Snapshots of a shrinking partner: Genome reduction in *Serratia symbiotica*

Alejandro Manzano-Marín¹,* and Amparo Latorre¹,²,*

¹Institut Cavanilles de Biodiversitat i Biologia Evolutiva - Universitat de València, Genètica Evolutiva, Paterna, 46980, Spain
²Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunitat Valenciana (FISABIO), Genómica y Salud, València, 46020, Spain
*alejandro.manzano@uv.es
*amparo.latorre@uv.es

**ABSTRACT**

Genome reduction is pervasive among maternally-inherited endosymbiotic organisms, from bacteriocyte- to gut-associated ones. This genome erosion is a step-wise process in which once free-living organisms evolve to become obligate associates, thereby losing non-essential or redundant genes/functions. *Serratia symbiotica* (Gammaproteobacteria), a secondary endosymbiont present in many aphids (Hemiptera: Aphididae), displays various characteristics that make it a good model organism for studying genome reduction. While some strains are of facultative nature, others have established co-obligate associations with their respective aphid host and its primary endosymbiont (*Buchnera*). Furthermore, the different strains hold genomes of contrasting sizes and features, and have strikingly disparate cell shapes, sizes, and tissue tropism. Finally, genomes from closely related free-living *Serratia marcescens* are also available. In this study, we describe in detail the genome reduction process (from free-living to reduced obligate endosymbiont) undergone by *S. symbiotica*, and relate it to the stages of integration to the symbiotic system the different strains find themselves in. We establish that the genome reduction patterns observed in *S. symbiotica* follow those from other dwindling genomes, thus proving to be a good model for the study of the genome reduction process within a single bacterial taxon evolving in a similar biological niche (aphid-*Buchnera*).

**Introduction**

Obligate microbial symbionts (weather primary, secondary, tertiary, or other) are present in a variety of eukaryotic organisms, such as leeches (Annelida: Hirudinida)¹, gutless oligochaetes (Annelida: Oligochaeta)², and insects (Arthropoda: Insecta)³ (see⁴). These have the capacity to produce essential nutrients their hosts cannot synthesise nor obtain from their diet⁵–⁸, making them essential for the correct development and survival of their partners. On the other hand, facultative symbionts are dispensable, although under certain environmental challenges/niches, they can endow the host with desirable traits, ranging from defence against parasitoids or fungal parasites to survival after heat stress (reviewed in⁹,¹⁰). Moreover, these facultative endosymbionts can even affect the performance of its host on a certain food source (e.g. a plant)¹¹–¹³. Whichever the symbiont’s function, taxonomic position, or status (facultative or obligate), a common feature from maternally inherited endosymbiotic organisms is the possession of a reduced genome, when compared to their free-living counterparts⁵,¹⁴–²². The sequencing of these genomes has undoubtedly provided important clues into the distinct features these display along the erosion process. While mildly-reduced genomes such as the one from *Sodalis glossinidius*, (facultative endosymbiont of the tsetse fly *Glossina morsitans morsitans*), *Sodalis prierantoniensis* (primary obligate endosymbiont from the rice weevil *Sitophilus oryzae*), and *Hamiltonella defensa* (facultative endosymbiont from the aphid *Acyrthosiphon pisum*) show intermediate guanine-cytosine (hereafter GC) contents and a massive presence of both mobile elements (hereafter MEs) and pseudogenes²³–²⁵, highly reduced genomes such as the ones from *Buchnera* (primary obligate endosymbiont of aphids) and *Blochmannia* (primary obligate endosymbiont from carpenter ants) are highly compact with few pseudogenes and show no traces of MEs²⁶,²⁷. However valuable the study of these genomes is, very few examples are available for different bacterial strains, belonging to a single genus, holding differentially-reduced genomes. These examples include *Arsenohpomon symbionts* of a parasitic wasp (*Nasonia vitripennis*) (INSDC:AUCC00000000.1), the brown planthopper (*Nilaparvata lugens*)²⁸, and louse flies (Diptera: Hippoboscidae)²⁹ (INSDC:CP013920.1); *Coxiella* symbionts of ticks³¹,³²; and *Sodalis* symbionts from the tsetse fly (and rice weevil)³³.  

*Serratia symbiotica*, a secondary endosymbiont harboured by many aphids³⁴–³⁷, is particular among currently sequenced bacterial symbionts. While strains harboured by *Aphis fabae* and *Acyrthosiphon pisum* (Aphidinae subfamily) are of facultative nature³⁰,³⁵,³⁶, strains from the aphids *Cinara tujafilina*, *Cinara cedri*, and *Tuberolachnus saliva* (Lachninae subfamily)
have established co-obligate associations with both the aphids and its primary obligate endosymbiont, Buchnera. This co-obligate association was putatively triggered by a loss of the riboflavin biosynthetic genes in Buchnera from the Lachninae last common ancestor. In addition, depending on the strain, its cell shape is either rod-like (strain CWBI-2.3 from Ap. fabae, hereafter SAF), filamentous (strain Tucson from Ac. pismum, hereafter SAP; and strain SCt-VLC from C. tujafilina, hereafter SCT), or spherical (strain SCc from Cinara cedri, hereafter SCc; and strain STs-Pazieg from T. salignus, hereafter STs). Not surprisingly, these strains hold genomes of contrasting sizes, ranging from 3.58 to 0.65 mega base pairs (hereafter Mbp).

In the current study, we analysed the genomes of currently-available S. symbiotica strains and compared them to the free-living insect pathogen S. marcescens strain Db11 (hereafter Db11). Through comparative genomics we investigated genome rearrangement, the enrichment, and loss, of MEs, and the erosion undergone by RNA features and the informational machinery in S. symbiotica. Additionally, we describe the diminution of certain genes and the possible functional consequences of these reductions. Finally, we relate all these features to different stages of the symbionts’ integration to the aphid-Buchnera symbiotic consortia and discuss the features which are convergent with other dwindling endosymbiotic genomes.

Results and Discussion

S. symbiotica strains and their shrinking genomes

Generally, “ancient” obligate endosymbionts hold highly reduced genomes as small as 112 kilo base pairs (hereafter kbp), conversely, more “recently” derived endosymbionts (including facultative ones) tend to display larger genomes, all the way up to the 4.5 Mbp genome of S. glossinidius (reviewed in ). Accordingly, the different genomes of S. symbiotica strains land within and along this spectrum, from the large 3.58 Mbp genome of the facultative SAF to the small 0.65 Mbp genome of the co-obligate STs (Figure 1). Similarly to the other large endosymbiotic genomes, SAF, SAP, and SCT’s display a large enrichment of MEs, both in terms of diversity and number of them. Interestingly, the composition of the insertion sequence (hereafter IS) families (the most common type of MEs found within these genomes) seems to be lineage-specific. While IS3

