Report

Current Biology

# Parallel Evolution in the Integration of a Co-obligate Aphid Symbiosis

## Highlights

- Aphids have independently evolved dependence on *Serratia symbiotica* at least 4 times
- The integration of the new co-obligate symbiont proceeds in a predictable manner
- Loss of the riboflavin and peptidoglycan pathways in *Buchnera* leads to co-dependence
- Amino acid synthesis is taken over by *Serratia* in a second phase of complementarity

## Authors

David Monnin, Raphaella Jackson, E. Toby Kiers, Marie Bunker, Jacintha Ellers, Lee M. Henry

## Correspondence

l.henry@qmul.ac.uk

## In Brief

Dependence on multiple nutrient-provisioning symbionts has evolved numerous times in insects. Monnin et al. provide evidence from the symbionts of aphids that these dependencies evolve in a predictable manner. The repeated losses of the same metabolic pathways bind the symbionts into co-dependence, and integration follows in a stepwise manner.

Monnin et al., 2020, Current Biology *30*, 1–9
May 18, 2020 
https://doi.org/10.1016/j.cub.2020.03.011

CellPress

Please cite this article in press as: Monnin et al., Parallel Evolution in the Integration of a Co-obligate Aphid Symbiosis, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.03.011

Current Biology
Report



# Parallel Evolution in the Integration of a Co-obligate Aphid Symbiosis

David Monnin,[1] Raphaella Jackson,[1] E. Toby Kiers,[2] Marie Bunker,[3] Jacintha Ellers,[2] and Lee M. Henry[1,4,*]
[1]School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK
[2]Department of Ecological Science, Vrije Universiteit Amsterdam, Amsterdam 1081 HV, the Netherlands
[3]Biology Department, University of Puget Sound, 1500 N. Warner St., Tacoma, WA 98416, USA
[4]Lead Contact
*Correspondence: l.henry@qmul.ac.uk
https://doi.org/10.1016/j.cub.2020.03.011

## SUMX

**Insects evolve dependence—often extreme—on microbes for nutrition. This includes cases in which insects harbor multiple endosymbionts that function collectively as a metabolic unit [1–5]. How do these dependences originate [6], and is there a predictable sequence of events leading to the integration of new symbionts? While co-obligate symbioses, in which hosts rely on multiple nutrient-provisioning symbionts, have evolved numerous times across sap-feeding insects, there is only one known case in aphids, involving *Buchnera aphidicola* and *Serratia symbiotica* in the Lachninae subfamily [7–9]. Here, we identify three additional independent transitions to the same co-obligate symbiosis in different aphids. Comparing recent and ancient associations allow us to investigate intermediate stages of metabolic and anatomical integration of *Serratia*. We find that these uniquely replicated evolutionary events support the idea that co-obligate associations initiate in a predictable manner—through parallel evolutionary processes. Specifically, we show how the repeated losses of the riboflavin and peptidoglycan pathways in *Buchnera* lead to dependence on *Serratia.* We then provide evidence of a stepwise process of symbiont integration, whereby dependence evolves first. Then, essential amino acid pathways are lost (at ~30–60 mya), which coincides with the increased anatomical integration of the companion symbiont. Finally, we demonstrate that dependence can evolve ahead of specialized structures (e.g., bacteriocytes), and in one case with no direct nutritional basis. More generally, our results suggest the energetic costs of synthesizing nutrients may provide a unified explanation for the sequence of gene losses that occur during the evolution of co-obligate symbiosis.**

## RESULTS

### Independent Transitions to Co-obligate Symbiosis

Sap-feeding insects have provided elegant case studies of the evolution of co-obligate symbioses, whereby organisms harbor multiple endosymbionts that function collectively as a metabolic unit. These include species of mealybugs that depend on endosymbionts, which in turn harbor their own endosymbionts, and cicadas, in which one symbiont has fragmented into distinct but interdependent lineages [1–5]. What processes drive multiple microbial species to join into co-obligate symbioses [6], and, more generally, is there a predictable, deterministic sequence of events leading to the genomic and anatomical integration of new symbionts?

The aphids are an ideal lineage to study early-stage co-obligate symbioses. The majority of aphid species harbor a single obligate symbiont, *Buchnera aphidicola*, and a second non-obligate symbiont *Serratia symbiotica* (hereafter referred to as *Buchnera* and *Serratia,* respectively). While *Serratia* is found at intermediate frequencies in numerous aphid species, the symbiont has transitioned to a co-obligate relationship with *Buchnera* in the Lachninae subfamily [7–9]. Such co-obligate functioning is marked by *Buchnera*'s losing metabolic capabilities, namely the ability to synthesize the essential nutrients riboflavin and, in some species, tryptophan [10]. Our aim was to determine whether (1) other cases of obligate co-dependences have arisxen across the aphids, and (2) to ask whether these transitions followed predictable genomic, metabolic, and anatomical trajectories. Such patterns can provide insight into the evolutionary processes that have led to the genome structure of more ancient multi-partner symbioses [6].

Using data on the symbiont prevalence in 131 aphid species from [11], we identified species that carry *Serratia* at a high frequency, and then tested aphid populations in both the United Kingdom and the Netherlands for obligate dependence on the symbiont. We defined species as having evolved obligate reliance on *Serratia* if (1) all individuals within populations carry the symbiont and (2) they experience a significant fitness reduction when the symbiont is removed. Symbionts that do not meet these criteria are referred to as "facultative," as they are not essential for host survival. We screened for the presence of *Serratia* using PCR and measured dependence by "curing" individual aphids with antibiotics that selectively removed *Serratia* without affecting *Buchnera*, and then counted their total offspring to determine the lifetime fecundity of the aphids in the presence and absence of the symbiont.

We identified ubiquitous *Serratia* symbioses in seven aphid species (Table S1), representing three independent co-obligate transitions in *Microlophium carnosum*, *Aphis urticata*, and in the *Periphyllus* genus. In the *Periphyllus* genus, *Serratia* was consistently present in the five species we surveyed, and we

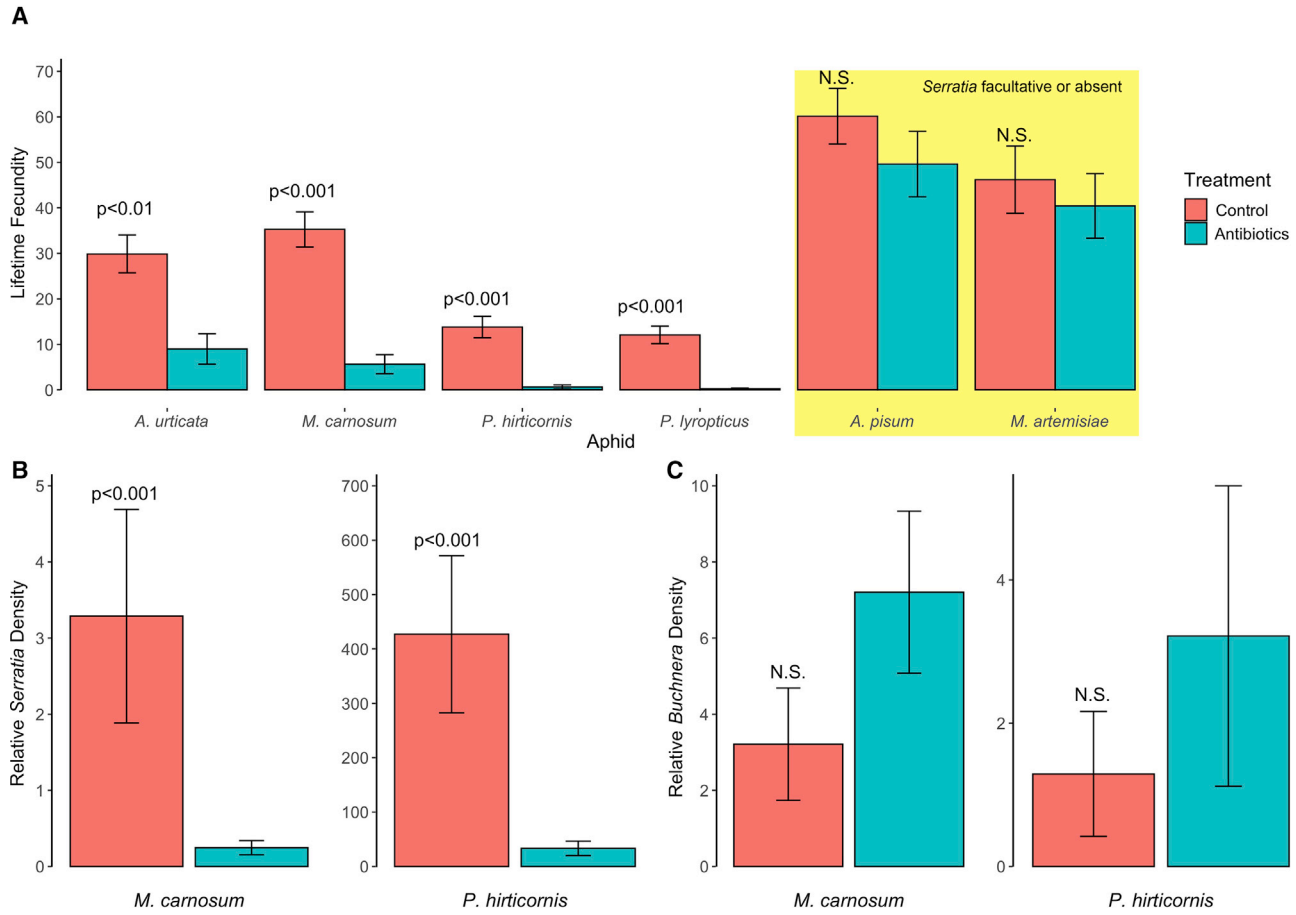


**Figure 1. Antibiotic Curing of *Serratia* in Aphids**

(A) Effect of antibiotic curing on aphid lifetime fecundity (mean ± standard error) in aphid species representing three independent evolutionary transitions to co-obligate dependence (*A. urticata*, *M. carnosum*, and *P. hirticornis* and *P. lyropictus* representing the *Periphyllus* genus). The yellow box highlights species that do not host *Serratia* as a co-obligate symbiont. Results synthesized from two independent trials on aphids from the United Kingdom and the Netherlands. See also Figure S1and Data S1A.

