**Appendix 1**

**Phylogenetic analysis among HGT event in *E. coli***

If an HGT event happened between a species of taxon A and the common ancestor of species B and B’ before the evolutionary splits of B and B’, we expect to observe two HGT events (e.g., between A and B and between A and B’). Indeed, we find that closely related strains often have exactly the same matches to distant families. As shown in [Figure 1—figure supplement 2](https://elifesciences.org/articles/62719/figures#fig1s2), pairs of *E. coli* strains with high average nucleotide identity (ANI) tend to share such matches. However, this effect is not very strong, probably because ANI does not accurately reflect the evolutionary distance between pairs of strains. To investigate this further, we retrieved the sequence of each match between *E. coli* strain and a species outside the *Escherichia* family. We then aligned all those matches (with blast) to all other *E. coli* strains and kept all alignments with an *E*-value <10−5

to assess, for each HGT event, its presence in each *E. coli* strain. The resulting matrix is shown in [Figure 1—figure supplement 3](https://elifesciences.org/articles/62719/figures#fig1s3). One can see that *E. coli* strains cluster to compact groups, possibly reflecting their phylogeny (see [Figure 1—figure supplement 3 (a)](https://elifesciences.org/articles/62719/figures#fig1s3)). The abundance of matches among *E. coli* strains exhibits a bi-modal distribution (see [Figure 1—figure supplement 3 (b)](https://elifesciences.org/articles/62719/figures#fig1s3)), possibly indicating the direction of HGT: matches that are abundant in *E. coli* have likely been transferred from *E. coli* to the distant family, whereas matches that are rare in *E. coli* were likely transferred from the distant family to *E. coli* . As shown in [Figure 1—figure supplement 4](https://elifesciences.org/articles/62719/figures#fig1s4), sharing a match to a different family (within a blast hit) is not strongly related to ANI distances; the association can be detected only statistically, as mentioned in [Figure 1—figure supplement 2](https://elifesciences.org/articles/62719/figures#fig1s2).

**Comparing bacterial and archaeal genomes**

It is known that bacteria and archaea have exchanged genetic material during their evolution ([Aravind et al., 1998](https://elifesciences.org/articles/62719" \l "bib4); [Nelson et al., 1999](https://elifesciences.org/articles/62719#bib57); [Garcia-Vallvé et al., 2000](https://elifesciences.org/articles/62719#bib33)). However, comparing all bacterial and archaeal RefSeq contigs longer than 106 bp, we find only several exact matches longer than 300 bp.

The longest one is of length 727 (see sequence 1 in [Supplementary file 8](https://elifesciences.org/articles/62719/figures#supp8)). This exact sequence exists in archaeon *Methanobacterium* sp. *MB1* and two bacteria: *Mahella australiensis 50–1 BON* and *Petrotoga mobilis SJ95*. The amino acid sequence of this match hits the (2Fe-2S)-binding protein, known to exist in both the bacterial and archaeal kingdoms.

*Mahella australiensis* gen. nov., sp. nov. (phylum: *Firmicutes*) is a moderately thermophilic anaerobic bacterium isolated from an Australian oil well ([Bonilla Salinas et al., 2004](https://elifesciences.org/articles/62719#bib8)). *Petrotoga mobilis* (phylum: *Thermotogae*) bacteria appear to be common members of the oil well microbial community. Petroleum reservoirs are a unique subsurface environment characterised by high temperatures, moderate to high salt concentrations, and abundant organic matter ([Hu et al., 2016](https://elifesciences.org/articles/62719#bib37)). *Methanobacterium* sp. *Mb1*, a hydrogenotrophic methanogenic archaeon, was isolated from a rural biogas plant producing methane-rich biogas from maize silage and cattle manure in Germany ([Maus et al., 2013](https://elifesciences.org/articles/62719" \l "bib53)).

We found a blast hit with more than 99% identity to this match in nine other species: *Pseudothermotoga elfii DSM 9442*, *Clostridium clariflavum DSM 19732*, *Fervidobacterium pennivorans strain DYC*, *Thermoanaerobacter* sp. *X513*, *Thermoanaerobacter* sp. *X514*, *Fervidobacterium pennivorans DSM 9078*, *Thermoanaerobacterium thermosaccharolyticum DSM 571*, and *Fervidobacterium islandicum strain AW-1*.

Just next to this match the same species share another match of length 496 (see sequence 2 in [Supplementary file 8](https://elifesciences.org/articles/62719/figures#supp8).fa). This match codes for signal peptidase II, also known to exist in both bacterial and archaeal kingdoms and, probably, plays some role in an antibiotic resistance ([Xiao and Wall, 2014](https://elifesciences.org/articles/62719#bib93)).

**Enrichment of gene functions in HGT**

Appendix 1—table 1

**Enrichment of different gene categories relative to the control set (see Gene enrichment analyses in Materials and methods).**

