The Neolithic revolution of bacterial genomes

Alex Mira, Ravindra Pushker and Francisco Rodríguez-Valera

Evolutionary Genomics Group, Division of Microbiology, Universidad Miguel Hernandez, San Juan 03550, Alicante, Spain

Current human activities undoubtedly impact natural ecosystems. However, the influence of Homo sapiens on living organisms must have also occurred in the past. Certain genomic characteristics of prokaryotes can be used to study the impact of ancient human activities on microorganisms. By analyzing DNA sequence similarity features of transposable elements, dramatic genomic changes have been identified in bacteria that are associated with large and stable human communities, agriculture and animal domestication: three features unequivocally linked to the Neolithic revolution. It is hypothesized that bacteria specialized in human-associated niches underwent an intense transformation after the social and demographic changes that took place with the first Neolithic settlements. These genomic changes are absent in related species that are not specialized in humans.

The Neolithic revolution

Before the Neolithic period, human survival was linked to the hunter-gatherer culture and populations were small and scattered [1]. Approximately 10 000 years ago, however, the advent of agriculture and animal husbandry brought the largest social revolution in the history of humankind [2]. Food resources were more abundant and constant, and the human species increased its population size at an extraordinary annual growth rate of 0.1% [3]. From the point of view of bacterial pathogens, humans suddenly became attractive hosts; they concentrated large populations on limited areas, which maximized the chance for transmission between longer-lived carriers. Thus, it is likely that human population growth and expansion during the Neolithic created a selective pressure that favoured pathogens that specialized in human hosts, originating what was probably the first wave of emerging human diseases [4].

One of the features that has emerged from fully sequenced genomes is that species that have specialized their niche undergo a process of reductive evolution through gene elimination [5–8]. Mobile elements such as insertion sequences (ISs) greatly increase in number following host restriction [9] and influence the process of genome reduction. After a niche change, many genes become obsolete and the selective pressure to preserve their function is dramatically reduced. In addition, host restriction can reduce the efficiency of selection in bacterial specialists because of reduced population sizes and genetic drift [10,11]. ISs might inactivate unnecessary or redundant genes in the new environment [12,13] by rapidly increasing the number of IS copies. The number of IS elements in bacterial pathogens could, therefore, indicate a recent host adaptation [14] or restriction [9]. Here, we show that recent expansions of ISs in bacteria are associated with humans or human activities that were developed in the Neolithic. The cultural revolution of the Neolithic can be dated precisely (for evolutionary time scales) at ~10 000 years before present, which provides an outstanding model for prokaryotic evolution in the short-to-medium time scale.

Recent expansions of mobile elements

We investigated paralogous genes, including IS elements, across 255 fully sequenced prokaryotic genomes. Paralogues are defined as protein-coding sequences within a genome that have at least 30% sequence identity over...
> 60% of their lengths [15–17]. To estimate DNA sequence similarity, global alignments were performed on the gene pairs using Stretcher [18]. When a frequency histogram of paralogue sequence similarity was plotted, we observed that most paralogues across taxa display 40–60% DNA sequence similarity, followed by a second group of duplicated genes at 90–100% similarity. This second group presumably contains recently formed paralogues and a few genes with a high protein-dosage requirement, the DNA sequence of which is kept identical by gene conversion mechanisms [19]. However, a closer look reveals that most of the members in this group are ISs or IS-transposase genes. Interestingly, 69 out of 89 bacteria in which > 75% of the recently formed paralogous genes correspond to ISs are species that are uniquely associated with humans (Figure 1). Most of the other cases correspond to host-associated species such as invertebrate symbionts or pathogens. These include the mutualist *Photorhabdus luminescens*, which is specifically associated with nematodes, and the reproductive parasite of insects *Wolbachia* *pipiens* (Figure 2a). This indicates that any niche specialization normally involves IS expansions [9]. A few cases of free-living species with recent IS expansions are also observed, such as the sediment-isolated species *Shewanella oneidensis* or *Geobacter metallireducens* (Figure 2b). Examples of pathogens specialized in humans (Figure 1) include *Bordetella pertussis*, a species that has recently undergone an extraordinary paralogue expansion only as a result of ISs (Figure 1a); however, *Bordetella bronchiseptica*, a species that can infect different hosts and live freely in the environment, has a normal distribution of DNA similarity across paralogues. The IS expansion not only inactivates genes by insertion but also increases the rate of genome rearrangements [11]. The human pathogen *Shigella flexneri*, which is closely related to the saprophytic *Escherichia coli* K12 strain, has recently undergone another IS expansion (Figure 1b). We hypothesize that mobile element expansion took place in Neolithic times when ancestors of these two species independently restricted their host range to humans. Three lines of evidence support the hypothesis that the expansion occurred at that moment. First, IS expansions are also observed in bacteria associated with domestication processes typical of the Neolithic transition. Second, the substitution rate between IS elements is virtually zero, which is consistent with a single, recent event. And third, divergence estimates between IS-loaded human specialists and IS-free generalists are consistent with the Neolithic hypothesis.