(B and C) Effect of antibiotic curing on *Serratia* density (B) and *Buchnera* density (C) compared to the density of host cells (mean ± standard error). See also Data S1B. "N.S." indicates no significant difference between treatments, "p<0.01" and "p<0.001" indicate a significant difference between treatments.

confirmed obligate dependence via curing *Serratia* in both *P. hirticornis* and *P. lyropictus.* These data suggest a single transition into an obligate relationship with *Serratia* at the origins of the *Periphyllus* genus (see below). Curing had the most dramatic effect in species of the *Periphyllus* genus, potentially reflecting a longer-term evolutionary association with *Serratia* (Figure 1A). We confirmed that the antibiotic treatments had no significant effect on the fecundity of our control aphid species *Acyrthosiphon pisum*, which harbors *Serratia* as a facultative symbiont, and the uninfected *Macrosiphoniella artemisiae* (Figures 1A and S1; Data S1A). Likewise, we confirmed with qPCR that the antibiotic treatment reduced *Serratia* density (Figure 1B; Data S1B), but did not reduce *Buchnera* density (Figure 1C; Data S1B).

We next estimated the origins of obligate dependence on *Serratia* using deep coverage 16S amplicon sequencing from our field-collected populations and previous data on *Serratia* associations in aphids [11]. First, we found evidence of a more ancient relationship between *Serratia* and aphids in the *Periphyllus* genus: amplicon sequencing confirmed that *Serratia* was absent from *Chaitophorus* aphids, a sister lineage to the *Periphyllus* clade (Data S1C). This suggests that dependence on *Serratia* originated at the divergence of these 2 genera an estimated 63–79 mya (Figures 2 and S2). Second, we found evidence of more recent origins of *Serratia* obligate dependence (<30 mya) in *A. urticata* and *M. carnosum.* Specifically, *Serratia* was either absent or present only as a facultative infection in some individuals in the species of *A. idaei* and *A. fabae* related to *A. urticata*. Lack of obligate dependence was likewise confirmed in *Macrosiphum euphorbiae* and *A. pisum*, related to *M. carnosum*.

## Genomic Basis of *Buchnera-Serratia* Metabolic Complementarity

We then asked whether evolving obligate dependence on *Serratia* was associated with a consistent genomic signature in aphids, and more specifically whether *Buchnera-Serratia* metabolic complementarity originates in a predictable manner across host lineages. We obtained whole-genome sequencing data for *M. carnosum*, *A. urticatasymbiont; aphid; co-obligate;Serratia*

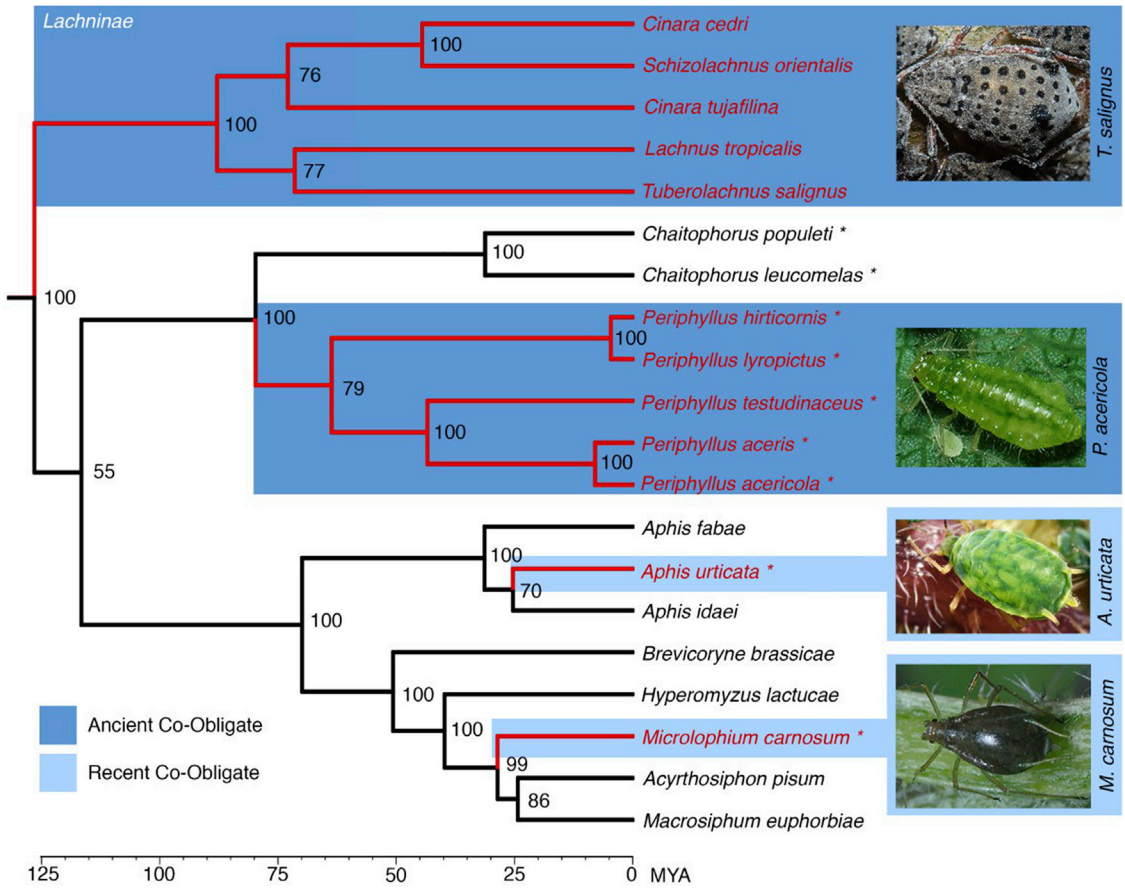


**Figure 2. A Dated Aphid Phylogeny Showing Lineages with Co-obligate Symbioses**

Branch colors represent aphid lineages that have evolved obligate dependence on *Serratia* (red) or those that carry the symbiont facultatively (black). Light blue shading represents more recent co-obligate associations, and dark blue represents ancient associations for the entire clade of host species. Asterisks denote data from this study (see Table S1). Other data are compiled from several studies [11–13]. Numbers on the nodes of the tree are bootstrap values. Node ages (in millions of years) and confidence intervals are reported in Figure S2. See Figure S4 and Data S1F for a phylogeny of the *Serratia symbiotica* strains. Photos from https://influentialpoints.com/.

*symbiotica;Buchnera aphidicola; evolution of dependence; metabolic complementation*, and three *Periphyllus* species. We then used previously published data from the Lachninae subfamily (*Cinara cedri*, *C. tujafilina* and *Tuberolachnus salignus*) to compare the gene losses in *Buchnera* from the four independent transitions into an obligate relationship with *Serratia*. This included the three new cases of co-obligate dependence identified here, and the previously identified cases in the Lachninae subfamily. Our analysis centered on the pathways and genes involved in essential nutrient provisioning to the host (Data S1D). Specifically, we focused on pathways that have experimental evidence for being essential for the aphid: riboflavin [14] and essential amino acids [15–21]. Of particular interest was the riboflavin pathway in *Buchnera*, as the loss of this pathway has been hypothesized to trigger the dependence on *Serratia* in the Lachninae aphids [10].

We found a consistent signature for the loss of the riboflavin pathway of *Buchnera* in both *M. carnosum* and aphids in the *Periphyllus* genus (Figure 3). In *M. carnosum*, *Buchnera* is missing one gene, part of the ribD complex, which is essential to the riboflavin pathway. In the *Periphyllus* genus, by contrast, the full pathway is missing, as it is in the Lachninae subfamily. Previous work in the Lachninae aphids suggests that *Buchnera* has also lost the capacity to synthesize the amino acid tryptophan in certain species (e.g., *C. cedri* and *T. salignus*). We find similar losses in the *Periphyllus* lineage. Here, the majority of genes in the tryptophan pathway have either been lost or pseudogenized, and orthologous gene copies, which are predicted to serve the same function, have been retained in the *Serratia* genome (Figure 3). Conversely, the tryptophan pathway has been retained in the *Buchnera* genomes of both *M. carnosum* and *A. urticata,* the more recent co-obligate relationships. This result suggests an advanced stage of functional losses in *Buchnera* of *Periphyllus* aphids, further supported by losses in several additional amino acid pathways that also appear to have been taken over by *Serratia*.

