE, DeRiemer K, Enciso JA, Wong MZ, Hannan M, de la Salmoniere YOL, Aman K, Kato-Maeda M, Small PM: **Functional and evolutionary genomics of *Mycobacterium tuberculosis*: insights from genomic deletions in 100 strains**. *Proc Natl Acad Sci USA* 2004, **101**:4865-4870.

On the basis of microarray assays of 100 *M. tuberculosis* clinical isolates, deletions are shown to be strongly clustered on the chromosome, on the scale of 15 kb and less. Some clusters are confined to a particular phylogenetic lineage and are possibly due to an insertion sequence element localized in one region. But most clusters involve particular regions that incur deletions in distinct lineages, indicating that region-specific properties have made them more vulnerable to deletion. This can be explained if the regions contain genes that are neutral or negative with respect to fitness in host-associated strains. This study is one of a few in which sufficient strains have been studied to determine whether the distribution of deletion frequencies is repeated in independent lineages of host-restricted bacteria.

48. Dale JW, Al-Ghusein H, Al-Hashmi S, Butcher P, Dickens AL, Drobniewski F, Forbes KJ, Gillespie SH, Lamprecht D, McHugh TD *et al.*: **Evolutionary relationships among strains of *Mycobacterium tuberculosis* with few copies of IS*6110***. *J Bacteriol* 2003, **185**:2555-2562.

49. • Feil EJ, Cooper JE, Grundmann H, Robinson DA, Enright MC, Berendt T, Peacock SJ, Smith JM, Murphy M, Spratt BG *et al.*: **How clonal is *Staphylococcus aureus*?**. *J Bacteriol* 2003, **185**:3307-3316.

This paper is one of several from this group that address the extent of clonality versus recombination in pathogenic bacteria, an issue that has repercussions for the extent of genetic drift and the fixation of novel mutations. The data presented here show that point mutations are 15 times more important than recombination in producing new alleles in *S. aureus*. The relative impact of recombination and mutation in producing novelty varies among species, with recombination much more important in the previously studied species *Neisseria meningitidis* and *Streptococcus pneumoniae*. This study also documents high levels of nonsynonymous substitutions between isolates belonging to the same clonal complex, suggesting that purifying selection is inefficient; in other words, high levels of genetic drift are occurring in these host-restricted lineages, enabling deleterious alleles to become common.

50. Lan R, Reeves PR: ***Escherichia coli* in disguise: molecular origins of *Shigella***. *Microbes Infect* 2002, **4**:1125-1132.

51. • Kozitskaya S, Cho SH, Dietrich K, Marre R, Naber K, Ziebuhr W: **The bacterial insertion sequence element IS*256* occurs preferentially in nosocomial *Staphylococcus epidermidis* isolates: association with biofilm formation and resistance to aminoglycosides**. *Infect Immun* 2004, **72**:1210-1215.

Clinical isolates of *Staphylococcus epidermidis*, which have a relatively restricted occurrence in hosts, show relatively high frequencies of IS*256*.

52. • Holden MT, Feil EJ, Lindsay JA, Peacock SJ, Day NP, Enright MC, Foster TJ, Moore CE, Hurst L, Atkin R *et al.*: **Complete genomes of two clinical *Staphylococcus aureus* strains: evidence for the rapid evolution of virulence and drug resistance**. *Proc Natl Acad Sci USA* 2004, **101**:9786-9791.

This comparison of a more generalized 'community acquired' strain with a more restricted clinical strain of *S. aureus* illustrates the link between mobile elements, genomic plasticity and the rapid evolution of host-restricted bacteria.

53. Achtman M, Zurth K, Morelli G, Torrea G, Guiyoule A, Carniel E: ***Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis***. *Proc Natl Acad Sci USA* 1999, **96**:14043-14048.

54. Girard JM, Wagner DM, Vogler AJ, Keys C, Allender CJ, Drickamer LC, Keim P: **Differential plague-transmission dynamics determine *Yersinia pestis* population genetic structure on local, regional, and global scales**. *Proc Natl Acad Sci USA* 2004, **101**:8408-8413.