**Pathogens of domesticated plants and livestock**

Pathogens of agricultural plants and domesticated animals also seem to have undergone recent episodes of IS expansion. Thus, the genomes burdened with ISs are those related to the three main changes that are
unequivocally linked to the Neolithic transition: agriculture, animal domestication and the formation of larger human communities. For example, several sequenced species of *Pseudomonas* are versatile microorganisms that live in soil or water. However, *Pseudomonas syringae* DC3000 (a pathovar that is specialized for tomato plants [20]) has undergone a massive episode of IS expansion, with most mobile elements sharing >98% sequence similarity (Figure 3a). Because crops are abundant and concentrated on certain regions, they are likely to represent an attractive niche for plant pathogens. Such pathogens include different species of *Xanthomonas* such as *Xanthomonas axonopodis* pv. citri, which is specialized in citrus trees and contains an extraordinary number of IS elements [21]. The rice pathogen *Xanthomonas oryzae* is even more extreme and contains >800 ISs [22]. Although the sequences of related species that are not plant specialists are unavailable for comparison, the reason behind this genomic change is likely to be related to the specialization of *Xanthomonas* in rice plants after the expansion of this staple food across Asia.

The genomes of bacteria specialized in farm animals also seem to be affected. The species *Burkholderia mallei* is an obligate parasite of horses, donkeys and mules, with no other known natural reservoir [23]. Thus, it seems likely that it specialized in these animals after the wild ancestors of horses were domesticated, which increased their population size and lifespan. The expansion of IS elements is also dramatic in this species (Figure 3b), whereas its sister species – the soil inhabitant *Burkholderia pseudomallei* – contains a limited number of IS elements [24]. Because the domestication of horses seems to have occurred relatively recently in human history (∼4500 years ago [25]), the speed of these genomic changes mediated by ISs must have been extraordinary.

Furthermore, bacteria associated with food production also seem to have undergone transposase amplification at a genomic scale. The domesticated lactic-acid bacterium *Lactococcus lactis* [26] is the most commonly used cheese starter in the world. This species has been used for generations to inoculate cheese, which has effectively maintained a domestic culture of the bacterium that is associated with human populations. When compared with sequenced related species, *L. lactis* shows a distinctive recent expansion of IS elements (Figure 3c). We predict that other bacterial species that are associated with traditional human food (such as those involved in yoghurt or wine production and the preservation of foods in brine or salt) might have undergone similar mobile-element expansions (e.g. Ref. [27]).
Predicting niche specialization