In contrast, *Buchnera* has retained the complete pathways to synthesize all of these essential nutrients in *A. urticata*. This is surprising, given the consistency of gene losses in *Buchnera* of *M. carnosum*, aphids in the *Periphyllus* genus, and the Lachninae aphids, all which are co-obligately dependent on *Serratia*. Compared to aphid lineages in which *Buchnera* is the sole

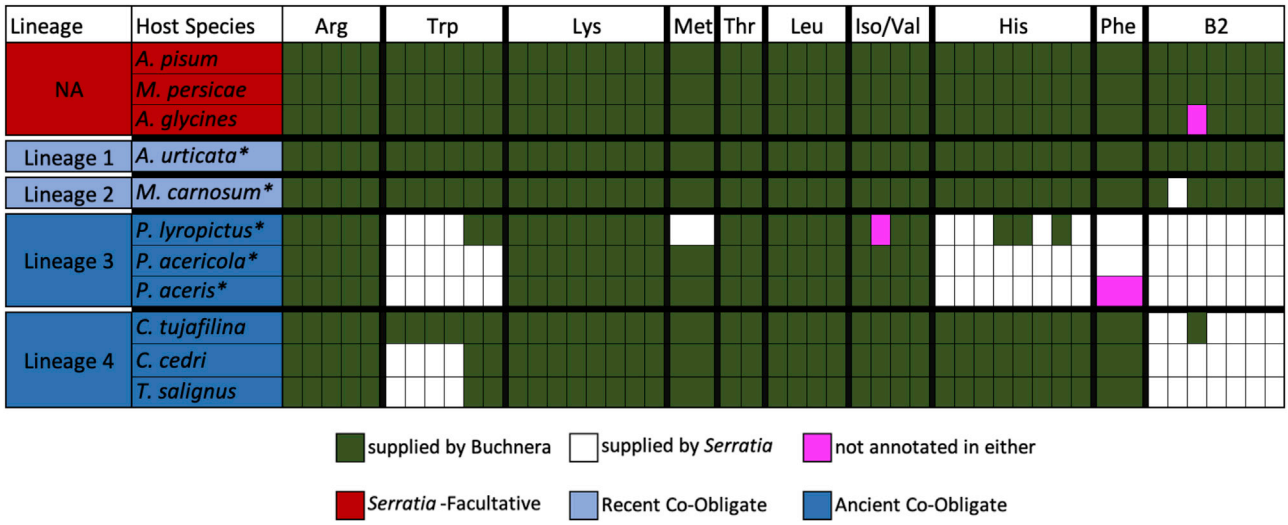


**Figure 3. Essential Nutrients-Provisioning Genes of *Buchnera* and *Serratia* Genomes**

Aphid lineages that have evolved dependence on *Serratia* are numbered. Coding capabilities of symbionts for essential amino acids and riboflavin are presented for each aphid species. Each cell represents a single gene with color-coding depicting whether the gene is present in the genome of *Buchnera* (green) or *Serratia* (white). The gene content of *Buchnera* from *A. pisum*, *M. persicae*, and *A. glycines* is included as a comparison to species in which *Buchnera* is the sole obligate symbiont. Gene names and additional information, including gene redundancies, can be found in Data S1D. Asterisks denote genomic data from this study. See Table S2 for a list of genes that were absent from *Buchnera* in *A. urticata*. See Table S3 for information on the metabolic cost of essential amino acid synthesis. See also Figure S3.

obligate symbiont (*A. pisum, Myzus persicae*, and *Aphis glycines*), the co-obligate association of *Serratia* and *Buchnera* in *A. urticata* has only six genes missing in *Buchnera,* in which there are orthologous gene copies in *Serratia.* None of the six genes has direct links to essential nutrient pathways (see Table S2 for more detail). This suggests that co-obligate dependence can arise in this system through alternative starting points, including non-nutritional pathways.

General genomic features of the different *Buchnera* strains likewise support the hypothesis that the co-obligate *Serratia* symbioses found in Lachninae and *Periphyllus* aphids are more ancient compared to *M. carnosum* and *A. urticata* (Figure S3). Both the genome size and GC content of *Buchnera* are highly reduced in the Lachninae and *Periphyllus* clades, which is suggestive of a more advanced degree of degradation. Gene redundancies are also indicative of the age of the co-obligate associations. In *M. carnosum* and *A. urticata*, the genomes of *Serratia* and *Buchnera* still contain a significant number of the same genes involved in synthesizing nutrients that are essential for the host aphid (72.5% and 39.2%, respectively). Conversely, in the *Periphyllus* lineages, both *P. acericola* and *P. aceris* have only 11.5% gene redundancy between the two symbionts. In *P. lyropictus,* there is a 47.1% overlap. The higher redundancy in *P. lyropictus* is likely due to *Serratia* being recently replaced by another *Serratia* strain within this aphid lineage.

As *Buchnera-Serratia* complementarity could also arise through pathways that are not essential for host nutrition, we investigated genes involved in additional pathways involved in translation, as it underlies essential functions in all bacteria [22] and peptidoglycan synthesis, which has been shown to be important in other symbiotic systems [23]. We also investigate the pathways to synthesize the precursors chorismate, homoserine, and vitamins $B_5$ and $B_9$, and lipoic acid. The additional B vitamins showed similar patterns of metabolic patchwork to riboflavin. However, genes to complete the pathways are either absent in both symbiont genomes ($B_5$ in *A. urticata* and the Lachninae) or are missing where *Buchnera* is the sole symbionts ($B_9$ in *A. pisum*), so it is unclear whether these vitamins are essential for the symbiosis. Genes involved in translation, chorismate, homoserine, and lipoic acid were for the most part conserved in all of the *Buchnera* genomes (Data S1E). The pathway for peptidoglycan, however, was entirely lost in the ancient co-obligate *Buchnera* (Lachninae and *Periphyllus* aphids). In contrast, only one or two genes were missing in the most recent lineages of *M. carnosum* and *A. urticata*. In all cases, the missing genes to synthesize peptidoglycan in *Buchnera* were retained on the *Serratia* genomes. This suggests that *Serratia* may be co-obligate because of its contributing role to the peptidoglycan synthesis of *Buchnera*, in addition to providing nutrients to the host. The only gene consistently missing in the peptidoglycan pathway from all co-obligate *Buchnera*, including *M. carnosum* and *A. urticata*, was *murF*. However, it is unclear whether this gene is essential for *Buchnera* as it is pseudogenized in *A. glycines*, which is a species not known to host any obligate symbiont other than *Buchnera*.

## Anatomical Integration in Co-obligate Symbiosis

Lastly, we studied the abundance and localization of symbionts within their host to look for anatomical signatures of co-obligate symbiosis. Likewise, we expected that a greater degree of metabolic reliance on *Serratia* in the *Periphyllus* aphids would correspond with greater anatomical integration—for example, through the formation of a specialized organ (bacteriome) to house *Serratia*. To test this idea, we performed fluorescent *in*

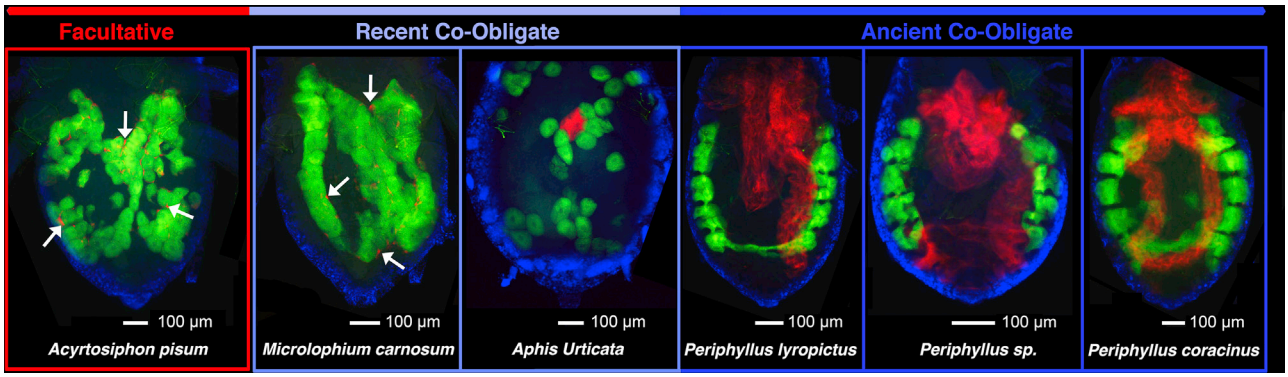


**Figure 4. FISH Images of Abdomens from Six Aphid Species**

*Buchnera* and *Serratia* are highlighted in green and red, respectively. The colored bar represents the degree of reliance on *Serratia*: facultative (red) or co-obligate (blue). The shades of blue represent the degree of anatomical integration with the host, from the least (*M. carnosum*) to the greatest (*Periphyllus* spp.). White arrows indicate the regions where *Serratia* is localized.

*situ* hybridization (FISH) using probes specifically targeting *Buchnera* and *Serratia*. As predicted, we found increased anatomical integration of *Serratia* in host lineages corresponding to a greater reliance on *Serratia* (Figure 4). In the most extreme case, we found that the *Periphyllus* aphids evolved a large organ (bacteriome) containing numerous bacteriocytes to house *Serratia* in their abdomens (Figure 4).

In contrast to high anatomical integration in the *Periphyllus* aphids, we found that both *M. carnosum* and *A. urticata* exhibit minimal integration of *Serratia*. In *A. urticata*, *Serratia* is localized in a small cluster of relatively large cells (~4), forming a small bacteriome surrounded by *Buchnera*-containing bacteriocytes. In *M. carnosum*, *Serratia* is the least integrated, with the symbiont being localized in sheath cells surrounding the *Buchnera*-containing bacteriocytes. This pattern is similar to the one found in *A. pisum* where *Serratia* maintains a consistently facultative relationship with its host.

We likewise expected *Serratia* abundance within the aphid to increase as the symbiont takes on a more metabolically demanding role. Here, we used qPCR to quantify the copies of *Serratia* genomes compared to the host aphid. In line with our predictions, we found a substantial increase in the abundance of *Serratia* that coincided with its greater metabolic role of synthesizing amino acids. Specifically, we found that the ratio of *Serratia* to host genome copies increased dramatically to an abundance ratio of 512:1 in *P. hirticornis*. This is compared to a 3:1 ratio found in the less integrated co-obligate of *M. carnosum* (Figure 1B).