Figure 1. Genome reduction in S. symbiotica. Serratia genomes are depicted as circular plots and are arranged from largest (leftmost) to smallest (rightmost). From outermost to innermost, the rings within the genome plots display features on the direct strand, the reverse one, and RNA genes. Inside the circles, coloured lines connect the same-family IS elements scattered throughout the genome, following the colour code at the very bottom of the image. The grey bars on top of the genome plots describe the lifestyle and genome reduction stage. Underneath the genome plots, the strain alias and the host, between parenthesis, are shown. Below, a table showing the genomic features of each strain and pie charts displaying the relative abundance of IS-family elements, with the two most abundant highlighted by name. Underneath, the colour code for the different IS elements.
and IS256 are the most prevalent in SAf and SAp (both facultative endosymbionts from Aphidinae aphids), IS481 and IS5 are the most common in SCt (co-obligate endosymbiont from C. tujafilina [Lachninae]). Conversely, the smaller genomes of SCc and STs lack any traces of MEs, congruent with similar-sized endosymbiotic genomes (see\textsuperscript{3}). Following the trend of many other dwindling genomes\textsuperscript{3}, all S. symbiotica have a GC content lower than that of their free-living counterpart, Db11. While this GC content is very similar among SAf, SAp, and SCt (52.1%), there is a marked decrease in SCc (29.2%) and even more so in STs (20.9%). Additionally, while there is a great enrichment of pseudogenes in SAf (126), SAp (550), SCt (916), and SCc (110), the small STs is almost deprived of these gene remnants. This genetic erosion comes along with a decrease in coding density. Accordingly, while SAf shows only a small decrease when compared to Db11 (87.9% to 78.2%), SAp, SCt, and SCc exhibit a marked drop down (56.8%, 53.4%, and 39.0%, respectively), mainly due to the increased pseudogenisation and "junk" DNA. On the other hand, the highly-reduced STs shows a high coding density (77.5%). This difference between SCc and STs is mainly due to high-amount "junk" DNA that is present in SCc’s genome, amounting to almost half of it\textsuperscript{36}. Finally, we also found a gradual loss (from free-living Db11 to co-obligate intracellular STs) of RNA features (rRNAs, tRNAs, and other non-coding RNAs [hereafter ncRNAs]), revealing their different levels of genomic erosion. As has been previously observed in other endosymbionts, genome erosion comes with a "disturbance" of the functional profile of the organism, when compared to their free-living relatives\textsuperscript{14,18}. Accordingly, prior analyses have described that while the functional profiles of free-living Serratia strains were very stable, a displacement of it was evident in SCc and SAp\textsuperscript{46}. Through a similar analysis using all five currently available S. symbiotica strains, we have found that while the recently-derived SAf, SAp, and SCt strains cluster together forming a sister group to Db11, the highly-reduced SCc and STs form a divergent cluster from the rest of Serratia strains (Figure 2A). These two S. symbiotica clusters differ mainly in the relative presence of MEs (category X) and translation-related genes (category J). While the former reflects the enrichment of SAf, SAp, and SCt’s in MEs, the latter evidences the common trend in highly reduced endosymbionts to retain housekeeping genes (e.g. category J includes all ribosomal proteins) (see\textsuperscript{3}).

![Figure 2](https://example.com/figure2.png)

**Figure 2. S. symbiotica functional profile displacement and genome rearrangement.** (A). Heat map showing the two-way clustering of the S. symbiotica COG profile’s differences, relative to the free-living S. symbiotica. On the right, one-letter code for the COG categories. (B) Graphic linear representation of the rearrangements undergone in S. symbiotica. Orange and red vertical bars mark the position of single-copy conserved genes. Contigs that do not have single-copy conserved genes are not displayed. (C) Unrooted tree as calculated by MGR for the minimum number of rearrangements undergone by S. symbiotica.

In the early stages of an endosymbiont’s genomic reduction, the genome’s enrichment in MEs can lead to rearrangement\textsuperscript{18,35}. These rearrangements get fixed in the endosymbiotic lineage once the MEs have been lost, as is observed by the general genome-wide synteny displayed in Buchnera\textsuperscript{15,47}, Blochmannia\textsuperscript{29}, or Blattabacterium\textsuperscript{48}. Free-living Serratia strains display general genome-wide synteny,\textsuperscript{46} on the contrary, S. symbiotica genomes display various rearrangements when compared to free-living Db11’s, and even among each other’s (Figure 2B and C). Interestingly, while the less-reduced genome of SAf displays the closest relationship (in terms of rearrangements) to Db11, the drastically-reduced genome of STs has accumulated the highest number of rearrangements. Also, SCc and STs’ genomes, which both lack MEs, display no synteny between them. These observations suggest that all S. symbiotica lineages have diverged before the loss of MEs, allowing a great number of lineage-specific reorganisation.

**Erosion of essential amino acid biosynthetic routes**

A general feature of endosymbiotic genomes is the loss of non-essential genes, leading to highly reduced genomes with a genetic repertoire specialised in the symbiotic function (reviewed in\textsuperscript{3}). In aphids, Buchnera, the primary obligate endosymbiont, is mainly in charge of producing essential amino acids (hereafter EAs) for its host. Therefore, it is expected that co-existing symbionts show degraded biosynthetic routes involved in the production of these compounds. By analysing these routes in S.
Symbiotica, the gradual degradation of genes and operon attenuators implicated in the synthesis of EAAs becomes immediately evident (Figure 3). The recently-derived SAf shows intact routes for most EAAs, with the notable exceptions of lysine and methionine. As already described in a previous study, there is a marked difference in the retention of leucine, arginine, and histidine biosynthetic-related genes even between the closely related facultative SAp and co-obligate SCt. Finally, by comparing SCc and STs against the other S. symbiotica strains and each other, it becomes evident that both have become highly dependent on Buchnera for the supply of EAAs, with the main difference between SCc and STs being the purging of the remaining pseudogenes in the latter.

![Figure 3. Erosion of essential amino acid biosynthetic genes in S. symbiotica.](https://example.com/figure3)

**Figure 3. Erosion of essential amino acid biosynthetic genes in S. symbiotica.** Inactivation tables showing the genes and leader sequences involved in the essential amino acid biosynthetic routes in S. symbiotica genomes compared to free-living Db11. At the top and left of the table, gene names for each enzymatic step and abbreviation for each Serratia strain, respectively. At the bottom of the table, black lines encompass the enzymatic steps required for the biosynthesis of each compound. Amino acid names are displayed using standard three-letter abbreviations. At the bottom-right, colour code for squares. Half-coloured boxes mean the genes catalysing the enzymatic step are present in different states.