As more information accumulates about species that have experienced genomic changes linked to ISs, it can be used in the reverse direction: recent lifestyle changes in bacteria could be predicted and ISs might even provide an approximate time scale for these genomic events. For
example, the genome of the legume symbiont *Bradyrhizobium japonicum* displays a large expansion of nearly identical transposases whereas another nodule-forming symbiont, *Mesorhizobium loti*, contains older transposases (Figure 4a). If our hypothesis is correct, the reason for this could be related to the formation of *B. japonicum* nodules in the widely cultivated soya bean. By contrast, *M. loti* has not been domesticated, is associated with several wild plants and has a wider host range [28]. The genome of the halophilic archaeon *Halobacterium salinarum* shows a recent IS expansion when compared with related species (Figure 4b). Although it could be claimed that this species is not related to humans, a closer look at its niche reveals that *H. salinarum* was isolated from salted fish. The extraction of salt is one of the oldest human activities for preserving food and was used by the ancient Chinese as long ago as 9000 BC [29]. As environments characterized by high salt concentrations expanded, a new human-associated niche extended across the globe and prokaryotes could have adapted to it. Similarly, genomic features such as IS expansions in other genomes could serve as hints of ecological changes.

The opposite scenario is also interesting. There might be cases of bacteria that are clearly specialized in humans but have no traces of IS expansion. *Mycobacterium leprae* (the causative agent of leprosy) could be one of the few examples. It contains no ISs even though they could have expanded across the two-thirds of the genome that are no longer necessary in its specialized niche [6]. Instead, the genome of *M. leprae* has undergone an incomplete reductive evolution process and has retained 1000 pseudogenes with a near-normal length. Perhaps the intracellular lifestyle has precluded it from incorporating ISs and the extraordinary number of pseudogenes is because of the absence of transposases, which has slowed down the pace of genome reduction. IS elements, when present, quickly inactivate genes [13] and regulatory elements [30], and serve as repetitive sequences that induce large deletions [31]. Thus, the expansion of ISs might function as a mechanism for facilitating ecological adaptation.

### Dating the IS expansion

Dating divergence in bacterial lineages is a difficult task as a result of (among other things) the absence of a fossil record, the difficulty of calibrating a molecular clock and the meagre knowledge about generation times in nature [32,33]. The ribosomal 16S rRNA gene is estimated to diverge at a rate of ~1–2% per 50 million years [32–35], using geological events that could be dated between 50 million and 500 million years. Although this gene seems to be too conserved to date relatively recent events, it is interesting to note that the observed divergences are similar among the different pairs studied here:

---

*Plots for all studied species are available from the authors upon request.*

---

*Example plots for *Bradyrhizobium japonicum* USDA 110 and *Mesorhizobium loti* MAFF303099; *Halobacterium salinarum* NRC-1 and *Haloarcula marismortui* ATCC 43049. Species on the left show recent IS element expansions that are not found in closely related species (right), which is hypothesized to be the result of niche specialization.*

---

Figure 4. IS expansions in species under potential human influence. DNA sequence similarity is indicated between pairs of paralogous genes (blue crosses) and between pairs of IS elements (red crosses) across selected bacterial species. Chromosomal location of the first gene for each pair is shown on the x-axis. (a) *Bradyrhizobium japonicum* USDA 110 and *Mesorhizobium loti* MAFF303099; (b) *Halobacterium salinarum* NRC-1 and *Haloarcula marismortui* ATCC 43049. Species on the left show recent IS element expansions that are not found in closely related species (right), which is hypothesized to be the result of niche specialization.
Table 1. Sequence divergence in IS-interrupted ORFs and in fully sequenced genomes of putative Neolithic species