## DISCUSSION

Dependence on multiple co-obligate symbionts has originated numerous times in the evolution of eukaryotes. However, the vast majority of co-obligate symbioses are ancient. Data on recent associations are needed to reveal the evolutionary processes that initiate dependence and provide insight into the intermediate steps leading to the extreme genomic and anatomical integration observed in ancient associations. By comparing ancient and recent associations in aphids, we find strong evidence that the mechanisms initially binding symbiotic partners in obligate relationships occur in a deterministic, predictable manner. Specifically, we find that dependence on *Serratia* originates through parallel evolutionary trajectories marked by repeated losses of the same nutrient pathways in *Buchnera* across multiple host lineages. Our genomic and FISH data show stepwise processes of symbiont integration, with the losses of essential amino acid pathways occurring between 30 and 60 million years after the co-obligate symbiosis evolves. This is followed by a second phase of dependence characterized by greater anatomical integration of *Serratia* in the more ancient obligate partnerships of the *Periphyllus* genus compared to the more recently adopted co-obligate associations of *A. urticata* and *M. carnosum*.

Our results provide the first evidence that *Buchnera* has repeatedly lost the capacity to produce the essential nutrient riboflavin in multiple aphid lineages. In each case in which the pathway to synthesize riboflavin has been lost, *Serratia* has retained genes to compensate for these metabolic changes in *Buchnera*. Studies have shown that an aphid's demand for riboflavin is relatively low compared to other nutrients, such as amino acids [14, 24, 25]. This may explain why riboflavin is lost first as the modest host demand for this vitamin may be easily met by a new symbiont even at a relatively low abundance. In several species within the Lachninae sub-family, the tryptophan pathway is also missing. This suggests that once the co-obligate symbiosis with *Serratia* is established, the loss of amino acid pathways in *Buchnera* can follow [10]. Our work confirms that the tryptophan pathway has likewise been lost in *Periphyllus* aphids and that the capacity to synthesize this amino acid is vulnerable to deletion. Tryptophan can be one of the most costly essential amino acids to synthetize (see Table S3) [26]; it has been hypothesized that these costs may explain why its loss is associated with the presence of a second obligate symbiont [6]. The second most energetically expensive amino acid synthesis pathways (phenylalanine, histidine, methionine, and isoleucine/valine) are likewise lost in the *Buchnera* of the *Periphyllus* aphid. Our finding is in line with work documenting the loss of energetically expensive amino acids and complementation by a companion symbiont in several ancient co-obligate symbioses, including *Sulcia* of some Auchenorrhyncha families (e.g., spittlebugs, cicadas)

and in *Carsonella*, the primary symbiont of Psyllids. In *Sulcia*, the amino acid pathways appear to have been lost 60 million years after co-dependence evolved, before the common ancestor of cicadas, sharpshooters, and spittlebugs [27].

We find amino acid pathways are only lost in *Buchnera* in the more ancient co-obligate associations of the *Periphyllus* and in some Lachninae species. In contrast, *Buchnera* has retained these functions in the more recent co-obligates of *A. urticata* or *M. carnosum*. This suggests that essential amino acids are only susceptible to deletions in the second phases of losses. This can happen once selection has been relaxed by the presence of a new obligate symbiont, 30–60 million years after dependence evolves. These results also provide strong support for the hypothesis that the energetic costs of synthesizing nutrients may provide a unified explanation for the sequence of gene losses that occur during the evolution of co-obligate symbiosis (Table S3).

In the case of *A. urticata*, we find that dependence on *Serratia* has evolved with no direct link to nutrient provisioning. Our results suggest that in this case, *Buchnera* may have become dependent on *Serratia* as a result of losing the capacity to synthesize peptidoglycan (Data S1E). Peptidoglycan is an important protective compound that is involved in bacterial cell division, shape, and integrity. The pathway to produce this polymer has also been shown to underlie the complementarity between symbiotic partners in other systems, such as in mealybugs [23]. In aphids, we find that the genes to synthesize peptidoglycan form a metabolic patchwork similar to those responsible for producing essential nutrients; the entire pathway is missing from the *Buchnera* of ancient co-obligates of Lachninae and *Periphyllus*, whereas only a single gene, *murF*, is missing in both of the more recent associations. This suggests that the sharing of peptidoglycan biosynthesis may represent an important first step in the formation of co-obligate symbioses. However, the peptidoglycan pathway is incomplete in *Buchnera* of certain aphid species (e.g., *A. glycines*), which may reflect compensatory adaptation on the part of the host [28]; therefore, its role in aphid symbioses requires further investigation.

More generally, our data support the hypothesis that lineages that recently acquired co-obligate symbionts will have cases of overlapping gene complexes. This is seen in both *A. urticata* and *M. carnosum*, in which *Serratia* shares many redundant genes with *Buchnera* in pathways for essential nutrients synthesis. While two of the three *Periphyllus* species showed high degrees of metabolic complementation, there was a single case (*P. lyropictus*) in which we did not document this pattern. This is likely due to *Serratia* being replaced within this lineage. Symbiont replacement is an important mechanism by which maladaptive symbionts are replaced with new functional ones [29, 30]. In the Lachninae aphids, *Serratia* has been replaced on multiple occasions, including by other more recently acquired *Serratia* strains [9, 12, 13].

The multiple independent origins of *Buchnera-Serratia* co-obligate symbioses also provide a unique opportunity to study the evolution of anatomical integration between hosts and their symbionts. Theory and data suggest that more ancient associations should be characterized by greater integration, both in terms of housing structures and symbiont densities within hosts [31, 32]. While our results generally support this hypothesis, we found that the evolution of specialized structures to house *Serratia* differed between the two most ancient aphid lineages. In the Lachninae, evidence suggests both co-obligate symbionts are located within the same bacteriome, although the arrangement of bacteriocytes differ among species. The *Periphyllus* aphids, by contrast, house the two symbionts in separate structures. This arrangement may have evolved *de novo* or through *Serratia* colonizing pre-existing bacteriocytes that evolved in complexity. The different configurations in the Lachninae and *Periphyllus* potentially represent alternative solutions to the same problem: allowing *Serratia* to reach densities that are high enough to efficiently perform its nutrient-provisioning role. Our curing data further support the idea that greater integration coincides with more reliance on *Serratia* in ancient co-obligates, as symbiont removal had the most dramatic fecundity effect in the *Periphyllus* genus (Figure 1A).

In the cases of *A. urticata* and *M. carnosum*, anatomical and genomic integration data suggest that *Serratia* is a more recently acquired co-obligate association. In *A. urticata*, *Serratia* is housed in a single small bacteriome. In *M. carnosum*, *Serratia* is not hosted in bacteriocytes at all, but rather in sheath cells that surround the large *Buchnera*-containing cells. This is a potential indication of its relatively recent transition to an obligate symbiont. According to this hypothesis, the lack of *Serratia* integration is indicative of its recent role as a facultative symbiont. These findings also suggest that at least in some cases, dependence evolves before the evolution of specialized structures to house symbionts.

Studying evolutionary transitions to obligate and co-obligate symbiosis is difficult because most events are characterized by single and ancient origins across large, diversified clades. This makes comparisons with outgroups less informative and prevents testing ideas on the relative importance of deterministic versus stochastic processes. The uniquely replicated evolutionary events of *Buchnera-Serratia* co-obligate symbiosis in different aphid lineages provide a degree of temporal resolution that demonstrates that co-obligate associations can form in a relatively predictable manner. Furthermore, our findings indicate that genomic integration may occur in advance of anatomical integration. Our results provide evidence that the evolutionary forces that bind multiple organisms into a single metabolic unit operate by deterministic stepwise processes. This allows us to better understand the role of symbioses in the evolution of complex organisms.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Assessment of *Serratia* prevalence
  - Curing experiments
  - Quantitative PCR
  - Phylogeny
  - Whole genomes sequencing, assembly and analysis
  - Fluorescent *in situ* hybridization
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND CODE AVAILABILITY

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2020.03.011.

## ACKNOWLEDGMENTS

The authors thank Alejandro Manzano-Marın for consulting on optimal parameters for genome assembly, Mario Dos Reis Barros for advice on molecular clocking, and three anonymous reviewers for their useful comments. This project was funded by L.H.,NERC IRF (NE/M018016/1); D.M., Marie Skłodowska-Curie Individual Fellowship (H2020-MSCA-IF-2017-796778-SYMOBLIGA); J.E., NWO VICI (865.12.003); and E.T.K. (Ammodo Funds).

## AUTHOR CONTRIBUTIONS

Conceptualization, D.M., L.M.H., and R.J.; Methodology, D.M., L.M.H., M.B., and R.J.; Software, R.J.; Validation, D.M., L.M.H., M.B., and R.J.; Formal Analysis, D.M., M.B., and R.J.; Investigation, D.M., L.M.H., M.B., and R.J.; Resources, D.M. and L.M.H.; Data Curation, R.J.; Writing – Original Draft, D.M., L.M.H., R.J., and E.T.K.; Writing – Review & Editing, D.M., J.E., L.M.H., M.B., R.J., and E.T.K.; Visualization, R.J.; Supervision, J.E. and L.M.H.; Project Administration, L.M.H.; Funding Acquisition, D.M. and L.M.H.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 13, 2020
Revised: February 18, 2020
Accepted: March 3, 2020
Published: April 2, 2020