### Decay of RNA features and the loss of regulation

Typically, highly-reduced endosymbionts retain only a small number of ncRNAs and other RNA features (see Supplementary Figs. S1 and S2). Through an annotation of these in the genomes of S. symbiotica and Db11, we have explored the erosion of RNA features (Figure 4: top panel). We found that in the recently-derived endosymbionts SAf, SAp, and SCt, many of these features are still retained, although differentially. This points towards drift acting behind the loss of these features at the early stages of genome reduction. As expected, these three genomes show the acquisition of ME related ncRNAs, which all belong to the large class of self-catalytic group II introns (RF00029, RF01999, RF02001, RF02003, RF02005, RF02012). In the intermediate and advanced stages of drastic genome reduction SCc and STs find themselves in, respectively, most of the RNA features have been lost. Conserved features across S. symbiotica are the 4.5S RNA component of the signal recognition particle (SRP) (ffs), the RNase P M1 RNA component (rnpB), the tmRNA (ssrA), the tpek11 small RNA (of unknown function), the leader sequence from the rnc-era transcription unit (coding for the ribonuclease 3 and the GTPase Era), and the alpha operon leader (coding for the 30S ribosomal subunits S13, S11, and S4; the 50S ribosomal subunit L17; and the DNA-directed RNA polymerase subunit alpha). The first three are interestingly also retained in other small genomes, but unidentifiable in some tiny genomes (Supplementary Fig. S1), hinting at these being essential functions retained until the last stages of genome reduction. Since most of these RNA features are related to the regulation of gene expression (small antisense RNAs, riboswitches, and leader sequences [including amino acid operon attenuators]), these losses would reflect a general trend of gene-regulation-loss in endosymbiotic genomes through the erosion of RNA features.

Regarding tRNAs, we observed a drastic reduction in tRNA-gene number, particularly marked in SCc and STs (Figure 4: bottom panel). These losses, as in other reduced endosymbionts (see Supplementary Fig. S2), mainly affect redundancy rather...
Figure 4. Decay of tRNA and other RNA features in *Serratia* genomes. (Top) Colour-coded diagram showing the decay of RNA features in different *S. symbiotica* genomes. On top of the matrix, gene names (for the first four columns) and RNA categories (for the rest) are indicated. On the bottom of the matrix, feature names are indicated for those features retained in SCc and STs. (Bottom) Colour-coded diagram showing the decay of tRNA features in different *S. symbiotica* genomes. On the top of the matrix, aminoacyl charging potential for each tRNA species (as inferred by TFAM). Each column represents a different anticodon. Asterisks indicate putative codon reassignments as judged by TFAM. fMet= N-Formylmethionine, kIle= lysylated isoleucine.

Contrasting the other *S. symbiotica* genomes, we were unable to detect a tRNA with aminoacyl charging potential for glutamate in SCc. This is similar to what is observed in other tiny genomes, where some tRNAs with certain aminoacyl charging potential are absent (Supplementary Fig. S2). However, the presence of a tRNA\(^{\text{Glu}}\) in a yet-identified plasmid cannot be discarded. Also, a loss of the selenocysteine tRNA is already present in the early co-obligate SCt, consistent with the loss of other selenocysteine-related genes, and completely absent in the smaller SCc and STs. It is important to remark that, in SAp, one of the tRNA\(^{\text{Met}}\) copies has undergone a mutation in its anticodon (C\(\text{AT} \rightarrow \text{AAT}\)), which could theoretically lead to the ATT codon to be recognised as coding for methionine. Finally, the tRNAs for formyl-methionine (tRNA\(^{\text{fMet}}\)), in charge of aminoacylation of the starting methionine, and lysylated isoleucine (tRNA\(^{\text{kIle}}\)) are conserved even in the two smallest *S. symbiotica*. This follows the trend observed in other reduced genomes (Supplementary Fig. S2) and points towards the essential nature of these tRNAs.

### Informational machinery

By analysing and comparing the informational machinery (ribosome-, transcription-, translation-, and DNA replication/repair-related genes) in *S. symbiotica* strains, both high preservation as well as gradual patterns of deterioration become evident in different categories. The ribosome, as well as the tRNA aminoacylation genes are mostly perfectly preserved (Figure 5: top). Marked differences include the presence of multiple copies of the three rRNA genes in SAf, SAp, and SCt, and the absence of two ribosomal proteins (*rpsI* and *rplM*) as well as the prolyl-tRNA synthetase gene (*proS*) in STs. While the retention of only one copy of the rRNA genes reflects the tendency of endosymbiotic, and other reduced genomes, to eliminate redundancy, the loss of the *rpsI* gene (coding for the 30S ribosomal subunit S9) reflects the loss of a non-essential gene. In *Escherichia coli*, it has been experimentally proven that a null mutant of the *rplI* gene is able to grow, albeit showing a slow growth phenotype\(^{49,51}\). Most intriguing are the losses of the *rplM* (coding for the 50S ribosomal subunit L13) and *proS* genes. The former has been described as essential in *E. coli*\(^{51}\), and its loss could be related to the loss of *rpsI*, that together with *rplM* forms an operon. The latter, could reflect a putative functional replacement of the ProS protein activity by another non-specific aminoacyl–tRNA synthetase. This phenomenon has been observed for the prolyl–tRNA synthetase of *Deinococcus radiodurans*, which has the ability to charge cysteine to tRNA\(^{\text{Cys}}\). This non-specific aminoacyl–tRNA synthetases have also been observed in archael organisms (reviewed in\(^{53}\)), suggesting this to be a common mechanism to cope with the lack of a specific aminoacyl–tRNA synthetase.

---

5/15
Figure 5. Informational machinery in *S. symbiotica* genomes. Circular plots displaying the different genes involved in functional categories (top left, ribosome; top right, tRNA aminoacylation; middle left, rRNA modifications; middle right, tRNA modification; bottom left, DNA replication and repair; bottom right, transcription and translation) in the informational machinery of *S. symbiotica* strains and Db11. Outer lines in each circular plot delimit the subcategory. Form outer to inner, the rings in the plot stand for the gene name, the colour-coded lines delimiting categories/complexes, and boxes standing for the presence or absence of the genes in Db11, SAf, SAP, SCt, SCc and STs.
Both rRNAs and tRNAs undergo a series of modifications that are required to produce the mature version of these ncRNAs (reviewed in 54, 55). By analysing the genes involved in both rRNA and tRNA modifications, we observed that while the recently-derived SAf, SAP, and SCI hold a rather complete set (with particularly marked losses of 23S rRNA methyltransferases), the highly-reduced SCC and STs retain only a small fraction of these genes (Figure 5: middle). With the notable exceptions of the fnt, tilS, trmD, tsaB, tsaC, tsaD, and tada genes (all retained in the small SCC and STs), individual knockout mutants of all the rRNA and tRNA modification-related genes in E. coli (except miaE, which is not present in this organism) have dimmed them non-essential56-59. The fnt and tilS genes code for the proteins responsible for the attachment of a formyl group to the free amino group of methionyl-tRNA^Met (for initiator methionine)60 and the modification of the wobble base of the CAU anticodon of the tRNA^{k1l61,62}, respectively. The retention of these two genes thereby insures both the correct charging of the initiator methionine in proteins (which is posttranslationally-removed) and the accurate recognition of the AUA codon as coding for Isoleucine. On the other hand, the genes tsaB, tsaC, and tsaD (along with the tsaE gene) are responsible for the biosynthesis of the threonylcarbamoyladenosine (t6A) residue at position 37 of ANN-decoding tRNAs.63 Interestingly, the tsaE gene (an ATPase), which has been found to be non-essential in E. coli under anaerobic conditions64, is missing from STs, thus the biosynthesis of t6A would be either putatively impaired or working in an unknown way. Finally, the tada gene, which codes for a tRNA-specific adenosine deaminase that is essential for viability in E. coli65, is retained even in the small STs.