<table>
<thead>
<tr>
<th>Species with IS-interrupted ORFs</th>
<th>Species with functional homologue</th>
<th>N(^a)</th>
<th>F(^b)</th>
<th>Average similarity(^d)</th>
<th>H(^c)</th>
<th>Average similarity(^f)</th>
<th>Average genomic dS(^g)</th>
<th>Divergence estimate(^w)</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yersinia pestis CO92</td>
<td>Yersinia pestis KIM</td>
<td>26</td>
<td>23</td>
<td>99.992</td>
<td>3442</td>
<td>99.872</td>
<td>0.00008 ± 0.00002</td>
<td>6000 years</td>
<td>[37,38]</td>
</tr>
<tr>
<td>Yersinia pestis CO92</td>
<td>Yersinia pseudotuberculosis</td>
<td>26</td>
<td>24</td>
<td>99.756</td>
<td>3413</td>
<td>99.154</td>
<td>0.00217 ± 0.00002</td>
<td>11000 years</td>
<td>[37,38]</td>
</tr>
<tr>
<td>Shigella flexneri 2a</td>
<td>Escherichia coli K12</td>
<td>77</td>
<td>56</td>
<td>97.399</td>
<td>3195</td>
<td>98.403</td>
<td>0.0436 ± 0.00076</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Burkholderia mallei</td>
<td>Burkholderia pseudomallei</td>
<td>43</td>
<td>41</td>
<td>99.584</td>
<td>2535</td>
<td>99.214</td>
<td>0.01647 ± 0.00036</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Bordetella pertussis</td>
<td>Bordetella pseudomallei</td>
<td>122</td>
<td>60</td>
<td>98.914</td>
<td>2880</td>
<td>98.713</td>
<td>0.0451 ± 0.00124</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Shigella flexneri 301</td>
<td>Shigella flexneri 245T</td>
<td>77</td>
<td>58</td>
<td>99.693</td>
<td>3449</td>
<td>99.859</td>
<td>0.00047 ± 0.00006</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Shigella flexneri 301</td>
<td>Shigella sonnei</td>
<td>77</td>
<td>64</td>
<td>98.185</td>
<td>3034</td>
<td>98.361</td>
<td>0.03693 ± 0.0008</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Pseudomonas syringae tomatto</td>
<td>Pseudomonas syringae phaseolicola</td>
<td>43</td>
<td>21</td>
<td>86.549</td>
<td>3988</td>
<td>93.81</td>
<td>0.6279 ± 0.0104</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

\(^a\) Abbreviation: N/A, not applicable.
\(^b\) Number of genes interrupted by ISs.
\(^c\) Number of IS-interrupted ORFs with a functional homologue in a related strain.
\(^d\) Average DNA similarity between the IS-interrupted ORFs and functional homologues.
\(^e\) Number of homologous genes identified by reciprocal best-hit method.
\(^f\) Average DNA similarity of all homologues between the two genomes.
\(^g\) dS, synonymous substitution rate per synonymous site, estimated by maximum likelihood [42].
\(^w\) Estimates in Y. pestis are included for reference.

The work presented here shows that expansions of transposable elements in bacterial genomes can be reliable indicators of specialization to human-related niches. We anticipate that parallel genomic changes could be taking place in other human-related organisms such as domesticated fungal species involved in traditional food production or eukaryotic unicellular parasites. The remarkable speed at which bacteria have adapted to human cultural evolution underscores the evolutionary plasticity of the prokaryotic cell, which can...
probably be matched only by viral evolution [41]. The central role of IS elements described here and the use of their sequence divergence to detect other (not necessarily human-related) recent environmental specializations. Furthermore, we propose that analysis of the type of genes inactivated in the IS insertion process might give hints as to the nature of the environmental change. Although the pseudogenes generated by IS elements are quickly eroded by deleterional bias [8], their function could give useful information in recent specialization processes. In addition, older events would leave a trace of divergent ISs. The relatively new science of comparative genomics might provide a powerful tool to understand better the ecology and evolution of prokaroytes. This will be better elucidated when more genome sequences are available from non-human pathogens.

Acknowledgements
A.M. is a recipient of a ‘Ramón y Cajal’ research contract from the Spanish Ministry of Science and Education. Support from I.S. Carlos III (04/1319) and the EU project GEMINI (QLK3-CT-2002-02056) is also acknowledged. We thank S.B. Ingham for help with the figures and G. Nyiro and F. Kunst for providing information about host specificity in Wolbachia and Photobacterium species.

References
14 Parkhill, J. et al. (2001) Genome sequence of Yersinia pestis, the causative agent of plague. Nature 413, 523–527