## REFERENCES

1. McCutcheon, J.P., McDonald, B.R., and Moran, N.A. (2009). Convergent evolution of metabolic roles in bacterial co-symbionts of insects. Proc. Natl. Acad. Sci. USA *106*, 15394–15399.
2. McCutcheon, J.P., and von Dohlen, C.D. (2011). An interdependent metabolic patchwork in the nested symbiosis of mealybugs. Curr. Biol. *21*, 1366–1372.
3. Takiya, D.M., Tran, P.L., Dietrich, C.H., and Moran, N.A. (2006). Co-cladogenesis spanning three phyla: leafhoppers (Insecta: Hemiptera: Cicadellidae) and their dual bacterial symbionts. Mol. Ecol. *15*, 4175–4191.
4. Wu, D., Daugherty, S.C., Van Aken, S.E., Pai, G.H., Watkins, K.L., Khouri, H., Tallon, L.J., Zaborsky, J.M., Dunbar, H.E., Tran, P.L., et al. (2006). Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. PLoS Biol. *4*, e188.
5. Rao, Q., Rollat-Farnier, P.A., Zhu, D.T., Santos-Garcia, D., Silva, F.J., Moya, A., Latorre, A., Klein, C.C., Vavre, F., Sagot, M.F., et al. (2015). Genome reduction and potential metabolic complementation of the dual endosymbionts in the whitefly Bemisia tabaci. BMC Genomics *16*, 226.
6. Douglas, A.E. (2016). How multi-partner endosymbioses function. Nat. Rev. Microbiol. *14*, 731–743.
7. Pérez-Brocal, V., Gil, R., Ramos, S., Lamelas, A., Postigo, M., Michelena, J.M., Silva, F.J., Moya, A., and Latorre, A. (2006). A small microbial genome: the end of a long symbiotic relationship? Science *314*, 312–313.
8. Lamelas, A., Pérez-Brocal, V., Gómez-Valero, L., Gosalbes, M.J., Moya, A., and Latorre, A. (2008). Evolution of the secondary symbiont "Candidatus serratia symbiotica" in aphid species of the subfamily lachninae. Appl. Environ. Microbiol. *74*, 4236–4240.
9. Manzano-Marín, A., Simon, J.C., and Latorre, A. (2016). 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 Biol. Evol. *8*, 1440–1458.
10. Manzano-Marín, A., and Latorre, A. (2014). 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 Biol. Evol. *6*, 1683–1698.
11. Henry, L.M., Maiden, M.C.J., Ferrari, J., and Godfray, H.C.J. (2015). Insect life history and the evolution of bacterial mutualism. Ecol. Lett. *18*, 516–525.
12. Manzano-Marín, A., Szabó, G., Simon, J.C., Horn, M., and Latorre, A. (2017). Happens in the best of subfamilies: establishment and repeated replacements of co-obligate secondary endosymbionts within Lachninae aphids. Environ. Microbiol. *19*, 393–408.
13. Meseguer, A.S., Manzano-Marín, A., Coeur d'Acier, A., Clamens, A.L., Godefroid, M., and Jousselin, E. (2017). Buchnera has changed flatmate but the repeated replacement of co-obligate symbionts is not associated with the ecological expansions of their aphid hosts. Mol. Ecol. *26*, 2363–2378.
14. Nakabachi, A., and Ishikawa, H. (1999). Provision of riboflavin to the host aphid, Acyrthosiphon pisum, by endosymbiotic bacteria, Buchnera. J. Insect Physiol. *45*, 1–6.
15. Douglas, A.E., and Prosser, W.A. (1992). Synthesis of the essential amino acid tryptophan in the pea aphid (Acyrthosiphon pisum) symbiosis. J. Insect Physiol. *38*, 565–568.
16. Douglas, A.E. (1988). Sulphate utilization in an aphid symbiosis. Insect Biochem. *18*, 599–605.
17. Febvay, G., Liadouze, I., Guillaud, J., and Bonnot, G. (1995). Analysis of energetic amino acid metabolism in Acyrthosiphon pisum: a multidimensional approach to amino acid metabolism in aphids. Arch. Insect Biochem. Physiol. *29*, 45–69.
18. Liadouze, I., Febvay, G., Guillaud, J., and Bonnot, G. (1996). Metabolic fate of energetic amino acids in the aposymbiotic pea aphid Acyrthosiphon pisum (Harris) (Homoptera: Aphididae). Symbiosis *21*, 115–127.
19. Brinza, L., Viñuelas, J., Cottret, L., Calevro, F., Rahbé, Y., Febvay, G., Duport, G., Colella, S., Rabatel, A., Gautier, C., et al. (2009). Systemic analysis of the symbiotic function of Buchnera aphidicola, the primary endosymbiont of the pea aphid Acyrthosiphon pisum. C. R. Biol. *332*, 1034–1049.
20. Wilson, A.C.C., Ashton, P.D., Calevro, F., Charles, H., Colella, S., Febvay, G., Jander, G., Kushlan, P.F., Macdonald, S.J., Schwartz, J.F., et al. (2010). Genomic insight into the amino acid relations of the pea aphid, Acyrthosiphon pisum, with its symbiotic bacterium Buchnera aphidicola. Insect Mol. Biol. *19* (*Suppl 2*), 249–258.
21. Shigenobu, S., and Wilson, A.C.C. (2011). Genomic revelations of a mutualism: the pea aphid and its obligate bacterial symbiont. Cell. Mol. Life Sci. *68*, 1297–1309.
22. Gil, R., Silva, F.J., Peretó, J., and Moya, A. (2004). Determination of the core of a minimal bacterial gene set. Microbiol. Mol. Biol. Rev. *68*, 518–537.
23. Bublitz, D.C., Chadwick, G.L., Magyar, J.S., Sandoz, K.M., Brooks, D.M., Mesnage, S., Ladinsky, M.S., Garber, A.I., Bjorkman, P.J., Orphan, V.J., and McCutcheon, J.P. (2019). Peptidoglycan Production by an Insect-Bacterial Mosaic. Cell *179*, 703–712.e7.
24. Sasaki, T., Hayashi, H., and Ishikawa, H. (1991). Growth and reproduction of the symbiotic and aposymbiotic pea aphids, Acyrthosiphon pisum maintained on artificial diets. J. Insect Physiol. *37*, 749–756.
25. Russell, C.W., Poliakov, A., Haribal, M., Jander, G., van Wijk, K.J., and Douglas, A.E. (2014). Matching the supply of bacterial nutrients to the nutritional demand of the animal host. Proc. Biol. Sci. *281*, 20141163.
26. Akashi, H., and Gojobori, T. (2002). Metabolic efficiency and amino acid composition in the proteomes of Escherichia coli and Bacillus subtilis. Proc. Natl. Acad. Sci. USA *99*, 3695–3700.

27. McCutcheon, J.P., and Moran, N.A. (2010). Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. Genome Biol. Evol. *2*, 708–718.
28. Chong, R.A., Park, H., and Moran, N.A. (2019). Genome Evolution of the Obligate Endosymbiont Buchnera aphidicola. Mol. Biol. Evol. *36*, 1481–1489.
29. Bennett, G.M., and Moran, N.A. (2015). Heritable symbiosis: the advantages and perils of an evolutionary rabbit hole. Proc. Natl. Acad. Sci. USA *112*, 10169–10176.
30. Sudakaran, S., Kost, C., and Kaltenpoth, M. (2017). Symbiont Acquisition and Replacement as a Source of Ecological Innovation. Trends Microbiol. *25*, 375–390.
31. Frank, S.A. (1997). Models of symbiosis. Am. Nat. *150* (*Suppl 1*), S80–S99.
32. Kiers, E.T., and West, S.A. (2015). Evolving new organisms via symbiosis. Science *348*, 392–394.
33. Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y., and Ishikawa, H. (2000). Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS. Nature *407*, 81–86.
34. Jiang, Z., Jones, D.H., Khuri, S., Tsinoremas, N.F., Wyss, T., Jander, G., and Wilson, A.C. (2013). Comparative analysis of genome sequences from four strains of the Buchnera aphidicola Mp endosymbion of the green peach aphid, Myzus persicae. BMC Genomics *14*, 917.
35. Cassone, B.J., Wenger, J.A., and Michel, A.P. (2015). Whole genome sequence of the soybean aphid endosymbiont Buchnera aphidicola and genetic differentiation among biotype-specific strains. J Genomics *3*, 85–94.
36. Lamelas, A., Gosalbes, M.J., Manzano-Marín, A., Peretó, J., Moya, A., and Latorre, A. (2011). Serratia symbiotica from the aphid Cinara cedri: a missing link from facultative to obligate insect endosymbiont. PLoS Genet. *7*, e1002357.
37. Gil, R., Sabater-Muñoz, B., Perez-Brocal, V., Silva, F.J., and Latorre, A. (2006). Plasmids in the aphid endosymbiont Buchnera aphidicola with the smallest genomes. A puzzling evolutionary story. Gene *370*, 17–25.
38. Foray, V., Grigorescu, A.S., Sabri, A., Haubruge, E., Lognay, G., Francis, F., Fauconnier, M.L., Hance, T., and Thonart, P. (2014). Whole-genome sequence of *Serratia symbiotica* strain CWBI-2.3$^{T}$, a free-living symbiont of the black bean aphid *Aphis fabae*. Genome Announc. *2*, e00767-14.
39. Manzano-Marín, A., Coeur d'acier, A., Clamens, A.L., Orvain, C., Cruaud, C., Barbe, V., and Jousselin, E. (2018). A freeloader? The highly eroded yet large genome of the Serratia symbiotica symbiont of Cinara strobi. Genome Biol. Evol. *10*, 2178–2189.
40. Burke, G.R., and Moran, N.A. (2011). Massive genomic decay in Serratia symbiotica, a recently evolved symbiont of aphids. Genome Biol. Evol. *3*, 195–208.
41. Iguchi, A., Nagaya, Y., Pradel, E., Ooka, T., Ogura, Y., Katsura, K., Kurokawa, K., Oshima, K., Hattori, M., Parkhill, J., et al. (2014). Genome evolution and plasticity of Serratia marcescens, an important multidrug-resistant nosocomial pathogen. Genome Biol. Evol. *6*, 2096–2110.
42. Ferrari, J., West, J.A., Via, S., and Godfray, H.C.J. (2012). Population genetic structure and secondary symbionts in host-associated populations of the pea aphid complex. Evolution *66*, 375–390.
43. Dunbar, H.E., Wilson, A.C.C., Ferguson, N.R., and Moran, N.A. (2007). Aphid thermal tolerance is governed by a point mutation in bacterial symbionts. PLoS Biol. *5*, e96.
44. Burke, G., Fiehn, O., and Moran, N. (2010). Effects of facultative symbionts and heat stress on the metabolome of pea aphids. ISME J. *4*, 242–252.
45. Koga, R., Tsuchida, T., and Fukatsu, T. (2003). Changing partners in an obligate symbiosis: a facultative endosymbiont can compensate for loss of the essential endosymbiont Buchnera in an aphid. Proc. Biol. Sci. *270*, 2543–2550.
46. Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. *75*, 7537–7541.
47. Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. *41*, D590–D596.
48. Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D., and Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. *47* (W1), W636–W641.
49. Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. *59*, 307–321.
50. Rambaut, A. (2007). FigTree, a graphical viewer of phylogenetic trees. http://tree.bio.ed.ac.uk/software/figtree/.
51. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. *24*, 1586–1591.
52. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.
53. Nurk, S., Bankevich, A., Antipov, D., Gurevich, A.A., Korobeynikov, A., Lapidus, A., Prjibelski, A.D., Pyshkin, A., Sirotkin, A., Sirotkin, Y., et al. (2013). Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. J. Comput. Biol. *20*, 714–737.
54. Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nat. Methods *12*, 59–60.
55. Lee, M.D. (2019). GToTree: a user-friendly workflow for phylogenomics. Bioinformatics *35*, 4162–4164.
56. R Development Core Team (2019). R: A language and environment for statistical computing (R Foundation for Statistical Computing).
57. Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., and de Hoon, M.J. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics *25*, 1422–1423.
58. Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics *11*, 119.
59. Shen, W., and Xiong, J. (2019). TaxonKit: a cross-platform and efficient NCBI taxonomy toolkit. bioRxiv. https://doi.org/10.1101/513523.
60. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.
61. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973.
62. Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS One *5*, e9490.
63. Foottit, R.G., Maw, H.E.L., VON Dohlen, C.D., and Hebert, P.D.N. (2008). Species identification of aphids (Insecta: Hemiptera: Aphididae) through DNA barcodes. Mol. Ecol. Resour. *8*, 1189–1201.
64. Heie, O.E. (1980). The Aphidoidea (Hemiptera) of Fennoscandia and Denmark. I. General Part. The Families Mindaridae, Hormaphididae, Thelaxidae, Anoeciidae, and Pemphigidae (Fauna Entomologica Scandinavica) (Scandinavian Science Press).
65. Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M., et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. *6*, 1621–1624.
66. Koga, R., Tsuchida, T., Sakurai, M., and Fukatsu, T. (2007). Selective elimination of aphid endosymbionts: effects of antibiotic dose and host genotype, and fitness consequences. FEMS Microbiol. Ecol. *60*, 229–239.
67. von Dohlen, C.D., and Moran, N.A. (1995). Molecular phylogeny of the Homoptera: a paraphyletic taxon. J. Mol. Evol. *41*, 211–223.