In regards to DNA replication and repair, the gene losses are particularly marked in the most genomically reduced symbionts, SCC and STs, affecting mostly DNA repair-related genes (Figure 5: bottom left). This is also observed in other reduced endosymbionts (see43), and is possibly related to the triggering of more drastic genome erosion (reviewed in3). DNA replication-related losses affect the non-essential holE gene of the DNA polymerase, the priA-dependent primosome (retaining an elementary DNA-dependent one [missing the auxiliary Hup proteins]), and the gyrA subunit of the DNA gyrase. These latter, although identified as essential in E. coli57, has also been found to be missing from tiny genomes43, thereby suggesting its function could be taken over by an alternative enzyme or it actually being non-essential in some endosymbiotic organisms.

In terms of transcription- and translation-related genes, a high-degree of retention in all S. symbiotica genomes can be observed (Figure 5: bottom right). Gene losses mainly affect the sigma factors, with STs retaining only the rpoD and rpoH genes, coding for 70 and 32, respectively. While the former is generally preserved in endosymbionts43, the latter is missing from endosymbionts such as Blattabacterium and Nasuia. 32 is required for the normal expression of heat shock genes and for the heat shock response through the regulation of the synthesis of heat shock proteins66, and thus its retention/loss could be specific of certain endosymbiotic systems.

Dwindling genes: stripping proteins down to the bones

Through the manual curation of the annotation of SCI, SCC, and STs endosymbionts35,37, we noted that some genes (atpC, cysJ, deaD, dnaX, ftsN, hscA, metG, pcnB, rnr, and tolC) seemed to be shorter in STs, and sometimes consistently shrunken across S. symbiotica, compared to those of free-living E. coli and even Db11. However, while these genes showed truncated or missing domains, they displayed a high degree of sequence conservation, when compared to Db11. Thorough examination of these shrunken genes revealed that experimental evidence, mainly form E. coli, have proven that truncated versions of these proteins were able to function with few to none obvious phenotypic consequences (details recorded in the annotation files available from the INSDC). Particularly evident is the loss of non-essential domains in six proteins: AceF, DnaX, FtsK, FtsN, and Rnr (Figure 6). The AceF protein (E2 component of pyruvate dehydrogenase complex) has undergone the loss of one or two biotin/lipoyl domains (PF00364) in all S. symbiotica genomes, namely STs retains only one. In E. coli, it has been shown, through the in vitro deletion of biotin/lipoyl domains, that one single domain suffices with respect to enzyme activity and protein function67. The tau subunit of the DNA polymerase III is coded by the dnaX gene, however an alternative isoform, denominated gamma subunit, is produced due to a programmed ribosomal frameshifting, which leads to a premature stop codon in the -1 frame at codon 43068. In vitro experiments with the shorter isoform, which lacks the tau 4 and 5 domains (PF12168 and PF12170), indicate that gamma is to be sufficient for replication69. The most drastic gene diminutions are observed in the ftsK and ftsN genes (whose products are involved in cell division), where SCC and STs preserve only a very small portion of the original gene. Independent in vivo experiments in E. coli mutants coding only for truncated FtsK (amino acids 1-200)70 or FtsN (amino acids 1-119)71 proteins, have corroborated that these tiny versions are sufficient for cell division, although short to long filamentous cells were observed to occur. Regarding MetG (Methionyl tRNA–synthetase), both SCC and STs are lacking the C-terminal putative tRNA binding domain (PF01588). Genetic complementation studies and characterization of C-terminally truncated enzymes in E. coli, established that MetG can be reduced to 547 residues without significant effect on either the activity or stability of the enzyme72. Finally, a deletion of the C-terminal basic domain of the Rnr protein (ribonuclease R) can be observed only in STs. This could lead to an increase in activity of this enzyme, since assays using purified truncated Rnr proteins from mutant E. coli, lacking the 83 residues from the C-terminus, were shown to display higher affinity and circa 2-fold higher activity than full length wild-type Rnr (on poly[A], A[17] and A[4] substrates)73. Through the alignment of the aforementioned putatively-functional proteins against other small and tiny genomes, we corroborated most of these gene
diminutions are common among these organisms (Supplementary Data S1). This suggests that selection might favour gene diminution (the retention of only essential domains of a protein), relaxing selective constraints in non-essential gene regions, thus further contributing to genome reduction.

**Figure 6. Diminution of genes in S. symbiotica.** Graphic representation of the proteins that show evident loss of non-essential domains (as judged by experimental evidence in *E. coli*) in *S. symbiotica*. Domains in each frame are represented by coloured boxes, with similar colours used for repeated domains in each protein. on top of each box, the domain's is provided. Above each frame, the protein's name is stated. On the bottom-right of every frame, a scale bar is provided.

**Conclusion**

*S. symbiotica* strains analysed here have all been evolving under a similar environment, the aphid-*Buchnera* symbiotic system. We have established that *S. symbiotica* strains can be considered to be along the genome reduction spectrum from a free-living bacterium to a drastically-reduced endosymbiont, thus providing "snapshots" of the genome reduction process. SAF would thus represent the very first stages of genome reduction, having not yet lost its ability to be grown in axenic culture and having undergone a mild genome shrinkage and few rearrangements, when compared to the free-living Db11. SAP and SCT would be a stage further down the path, having a more reduced genome than SAF and showing a massive enrichment in both pseudogenes and MEs. However, SCT has already done the transition to becoming a co-obligate endosymbiont, and thus shows more drastic gene losses in the EAAs’ biosynthetic pathways. SCC and STS find themselves in more advanced stages of genome reduction and integration to their symbiotic systems, having established a series of metabolic dependencies and complementation with *Buchnera* for the synthesis of several essential compounds. Nonetheless, SCC differs greatly from STS in genome size, which is explained by the former being in a recent stage of an advanced genome erosion, thus retaining several pseudogenes and "junk"
DNA. Both SCc and STs display a drastic genome-wide gene loss, and particularly in their ncRNA repertoire and informational machinery. Through the comparison of these S. symbiota strains, we were able to hint at essential retained functions, which not surprisingly are shared with other highly-reduced endosymbionts. The detailed study of protein diminution in S. symbiota revealed a common tendency of endosymbionts to loose non-essential protein domains, and thus constituting an additional route towards genome reduction. We expect the further study of this particular endosymbiont of aphids will continue to provide important clues into the intriguing process of genome reduction.

Methods

Annotation of protein-coding genes
All protein-coding genes that were not found in their respective S. symbiota genomes were searched for using the online version of tblastn74 with S. marcescens’ or E. coli’s protein as query. All positive hits with an e-value \( \leq 10^{-3} \) were then manually curated. Domains within a protein were annotated using the InterProScan75 webserver and through alignments against E. coli’s proteins using MAFFT v7.22076. Circular representations of presence/absence of genes were done using circos v0.6777 and edited in Inkscape v0.91. COG categories for proteins were assigned using blastx and ad hoc perl scripts to select the best non-overlapping hits with an e-value threshold of \( \leq 10^{-3} \). Then, COG categories absolute counts were converted to relative ones per organism. Finally, in order to analyse the disturbance of S. symbiota’s functional profiles from that of free-living Db11, we subtracted Db11’s relative frequency per COG from each value within the same row. Visual display of COG categories was done using R and the gplots library, followed by manual editing in Inkscape.

Annotation of RNA features
tRNA features were annotated using tRNAscan-SE v1.3.178 (-B option for the bacterial model) and TFAM v1.4, followed by manual curation. All other RNA features were searched for using Infernal v1.1.79 (--cut_tc --mid) against the Rfam v12.0 database.80 All hits with an e-value lower than \( \leq 10^{-3} \) were considered and manually curated. Visual displays were done using R and the gplots library, followed by manual editing in Inkscape. Plain-text source files used for the plotting of RNA features can be found in https://figshare.com/s/193f7513676a83b52241.

Rearrangement analysis
Single copy shared proteins among S. symbiota strains and Db11 were calculated as in37. Briefly, we used OrthoMCL v2.0.981 to build the orthologous groups of proteins, followed by manual curation aimed at joining rapidly-evolving proteins such as outer membrane proteins. These proteins were then used as rearrangement markers for calculating a minimal rearrangement phylogeny in MGR v2.0.82 (no heuristics). Scaffold for unfinished genomes (SAf, SAP, and SCl) were arranged as to minimise the distance against Db11. Tree visualisation was done in FigTree v1.4.1. Rearrangement graphic was done in R using the genoPlotR library.83 All graphics were edited in Inkscape.

References
1. Perkins, S. L., Budinoff, R. B. & Siddall, M. E. New Gammaproteobacteria associated with blood-feeding leeches and a broad phylogenetic analysis of leech endosymbionts. Applied and environmental microbiology 71, 5219–5224 (2005). URL http://aem.asm.org/content/71/9/5219.short.
2. Giere, O. & Langheld, C. Structural organisation, transfer and biological fate of endosymbiotic bacteria in gutless oligochaetes. Marine Biology 93, 641–650 (1987). URL http://link.springer.com/10.1007/BF00392801.
3. Moran, N. A., McCutcheon, J. P. & Nakabachi, A. Genomics and Evolution of Heritable Bacterial Symbionts. Annual Review of Genetics 42, 165–190 (2008). URL http://www.annualreviews.org/doi/abs/10.1146/annurev.genet.41.110306.130119.
4. Buchner, P. Endosymbiose der Tiere mit Pflanzlichen Mikroorganismen (Birkhäuser Basel, Basel, 1953). URL http://link.springer.com/10.1007/978-3-0348-6958-4.
5. Nogge, G. Significance of symbionts for the maintenance of an optimal nutritional state for succesful reproduction in haematophagous arthropods. Parasitology. Workshop Proceedings of the Third European Multicolloquium of Parasitology (1980) 82, 101–104 (1981).
6. Douglas, A. E. Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria Buchnera. Annu Rev Entomol 43, 17–37 (1998). URL http://www.annualreviews.org/doi/abs/10.1146/annurev.ento.43.1.17.

9/15
7. Woyke, T. et al. Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* **443**, 950–5 (2006). URL http://www.nature.com/doifinder/10.1038/nature05192.

8. Manzano-Marin, A., Oceguera-Figueroa, A., Latorre, A., Jiménez-Garcia, L. F. & Moya, A. Solving a Bloody Mess: B-Vitamin Independent Metabolic Convergence among Gammaproteobacterial Obligate Endosymbionts from Blood-Feeding Arthropods and the Leech *Haementeria officinalis*. *Genome Biology and Evolution* **7**, 2871–2884 (2015). URL http://gbe.oxfordjournals.org/lookup/doi/10.1093/gbe/evv188.

9. Oliver, K. M., Degnan, P. H., Burke, G. R. & Moran, N. A. Facultative Symbionts in Aphids and the Horizontal Transfer of Ecologically Important Traits. *Annual Review of Entomology* **55**, 247–266 (2010). URL http://www.annualreviews.org/doi/abs/10.1146/annurev-ento-112408-085305.

10. Oliver, K. M., Smith, A. H. & Russell, J. A. Defensive symbiosis in the real world - advancing ecological studies of heritable, protective bacteria in aphids and beyond. *Functional Ecology* **28**, 341–355 (2014). URL http://doi.wiley.com/10.1111/1365-2435.12133.

11. Leonardo, T. E. & Muiru, G. T. Facultative symbionts are associated with host plant specialization in pea aphid populations. *Proceedings. Biological sciences / The Royal Society* **270 Suppl**, S209–S212 (2003). URL http://rspb.royalsocietypublishing.org/cgi/doi/10.1098/rsbl.2003.0064.

12. Tsuchida, T., Koga, R. & Fukatsu, T. Host Plant Specialization Governed by Facultative Symbiont. *Science* **303**, 1989–1989 (2004). URL http://www.sciencemag.org/cgi/doi/10.1126/science.1094611.

13. Ferrari, J., Scarborough, C. L. & Godfray, H. C. J. Genetic Variation in the Effect of a Facultative Symbiont on Host-Plant Use by Pea Aphids. *Oecologia* **153**, 323–329 (2007). URL http://www.jstor.org/stable/40210868.

14. Gil, R. et al. The genome sequence of *Blochmannia floridanus*: Comparative analysis of reduced genomes. *Proceedings of the National Academy of Sciences* **100**, 9388–9393 (2003). URL http://www.pnas.org/cgi/doi/10.1073/pnas.1533499100.

15. van Ham, R. C. H. J. et al. Reductive genome evolution in *Buchnera aphidicola*. *Proceedings of the National Academy of Sciences* **100**, 581–586 (2003). URL http://www.pnas.org/cgi/doi/10.1073/pnas.0235981100.

16. Burke, G. R. & Moran, N. A. Massive Genomic Decay in Serratia symbiotica, a Recently Evolved Symbiont of Aphids. *Genome Biology and Evolution* **3**, 195–208 (2011). URL http://gbe.oxfordjournals.org/cgi/doi/10.1093/gbe/evr002.

17. Nikoh, N., Hosokawa, T., Oshima, K., Hattori, M. & Fukatsu, T. Reductive Evolution of Bacterial Genome in Insect Gut Environment. *Genome Biology and Evolution* **3**, 702–714 (2011). URL http://gbe.oxfordjournals.org/cgi/doi/10.1093/gbe/evr064.