68. Chen, R., Favret, C., Jiang, L., Wang, Z., and Qiao, G. (2016). An aphid lineage maintains a bark-feeding niche while switching to and diversifying on conifers. Cladistics *32*, 555–572.

69. O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. *44* (D1), D733–D745.

70. Darling, A.E., Mau, B., and Perna, N.T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One *5*, e11147.

71. Tanizawa, Y., Fujisawa, T., and Nakamura, Y. (2018). DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. Bioinformatics *34*, 1037–1039.

72. Riley, M., Abe, T., Arnaud, M.B., Berlyn, M.K.B., Blattner, F.R., Chaudhuri, R.R., Glasner, J.D., Horiuchi, T., Keseler, I.M., Kosuge, T., et al. (2006). Escherichia coli K-12: a cooperatively developed annotation snapshot–2005. Nucleic Acids Res. *34*, 1–9.

73. Caspi, R., Foerster, H., Fulcher, C.A., Kaipa, P., Krummenacker, M., Latendresse, M., Paley, S., Rhee, S.Y., Shearer, A.G., Tissier, C., et al. (2008). The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. *36*, D623–D631.

74. Karp, P.D., Latendresse, M., Paley, S.M., Krummenacker, M., Ong, Q.D., Billington, R., Kothari, A., Weaver, D., Lee, T., Subhraveti, P., et al. (2016). Pathway Tools version 19.0 update: software for pathway/genome informatics and systems biology. Brief. Bioinform. *17*, 877–890.

75. Emms, D., and Kelly, S. (2019). OrthoFinder2: phylogenetic orthology inference for comparative genomics. Genome Biol *20*, 238.

76. Koga, R., Tsuchida, T., and Fukatsu, T. (2009). Quenching autofluorescence of insect tissues for in situ detection of endosymbionts. Appl. Entomol. Zool. *44*, 281–291.

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Deposited Data | | |
| NCBI nucleotide database | https://www.ncbi.nlm.nih.gov/ | EU071358.1; EF073226.1; EF073230.1; EU358908.1; JF950580.1; EU358928.1; KX645787.1; KX645797.1; KT238076.1; FM174683.1; FM174684.1; DQ005157.1; AF163879.1; HM117788.1; KM501168.1; AF147812.1; AF275250.1; AF275216.1; AF275214.1; EU358868.1; EU358888.1; JX965987.1; KX507107.1; KX507113.1; JX965996.1; HM117803.1; AF275246.1; CP034882.1; KX620606.1; KX620618.1; KX620620.1; KX620627.1; CP000263.1; CP001817.1; CP034876.1; JX998110.1; CP033006.1; JX998124.1; LN890285.1; CP002301.1; AP010872.1; CP022500.1; JF883920.1; KR034269.1; EF073073.1; KY323028.1; KF638947.1; MH183024.1; KF639284.1; KX680185.1; KT237845.1; KU321598.1; JQ916729.2; KP189472.1; JN032720.1; JF883800.1; JQ916732.2; KT237876.1 |
| *Buchnera aphidicola* str. APS (*Acyrthosiphon pisum*) (enterobacteria) | [33] | GCF_000009605.1 |
| *Buchnera aphidicola* str. G002 (*Myzus persicae*) | [34] | GCF_000521565.1 |
| *Buchnera aphidicola* (*Aphis glycines*) (enterobacteria) | [35] | GCF_001280225.1 |
| *Buchnera aphidicola* (*Cinara tujafilina*) | [36] | GCF_000217635.1 |
| *Buchnera aphidicola* BCc (enterobacteria) | [7, 37] | GCF_000090965.1 |
| *Buchnera aphidicola* (*Tuberolachnus salignus*) (enterobacteria) | Centre de Biologie pour la Gestion des Populations (CBGP - UMR1062) | GCF_900016785.1 |
| *Serratia symbiotica* SCt-VLC (enterobacteria) | [10] | GCA_900002265.1 |
| *Serratia symbiotica* str. 'Cinara cedri' (enterobacteria) | Instituto Cavanilles de Biodiverasidad y Biologia evolutiva | GCA_000238975.1 |
| *Serratia symbiotica* (enterobacteria) | Centre de Biologie pour la Gestion des Populations (CBGP - UMR1062) | GCA_900016775.1 |
| *Buchnera aphidicola* and *Serratia symbiotica* (*A. urticata*, *M. carnosum*, *P. lyropictus*, *P. acericola*, *P. aceris*) | This study | PRJNA605335 |
| *Serratia symbiotica* (*Aphis fabae*) | [38] | GCA_000821185.1 |
| *Serratia symbiotica* (*A. pisum*, strain IS) | The Open University of Japan | GCA_008370165.1 |
| *Serratia symbiotica* (*Cinara strobi*) | [13, 39] | GCA_900380265.1 |
| *Serratia symbiotica* (*Aphis fabae*, strain 24.1) | Université Catholique de Louvain | GCA_009831665.1 |
| *Serratia symbiotica* (*Aphis passeriniana*) | Université Catholique de Louvain | GCA_009831785.1 |
| *Serratia symbiotica* (*Acyrthosiphon pisum*, Tucson) | [40] | GCA_000186485.2 |

(*Continued on next page*)

**Continued**

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| *Serratia symbiotica* (*Cinara fornacula*) | [13] | GCA_900155695.1 |
| *Xanthomonas axonopodis* Xac29-1 (g-proteobacteria) | Shanghai Jiao Tong University | GCF_000348585.1 |
| *Serratia marcescens* subsp. marcescens Db11 (enterobacteria) | [41] | GCF_000513215.1 |
| *Serratia plymuthica* (enterobacteria) | N/A | GCF_900478125.1 |
| Oligonucleotides | | |
| 16S *Serratia symbiotica* specific primers 16sS A1: 5′-AGAGTT TGATCMTGGCTCAG-3′; 16sS 2R: 5′-TTTGAGTT CCCGACTTTAACG-3′ | This study | N/A |
| 16S *Serratia* generalist primers 16sS 10F: 5′-AGTTTGATCA TGGCTCAGATTG-3′; 16sS R443R: 5′-CTTCTGCGAGTA ACGTCAATG-3′ | [42] | N/A |
| *Buchnera dnaK* primers: BHS70F2: 5′-ATGGGTAAAATTATTGGTATTG-3′; BHS70R2: BHS70R2: 5′-ATAGCTTGACGTTTAGCAGG-3′. | [43] | N/A |
| *Serratia dnaK* primers: ApRF1: 5′-TGGCGGGTGATGTGAAG-3′; ApRR1: 5′-CGGGATAGTG GTGTTTTTGG-3′ | [44] | N/A |
| Aphid *elongation factor 1-α* primers: ApEF1-alpha 107F: 5′-CTGA TTGTGCCGTGCTTATTG-3′; ApEF1-alpha 246R: 5′-TAT GGTGGTTCAGTAGAGTCC-3′. | [43] | N/A |
| *Buchnera* 16S probe (*A. pisum*, *A. urticata*, *M. carnosum*): 5′-Alexa Fluor 488-CCTCTTTTGGG TAGATCC-3′ | [45] | N/A |
| *Buchnera* 16S probe (*Periphyllus* spp.) probe: 5′-Alexa Fluor 488-CCTTTTTTGGGCAGATTC-3′ | This study | N/A |
| *Serratia* 16S probe (all species): 5′-Cy3-CCCGACTTTATCGCTGGC-3′ | [45] | N/A |
| Software and Algorithms | | |
| MOTHUR | [46] | https://www.mothur.org/ |
| SILVA alignment | [47] | https://www.mothur.org/ |
| MUSCLE | [48] | https://www.ebi.ac.uk/Tools/msa/muscle/ |
| PhyML | [49] | http://www.atgc-montpellier.fr/phyml/ |
| FigTree v1.4.4 | [50] | http://tree.bio.ed.ac.uk/software/figtree/ |
| mcmctree function in PAML | [51] | http://abacus.gene.ucl.ac.uk/software/paml.html#download |
| Trimmomatic | [52] | http://www.usadellab.org/cms/?page=trimmomatic |
| SPAdes v3.11.1 | [53] | http://cab.spbu.ru/software/spades/ |
| DIAMOND | [54] | http://www.diamondsearch.org/index.php |
| CODONCODE ALIGNER version 4.0.2 | CodonCode Corporation 2012, Centerville, MA, USA | https://www.codoncode.com/aligner/ |