18. Clayton, A. L. et al. A novel human-infection-derived bacterium provides insights into the evolutionary origins of mutualistic insect-bacterial symbioses. *PLoS genetics* **8**, e1002990 (2012). URL http://dx.plos.org/10.1371/journal.pgen.1002990.

19. Nakabachi, A. et al. Defensive Bacteriome Symbiont with a Drastically Reduced Genome. *Current biology* **23**, 1478–1484 (2013). URL http://dx.doi.org/10.1016/j.cub.2013.06.027.

20. Foray, V. et al. Whole-Genome Sequence of *Serratia symbiotica* Strain CWBI-2.3T, a Free-Living Symbiont of the Black Bean Aphid *Aphis fabae*. *Genome Announcements* **2**, e00767–14 (2014). URL http://genomea.asm.org/cgi/doi/10.1128/genomeA.00767-14.

21. Gottlieb, Y., Lalzar, I. & Klasson, L. Distinctive Genome Reduction Rates Revealed by Genomic Analyses of Two *Coxiella*-Like Endosymbionts in Ticks. *Genome biology and evolution* **7**, 1779–1796 (2015). URL http://gbe.oxfordjournals.org/lookup/doi/10.1093/gbe/evv108.

22. Smith, T. A., Driscoll, T., Gillespie, J. J. & Raghavan, R. A *Coxiella*-Like Endosymbiont Is a Potential Vitamin Source for the Lone Star Tick. *Genome Biology and Evolution* **7**, 831–838 (2015). URL http://gbe.oxfordjournals.org/cgi/doi/10.1093/gbe/evv016.

23. Toh, H. et al. Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. *Genome Research* **16**, 149–156 (2005). URL http://www.genome.org/cgi/doi/10.1101/gr.4106106.

24. Oakeson, K. F. et al. Genome degeneration and adaptation in a nascent stage of symbiosis. *Genome Biology and Evolution* **6**, 76–93 (2014). URL http://gbe.oxfordjournals.org/cgi/doi/10.1093/gbe/evt210.
25. Degnan, P. H., Yu, Y., Sisneros, N., Wing, R. A. & Moran, N. A. *Hamiltonella defensa*, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. *Proceedings of the National Academy of Sciences* **106**, 9063–9068 (2009). URL http://www.pnas.org/cgi/doi/10.1073/pnas.0900194106.

26. Williams, L. E. & Wernegreen, J. J. Genome evolution in an ancient bacteria-ant symbiosis: parallel gene loss among *Blochmannia* spanning the origin of the ant tribe Camponotini. *PeerJ* **3**, e881 (2015). URL https://peerj.com/articles/881.

27. Xue, J. *et al.* Genomes of the rice pest brown planthopper and its endosymbionts reveal complex complementary contributions for host adaptation. *Genome Biology* **15**, 521 (2014). URL http://genomebiology.com/2014/15/12/521.

28. Nováková, E., Husnák, F., Šochová, E. & Hypša, V. *Arsenophonus* and *Sodalis* Symbionts in Louse Flies; an Analogy to the *Wigglesworthia* and *Sodalis* System in Tsetse Flies. *Applied and Environmental Microbiology* **81**, 6189–6199 (2015). URL http://aem.asm.org/lookup/doi/10.1128/AEM.01487-15.

29. Chen, D.-Q. & Purcell, A. H. Occurrence and Transmission of Facultative Endosymbionts in Aphids. *Current Microbiology* **34**, 220–225 (1997). URL http://link.springer.com/10.1007/s002849900172.

30. Russell, J. A., Latorre, A., Sabater-Munoz, B., Moya, A. & Moran, N. A. Side-stepping secondary symbionts: widespread horizontal transfer across and beyond the Aphidoidea. *Molecular Ecology* **12**, 1061–1075 (2003). URL http://doi.wiley.com/10.1046/j.1365-294X.2003.01780.x.

31. Lamelas, A. *et al.* Evolution of the Secondary Symbiont “*Candidatus Serratia symbiotica*” in Aphid Species of the Subfamily Lachninae. *Applied and Environmental Microbiology* **74**, 4236–4240 (2008). URL http://aem.asm.org/cgi/doi/10.1128/AEM.01487-15.

32. Burke, G. R., Normark, B. B., Favret, C. & Moran, N. A. Evolution and diversity of facultative symbionts from the aphid subfamily Lachninae. *Applied and Environmental Microbiology* **75**, 5328–5335 (2009). URL http://aem.asm.org/cgi/doi/10.1128/AEM.00717-09.

33. Chandler, S. M., Wilkinson, T. L. & Douglas, A. E. Impact of plant nutrients on the relationship between a herbivorous insect and its symbiotic bacteria. *Proceedings. Biological sciences / The Royal Society* **275**, 565–70 (2008). URL http://rspb.royalsocietypublishing.org/content/275/1634/565.

34. Chen, D.-Q., Montllor, C. B. & Purcell, A. H. Fitness effects of two facultative endosymbiotic bacteria on the pea aphid, *Acyrthosiphon pisum*, and the blue alfalfa aphid, *A. kondoi*. *Entomologia Experimentalis et Applicata* **95**, 315–323 (2000). URL http://doi.wiley.com/10.1046/j.1570-7458.2000.00670.x.

35. Manzano-Marín, A. & Latorre, A. Settling Down: The Genome of *Serratia symbiotica* from the Aphid *Cinara tujafilina* Zooms in on the Process of Accommodation to a Cooperative Intracellular Life. *Genome Biology and Evolution* **6**, 1683–1698 (2014). URL http://gbe.oxfordjournals.org/cgi/doi/10.1093/gbe/evu133.

36. Lamelas, A. *et al.* *Serratia symbiotica* from the Aphid *Cinara cedri*: A Missing Link from Facultative to Obligate Insect Endosymbiont. *PLoS Genetics* **7**, e1002357 (2011). URL http://dx.plos.org/10.1371/journal.pgen.1002357.

37. Manzano-Marín, A., Simon, J.-C. & Latorre, A. Reinventing the wheel and making it round again: Evolutionary convergence in *Buchnera*-Serratia symbiotic consortia between the distantly related Lachninae aphids *Tuberolachnus salignus* and *Cinara cedri*. *Genome Biology and Evolution* evw085 (2016). URL http://gbe.oxfordjournals.org/lookup/doi/10.1093/gbe/evw085.

38. Sabri, A. *et al.* Isolation, pure culture and characterization of *Serratia symbiotica* sp. nov., the R-type of secondary endosymbiont of the black bean aphid *Aphis fabae*. *International Journal of Systematic and Evolutionary Microbiology* **61**, 2081–2088 (2011). URL http://ijs.microbiologyresearch.org/content/journal/ijs.10.1099/ijs.0.024133-0.

39. Moran, N. A., Russell, J. A., Koga, R. & Fukatsu, T. Evolutionary Relationships of Three New Species of Enterobacteriaceae Living as Symbionts of Aphids and Other Insects. *Applied and Environmental Microbiology* **71**, 3302–3310 (2005). URL http://aem.asm.org/cgi/doi/10.1128/AEM.71.6.3302-3310.2005.

40. Kikuchi, Y. Endosymbiotic Bacteria in Insects: Their Diversity and Culturability. *Microbes and Environments* **24**, 195–204 (2009). URL http://jol.jic.jst.go.jp/JST.JSTAGE/jsme2/ME09140S?f=from=CrossRef.

41. Flyg, C., Kenne, K. & Boman, H. G. Insect Pathogenic Properties of *Serratia marcescens*: Phage-resistant Mutants with a Decreased Resistance to Cecropia Immunity and a Decreased Virulence to *Drosophila*. *Microbiology* **120**, 11/15.
173–181 (1980). URL http://mic.microbiologyresearch.org/content/journal/micro/10.1099/00221287-120-1-173.

42. Bennett, G. M. & Moran, N. A. Small, smaller, smallest: The origins and evolution of ancient dual symbioses in a phloem-feeding insect. * Genome Biology and Evolution* 5, 1675–1688 (2013). URL http://gbe.oxfordjournals.org/cgi/doi/10.1093/gbe/evt118.

43. Moran, N. A. & Bennett, G. M. The Tiniest Tiny Genomes. * Annual Review of Microbiology* 68, 195–215 (2014). URL http://www.annualreviews.org/doi/abs/10.1146/annurev-micro-091213-112901.

44. Degnan, P. H. *et al.* Dynamics of genome evolution in facultative symbionts of aphids. *Environmental Microbiology* 12, 2060–2069 (2009). URL http://doi.wiley.com/10.1111/j.1462-2920.2009.02085.x.

45. Hansen, A. K., Vorburger, C. & Moran, N. A. Genomic basis of endosymbiont-confferred protection against an insect parasitoid. * Genome Research* 22, 106–114 (2012). URL http://genome.cshlp.org/cgi/doi/10.1101/gr.125351.111.

46. Manzano-Marín, A., Lamelas, A., Moya, A. & Latorre, A. Comparative Genomics of *Serratia* spp.: Two Paths towards Endosymbiotic Life. * PLoS ONE* 7, e47274 (2012). URL http://dx.plos.org/10.1371/journal.pone.0047274.

47. Tamas, I. *et al.* 50 Million Years of Genomic Stasis in Endosymbiotic Bacteria. *Science* 296, 2376–2379 (2002). URL http://scitation.aip.org/content/aip/journal/physrevlett/89/1/10.1103/PhysRevLett.89.028102.

48. Patino-Navarrete, R., Moya, A., Latorre, A. & Peretó, J. Comparative Genomics of *Blattabacterium cuenoti*: The Frozen Legacy of an Ancient Endosymbiotic Genome. * Genome Biology and Evolution* 5, 351–361 (2013). URL http://gbe.oxfordjournals.org/cgi/doi/10.1093/gbe/evt011.

49. Mendonça, A. G., Alves, R. J. & Pereira-Leal, J. B. Loss of Genetic Redundancy in Reductive Genome Evolution. * PLoS Computational Biology* 7, e1001082 (2011). URL http://dx.plos.org/10.1371/journal.pcbi.1001082.

50. Bubunenko, M., Baker, T. & Court, D. L. Essentiality of Ribosomal and Transcription Antitermination Proteins Analyzed by Systematic Gene Replacement in *Escherichia coli*. * Journal of Bacteriology* 189, 2844–2853 (2007). URL http://jb.asm.org/cgi/doi/10.1128/JB.01713-06.

51. Shoji, S., Dambacher, C. M., Shajani, Z., Williamson, J. R. & Schultz, P. G. Systematic Chromosomal Deletion of Bacterial Ribosomal Protein Genes. * Journal of Molecular Biology* 413, 751–761 (2011). URL http://linkinghub.elsevier.com/retrieve/pii/S00221287-120-1-173.

52. Zhang, C.-M. & Hou, Y.-M. Synthesis of Cysteinyl-tRNA-Cys by a Prolyl-tRNA Synthetase. * RNA Biology* 1, 34–40 (2014). URL http://www.tandfonline.com/doi/abs/10.4161/rna.1.1.910.

53. Jacquin-Becker, C. *et al.* Cysteinyl-tRNA formation and prolyl-tRNA synthetase. *FEBS Letters* 514, 34–36 (2002). URL http://doi.wiley.com/10.1016/S0014-5793(02)029023-1.

54. Decatur, W. A. & Fournier, M. J. RNA modifications and ribosome function. *Trends in Biochemical Sciences* 27, 344–351 (2002). URL http://linkinghub.elsevier.com/retrieve/pii/S096800402021096.

55. Hagervall, T. G. & Björk, G. R. Transfer RNA Modification. * EcoSal Plus* 1 (2013). URL http://www.asmscience.org/content/journal/ecosalplus/10.1128/ecosalplus.4.6.2.

56. Gerdes, S. Y. *et al.* Experimental Determination and System Level Analysis of Essential Genes in *Escherichia coli* MG1655. * Journal of Bacteriology* 185, 5673–5684 (2003). URL http://jb.asm.org/cgi/doi/10.1128/JB.185.19.5673-5684.2003.

57. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. * Molecular Systems Biology* 2, 2006.0008 (2006). URL http://msb.embopress.org/cgi/doi/10.1038/msb4100050.

58. Joyce, A. R. *et al.* Experimental and Computational Assessment of Conditionally Essential Genes in *Escherichia coli*. *Journal of Bacteriology* 188, 8259–8271 (2006). URL http://jb.asm.org/cgi/doi/10.1128/JB.00740-06.

59. Purta, E., Kaminska, K. H., Kasprzak, J. M., Bujnicki, J. M. & Douthwaite, S. YbeA is the m3 methyltransferase RlmH that targets nucleotide 1915 in 23S rRNA. *RNA* 14, 2234–2244 (2008). URL http://www.rnajournal.org/cgi/doi/10.1261/rna.0198108.

60. Kahn, D., Fromant, M., Fayat, G., Dessen, P. & Blanquet, S. Methionyl-Transfer-RNA Transformylase from *Escherichia coli*. Purification and Characterisation. *European Journal of Biochemistry* 105, 489–497 (1980). URL http://doi.wiley.com/10.1111/j.1432-1033.1980.tb04524.x.
61. Muramatsu, T. et al. Codon and amino-acid specificities of a transfer RNA are both converted by a single post-transcriptional modification. *Nature* **336**, 79–81 (1988). URL http://www.nature.com/doifinder/10.1038/33679a0.

62. Grosjean, H. & Björk, G. R. Enzymatic conversion of cytidine to lysidine in anticodon of bacterial tRNAlle – an alternative way of RNA editing. *Trends in Biochemical Sciences* **29**, 165–168 (2004). URL http://linkinghub.elsevier.com/retrieve/pii/S0968004004000532.