(*Continued on next page*)

*Continued*

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| CLC genomic workbench 12.0 | QIAGEN, Venlo, Netherlands | https://digitalinsights.qiagen.com/products-overview/analysis-and-visualization/qiagen-clc-genomics-workbench/?gclid=EAIaIQobChMIqJPWgam45wIVArTtCh3CpQuWEAAYASAAEgKAUvD_BwE |
| GToTree v.1.4.7 | [55] | https://github.com/AstrobioMike/GToTree/wiki/installation |
| R 3.6.0 | [56] | https://www.r-project.org/ |
| Biopython | [57] | https://biopython.org/ |
| HMMER3 v3.2.1 | N/A | http://hmmer.org/ |
| Prodigal v2.6.3 | [58] | https://omictools.com/prodigal-tool |
| TaxonKit v0.3 | [59] | https://bioinf.shenwei.me/taxonkit/download/ |
| Muscle v3.8 | [60] | https://www.drive5.com/muscle/ |
| Trimal v1.4 | [61] | http://trimal.cgenomics.org/ |
| FastTree v2.1 | [62] | http://www.microbesonline.org/fasttree/ |

## LEAD CONTACT AND MATERIALS AVAILABILITY

This study did not generate new unique reagents. Further information and requests for resources should be directed to the Lead Contact, Lee Henry (l.henry@qmul.ac.uk).

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

The following aphid species were used in this study: Acyrthosiphon pisum, Macrosiphoniella artemisiae, Aphis urticata, Microlophium carnosum, Chaitophorus leucomelas, C. populeti, Periphyllus acericola, P. aceris, P. coracinus, P. hirticornis, P. lyropictus and an unknown Periphyllus species (referred to as Periphyllus sp. in the main text). Clonal lines of aphids were maintained in the lab at 15°C with a 16 hours light (Sylvania Gro-Lux F36W/GRO-T8 bulb) 8h dark cycle on a leaf of their host plant: *Vicia faba* (A. pisum), Artemisia vulgaris (M. artemisiae), Urtica dioica (A. urticata and M. carnosum), Acer spp. (Periphyllus spp.) embedded in agar in a Petri dish to keep the leaves fresh. Leaves were changed as needed.

## METHOD DETAILS

### Assessment of *Serratia* prevalence

#### *Collection and identification of aphid samples*

Aphids were collected in the UK and the Netherlands between 2011 and 2019. They were dislodged by beating plants over a white tray or removed manually from the plant, before being placed in 100% ethanol or collected live for curing experiments. Resampling of the same aphid clones was minimized by separating collections from the same plant species by at least 10 m. Aphids were identified by barcoding based on data from [63] and morphological examination following [64]. We sequenced the COI barcoding region using standard protocols for DNA extraction, amplification and editing and performed alignments with MUSCLE in CODONCODE ALIGNER version 4.0.2 (CodonCode Corporation 2012, Centerville, MA, USA). Genomic DNA was extracted from individual specimens using DNeasy Blood and Tissue kits (QIAGEN, Venlo, Netherlands) and we amplified an approximately 700 bp DNA fragment of the cytochrome *c* oxidase I (COI) mitochondrial gene using Lep F (5′-ATTCAACCAATCATAAAGATATTGG-3′) and Lep R (5′-TAAACTTCTGGATGTCCAAAAAATCA-3′) primers, which was sequenced in the forward direction. Aphids were identified to species by comparing COI sequence data to the online databases BOLD (http://www.boldsystems.org/) and GenBank using BLAST. Morphological examination was carried out by macerating individual aphids in KOH (Potassium Hydroxide) and mounting them on microscope slides.

#### *Diagnostic screening for* Serratia symbiotica

We confirm all *A. urticata*, *M. carnosum* and five species of *Periphyllus* aphids ubiquitously carry *Serratia*, in the UK and the Netherlands, by amplifying a partial region of the 16S rRNA gene using the specific primers 16sS A1 (5′-AGAGTTTGATCMTGGCTCAG-3′) and 16sS 2R (5′-TTTGAGTTCCCGACTTTAACG-3′) and sequencing the amplicons. The PCR cycling conditions were as follow: 3 minutes at 95°C followed by 30 cycles of 30 s at 95°C, 1 minute at 52°C and 1 minute at 72°C, and finally 5 minutes at 72°C. To confirm the primers were only amplifying *S. symbiotica* we compared sequences to published records on GenBank using BLAST. If the specific primers failed to amplify due to primer binding specificity we used more general *Serratia* primers that amplify diverse *Serratia* species: 16sS 10F (5′-AGTTTGATCATGGCTCAGATTG-3′) and 16sS R443R (5′-CTTCTGCGAGTAACGTCAATG-3′),

and confirmed the presence of *S. symbiotica* by comparing sequences to those on GenBank using BLAST. PCR cycling conditions were as follows, 2 minutes at 94°C, followed by 10 cycles of one minute at 94°C, 1 minute at 65°C-55°C (touchdown in 1°C steps) and 2 minutes at 72°C 2:00, followed by 25 cycles of 1 minute at 94°C, 1 minute at 55°C and 2 minutes at 72°C, and finally, 6 minutes at 72°C [42].

To estimate the origins of the co-dependence in *Periphyllus* aphids we sampled for *Serratia* in the *Chaitophorus* genus, which is a sister taxon to *Periphyllus* [11]. We deep-sequenced the universal bacterial 16S rRNA gene in *Chaitophorus populeti* (7 samples) and *C. leucomelas* (1 sample) to confirm *Serratia* was not ubiquitously present in these aphids, indicating they had not evolved co-dependence on the symbiont or been replaced. We PCR amplified the V4 region of the bacteria 16S rRNA gene following standard protocols [65], and deep-sequenced the amplicons using the Illumina MiSeq2000 platform. 16S rRNA analysis was performed using the standard operating procedure for MOTHUR [46], including read joining filtering, and Operational taxonomic unit (OTU) selection at 99 percent. Taxonomic assignments of the reads were performed using the full length SILVA alignment [47] available from mothur.org. As a final step samples were filtered using R to only consider OTUs at a one percent relative abundance or higher in the sample. No OTU(s) corresponding to *Serratia symbiotica* were found in any of these samples. The total and absolute number of OTU reads for each symbiont species are presented in Data S1C.

## Curing experiments

To selectively cure aphids of *Serratia* while not affecting *Buchnera*, we used antibiotics that specifically target cell walls, which are reduced in *Buchnera* [66]. Curing experiments were conducted on *A. urticata, M. carnosum,* and the two *Periphyllus* species that we were able to successfully culture in the lab: *P. hirticornis* and *P. lyropictus* in both the UK and the Netherlands. We were unable to culture the remaining *Periphyllus* species in the lab so they were not include in the curing assays. *Acyrthosiphon pisum* (UK and NL) and *M. artemisiae* (UK only) were included to confirm the antibiotic treatments had no consistent negative fitness effects on species that harbor *Serratia* as a facultative symbiosis (*A. pisum*) or were uninfected by the symbiont (*M. artemisiae*). The antibiotic solution was obtained by mixing 10 mg/mL of Ampicillin sodium salt, 5 mg/mL Cefotaxime sodium salt, and 5 mg/mL Gentamicin in water. All antibiotics were obtained from Sigma-Aldrich (St. Louis, MO, USA). A single leaf of the host plant was cut and placed in a 0.5 mL Eppendorf tube filled with either one of two treatments i) the antibiotic solution or ii) water (control treatment). For the curing assay we placed 3-5 one or two day old aphid nymphs on a leaf and left them to feed for five (UK experiment) or three (Dutch experiment) days on either water or antibiotic solution. At the end of the treatment, aphids were transferred to their own individual Petri dishes, each containing a single leaf in agar from their appropriate host plants. The leaves were changed each week and the lifetime fecundity of each aphid was recorded by counting the number of offspring. In the UK experiment, a sub-sample of first-generation aphids (9 control and 11 treated *M. carnosum* and 10 control and 10 treated *P. hirticornis*) were sacrificed at ~20 days old to quantify symbiont density using qPCR. The experiment was performed in the temperature and lighting conditions described previously.