63. Deutsch, C., El Yacoubi, B., de Crécy-Lagard, V. & Iwata-Reuyl, D. Biosynthesis of Threonylcarbamoyl Adenosine (t 6 A), a Universal tRNA Nucleoside. *Journal of Biological Chemistry* **287**, 13666–13673 (2012). URL http://www.jbc.org/lookup/doi/10.1074/jbc.M112.344028.

64. Mangat, C. S. & Brown, E. D. Known Bioactive Small Molecules Probe the Function of a Widely Conserved but Enigmatic Bacterial ATPase, YjeE. *Chemistry & Biology* **15**, 1287–1295 (2008). URL http://linkinghub.elsevier.com/retrieve/pii/S1074552108004109.

65. Wolf, J., Gerber, A. P. & Keller, W. tadA, an essential RNA-specific adenosine deaminase from *Escherichia coli*. *The EMBO journal* **21**, 3841–51 (2002). URL http://emboj.embopress.org/cgi/doi/10.1093/emboj/cdf362.

66. Grossman, A. D., Straus, D. B., Walter, W. A. & Gross, C. A. Sigma 32 synthesis can regulate the synthesis of heat shock proteins in *Escherichia coli*. *Genes & Development* **1**, 179–184 (1987). URL http://www.genesdev.org/cgi/doi/10.1101/gad.1.2.179.

67. Guest, J. R., Lewis, H. M., Graham, L. D., Packman, L. C. & Perham, R. N. Genetic reconstruction and functional analysis of the repeating lipoyl domains in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*. *Journal of Molecular Biology* **185**, 743–754 (1985). URL http://linkinghub.elsevier.com/retrieve/pii/0022836585900592.

68. Tsuchihashi, Z. & Kornberg, A. Translational frameshifting generates the gamma subunit of DNA polymerase III holoenzyme. *Proceedings of the National Academy of Sciences* **87**, 2516–2520 (1990). URL http://www.pnas.org/cgi/doi/10.1073/pnas.87.7.2516.

69. Walker, J. R. et al. *Escherichia coli* DNA polymerase III tau- and gamma-subunit conserved residues required for activity in vivo and in vitro. *Journal of bacteriology* **182**, 6106–13 (2000). URL http://jb.asm.org/cgi/doi/10.1128/JB.182.21.6106-6113.2000.

70. Draper, G. C., McLennan, N., Begg, K., Masters, M. & Donachie, W. D. Only the N-terminal domain of FtsK functions in cell division. *Journal of bacteriology* **180**, 4621–7 (1998). URL http://jb.asm.org/content/180/17/4621.abstract.

71. Gerding, M. A. et al. Self-Enhanced Accumulation of FtsN at Division Sites and Roles for Other Proteins with a SPOR Domain (DamX, DedD, and RlpA) in *Escherichia coli* Cell Constriction. *Journal of Bacteriology* **191**, 7383–7401 (2009). URL http://jb.asm.org/cgi/doi/10.1128/JB.00811-09.

72. Mellot, P., Mechulam, Y., Le Corre, D., Blanquet, S. & Fayat, G. Identification of an amino acid region supporting specific methionyl-tRNA synthetase: tRNA recognition. *Journal of Molecular Biology* **208**, 429–443 (1989). URL http://linkinghub.elsevier.com/retrieve/pii/00228365990057X.

73. Vincent, H. A. & Deutscher, M. P. The Roles of Individual Domains of RNase R in Substrate Binding and Exoribonuclease Activity: The Nuclease Domain is Sufficient for Digestion of Structured RNA. *Journal of Biological Chemistry* **284**, 486–494 (2009). URL http://www.jbc.org/cgi/doi/10.1074/jbc.M806468200.

74. Johnson, M. et al. NCBI BLAST: a better web interface. *Nucleic Acids Research* **36**, W5–W9 (2008). URL http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gkn201.

75. Jones, P. et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014). URL http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/btu031.

76. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**, 772–780 (2013). URL http://mbe.oxfordjournals.org/cgi/doi/10.1093/molbev/mst010.

77. Krzywinski, M. et al. Circos: an information aesthetic for comparative genomics. *Genome research* **19**, 1639–45 (2009). URL http://genome.cshlp.org/content/early/2009/06/15/gr.092759.109.abstract.

78. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. *Nucleic Acids Research* **25**, 955–964 (1997). URL http://nar.oxfordjournals.org/lookup/doi/10.1093/nar.25.5.955.
79. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013). URL http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/btt509.

80. Nawrocki, E. P. et al. Rfam 12.0: updates to the RNA families database. *Nucleic Acids Research* **43**, D130–D137 (2015). URL http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gku1063.

81. Li, L. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Research* **13**, 2178–2189 (2003). URL http://www.genome.org/cgi/doi/10.1101/gr.1224503.

82. Bourque, G. & Pevzner, P. A. Genome-scale evolution: Reconstructing gene orders in the ancestral species. *Genome Research* **12**, 26–36 (2002). URL http://genome.cshlp.org/content/12/1/26.abstract.

83. Guy, L., Roat Kultima, J. & Andersson, S. G. E. genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* **26**, 2334–2335 (2010). URL http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/btq413.

Acknowledgements

This work has been funded by the Ministerio de Economía y Competitividad (Spain) co-financed by FEDER funds [BFU2015-64322-C2-1-R to A.L.]; the European Commission [Marie Curie FP7 PITN-GA-2010-264774-SYMBIOMICS to A.M.M.]; and the Consejo Nacional de Ciencia y Tecnología (Mexico) [Doctoral scholarship CONACYT 327211/381508 to A.M.M.]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions statement

A.M.M. and A.L. conceived and designed the study. A.M.M. analysed the data. A.M.M. and A.L. wrote, reviewed, and approved the final version of the manuscript.

Additional information

Accession codes

Accessions for all organisms used in this study.

| Species                  | Strain     | Genome size (Mbp) | Accession       |
|--------------------------|------------|-------------------|-----------------|
| *Serratia marcescens*    | Db11       | 5.11              | HG326223        |
| *Serratia symbiotica*    | CWBI-2.3   | 3.58              | GCA_000821185.1 |
| *Serratia symbiotica*    | Tucson     | 2.57              | GCA_000186485.2 |
| *Serratia symbiotica*    | SCt-VLC    | 2.49              | GCA_900002265.1 |
| *Serratia symbiotica*    | SCc        | 1.76              | CP002295        |
| *Serratia symbiotica*    | STs-Pazieg | 0.65              | LN890288        |
| *Baumannia cicadellinicola* | BGSS      | 0.76              | CP008985        |
| *Buchnera aphidicola*    | BCC        | 0.42              | CP000263        |
| *Nasua deltocephalina*   | PUNC       | 0.11              | CP013211        |
| *Sulcia muelleri*        | GWSS       | 0.25              | CP000770        |
| *Tremblaya phenacola*    | PAVE       | 0.17              | CP003982        |
| *Wigglesworthia glossinidia* | brevipalpis | 0.70              | BA000021        |
| *Zinderia insecticola*   | CARI       | 0.21              | CP002161        |
Competing financial interests
The authors declare no competing financial interests.