## Quantitative PCR

The relative density of *Buchnera* compared to the density of *Serratia* was measured by quantitative PCR on whole aphid DNA extracts. We used three single copy genes: one in the aphid nuclear genome (*elongation factor 1-α*) and one in each of the two symbionts (*dnaK* gene). The primers were as follows. *Buchnera dnaK* gene [43]: BHS70F2: 5′-ATGGGTAAAATTATTGGTATTG-3′; BHS70R2: BHS70R2: 5′-ATAGCTTGACGTTTAGCAGG-3′. *Serratia dnaK* gene [44]: ApRF1: 5′-TGGCGGGTGATGTGAAG-3′; ApRR1: 5′-CGGGATAGTGGTGTTTTTGG-3′. Aphid *elongation factor 1-α* [43]: ApEF1-alpha 107F: 5′-CTGATTGTGCCGTGCT TATTG-3′; ApEF1-alpha 246R: 5′-TATGGTGGTTCAGTAGAGTCC-3′. The quantification was performed on a CFX Connect Real-Time PCR Detection System (BioRad, Hercules, California, U.S.A.). The PCR reaction mixture included 10 μL Luna Universal qPCR Master Mix (New England BioLabs, Ipswich, Massachusetts, USA), 7 μL $H_2O$, 0.5 μL of each primer (10 nM), and 2 μL DNA. The cycling conditions were: 15 min activation at 95°C followed by 40 cycles at 95°C for 15 s, at 60°C for 1 min, and 95°C for 15 s. The mean efficiencies were calculated using a ten-fold series of dilutions from $10^2$ to $10^7$ copies of purified PCR products. The efficiencies were 95.7 for the aphid gene, 96.8 for the *Buchnera* gene, and 97.3 for the *Serratia* gene. Duplicate samples were used for the determination of DNA quantities. As the deviations between the duplicates were below 0.5 cycles, the mean Cp values were used to calculate starting quantity. For each sample, the starting quantity for the *Buchnera* gene was divided by the starting quantity for the aphid gene to obtain the *Buchnera* density, and the starting quantity for the *Serratia* gene was divided by the starting quantity for the aphid gene to obtain the *Serratia* density.

## Phylogeny

The aphid phylogeny was built using a concatenation of four genes: *Elongation factor 1 alpha*, *12S ribosomal RNA*, *16S ribosomal RNA* and *Cytochrome oxidase subunit 1*. Sequences were obtained from NCBI (Data S1G) and from the genomic data original to this study. Genes were concatenated using CLC genomic workbench 12.0 and aligned using MUSCLE [48]. Maximum Likelihood (ML) phylogeny was generated using the online PhyML server [49]. The phylogeny was bootstrapped 100 times, and rooted using sequences from *Adelges japonicus*, *A. couleyi*, *Candidatus* Ishikawaella capsulata and *Salmonella enterica*. The Adelgidae are basal to the Aphididae [67], *Candidatus* Ishikawaella capsulata and *Salmonella enterica* are outgroups belonging to the same family as *Buchnera aphidicola*. The tree was visualized using FigTree v1.4.4 [50]. *Chaitophorus saliniger* was included to improve node support then pruned from the phylogeny to only retain species for which *Serratia*-infection data are available.

We dated the phylogeny using the mcmctree function in PAML [51]. The calibrated the molecular clock using the estimated divergence (97.45-77.65 MYA) between the Lachnini (e.g., *T. salignus*) and Eulachnini (e.g., *C. cedri*) [68].

The *Serratia* phylogeny was built using GToTree v.1.4.7 under default parameters [55]. GToTree makes use of Biopython [57], HMMER3 v3.2.1 (hmmer.org), Prodigal v2.6.3 [58], TaxonKit v0.3 [59], Muscle v3.8 [60], Trimal v1.4 [61], and FastTree v2.1 [62]. Accession numbers as listed in Data S1F were provided alongside the GBK files generated by DFAST for the *Serratia* genomes we sequenced (see subsequent methods). Gammaproteobacteria was specified as the single copy gene set to use and species name information was added using the parameters -t -L Species.

## Whole genomes sequencing, assembly and analysis

We obtained whole genome sequences for all co-obligate aphid species available to us at the time: *A. urticata*, *M. carnosum*, *P. acericola*, *P. aceris* and *P. lyropictus* (in addition to *P. testudinaceus* and *P. hirticornis*). DNA was extracted from individual aphids (one sample per species) and sequenced at Centre for Genomic Research (University of Liverpool). The libraries were prepared using the Nextera XT kit, and sequenced on an Illumina HiSeq 4000 (paired-end, 2x150 bp reads). Seven samples were multiplexed on one lane. Two samples were discarded at this stage (*P. testudinaceus* and *P. hirticornis*) because the coverage of the symbiont genomes was not high enough to include in the comparative analysis. Average genome coverage for the endosymbionts of interest was ~950x for *Buchnera* and ~97x for *Serratia* (Data S1H). The aphid hosts genomes were not analyzed.

Reads were trimmed for quality and Illumina adaptors were removed using Trimmomatic [52] under default settings. Reads were assembled using SPAdes v3.11.1 [53] in two stages. In the first stage and assembly was built, using assembly only mode with other parameters as default. The reads were then mapped back to this assembly using bwa mem. Contigs were partitioned into *Buchnera* and *Serratia* bins based on a DIAMOND [54] search of the contigs against the NCBI's non-redundant Refseq protein database [69]. The reads mapping to the contigs in the *Buchnera* bin were then reassembled using SPAdes, this time using error correction and in careful mode with kmer sizes of 33, 55, 77, 99 and 127. The contigs of the resulting assembly were filtered by coverage and identity, then blasted against the NCBI's non-redundant Refseq nucleotide database [69]. The results were manually inspected and contigs belonging to species other than *Buchnera* were removed.

Three of the five genomes, the *Buchnera* strains belonging to the *Periphyllus* species, were not able to be fully closed due to repetitive high AT content regions. These genomes were aligned against *Buchnera aphidicola* strain APS [33] and each other using MAUVE [70] and inspected using Geneious Prime 2019.06.17 (https://www.geneious.com) to ensure the contigs were oriented correctly. Using the high level of gene synteny between *Buchnera* lineages we manual inspected the gaps of all five genomes to confirm they could not contain genes relevant to provisioning of essential nutrients. In only one case, the genome of *P. aceris* contained a gap with the genes *ribE* and *ribD*, which are part of the riboflavin pathway. In this case, the remaining genes in the riboflavin pathway were confirmed to be absent followed by a comparison with the *Buchnera* of the other two *Periphyllus* species. Both species had a similar deletion that spanned the area of the gap. Finally, the assembled sequencing data from before the binning stage was inspected to ensure these genes were not overlooked due to incorrect binning. In this way we confirmed the pathway was non-functioning in this *Buchnera* lineage

We annotated both *Serratia* and *Buchnera* genes using DFAST [71] with *Escherichia coli* K-12 genome annotations as a reference [72]. To ensure we did not miss any genes in our final assemblies, we also annotated the initial assembly prior to partitioned into *Buchnera* and *Serratia* bins, and confirmed they did not contain any additional genes involved in nutrient provisioning. Genes making up each pathway were determined using the Metacyc [73] pathway reference for *E. coli*. Metabolic pathways of the co-obligate *Buchnera* lineages were compared to three *Buchnera* strains where it is the sole obligate symbiont (strains APS, G002, BAg: Data S1H) to target regions that had been deleted from *Buchnera* that are in pathways involved in synthesizing essential nutrients. Where genes involved in nutrient pathways were deleted in *Buchnera*, we confirmed orthologous copies of the deleted genes had been retained in the *Serratia* genome. The presence of genes that are part of nutrient pathways in each *Buchnera* and *Serratia* genome was evaluated using Pathway Tools [74] and manual examination of the annotation files. See Data S1D for full results. We additionally evaluated the presence of genes, using the same methods, involved in the synthesis of peptidoglycan and genes involved in translation. See Data S1E for these results.

Additionally, in order to investigate what genes might be missing from the *A. urticata* strain of *Buchnera,* we used Orthofinder v2.2.7 [75] with default parameters to group orthologous genes between all of the *Buchnera* strains investigated (as listed in Data S1H).

## Fluorescent *in situ* hybridization

Whole mount FISH was performed following a protocol adapted from [76]. Aphids were fixed overnight in Carnoy's solution (60% ethanol, 30% chloroform, 10% glacial acetic acid). The aphids were then bleached in an alcoholic $H_2O_2$ solution (80% ethanol, 14% $H_2O$, 6% $H_2O_2$) for 3 days, changing the solution each day. The samples were then thoroughly washed in 90% ethanol and kept at −25°C until processed. The head of the aphids was removed to facilitate the penetration of the probes, and the samples were washed 4 times (30 minutes each) in PBSTw (Phosphate Buffer Saline with 0.02% of Tween 20), and then 3 times (5 minutes each) in hybridization buffer (20 mM Tris-HCl at pH 8.0, 0.9 M NaCl, 0.01% sodium dodecyl sulfate and 30% formamide). The samples were then incubated overnight at room temperature in hybridization buffer supplemented with 100 mM of each 16S rRNA fluorescent probe, one targeting *Buchnera*, one targeting *Serratia*. The probes were as follows: *Buchnera* (*A. pisum*, *A. urticata*, *M. carnosum*): 5′-Alexa Fluor 488-CCTCTTTTGGGTAGATCC-3′ [45], *Buchnera* (*Periphyllus* spp.): 5′-Alexa Fluor 488-CCTTTTTTGGGCAGATTC-3′, *Serratia* (all species): 5′-Cy3-CCCGACTTTATCGCTGGC-3′ [45]. Following a washing in PBSTw,

the samples were mounted on slides in vectashield hardset antifade mounting media with DAPI (to highlight the host body). Mounted samples were visualized using a Leica DMRA2 epi-fluorescent microscope. Monochrome pictures were obtained using a Hamamatsu Orca camera and the Volocity 6.3.1 software, and final color images were obtained using ImageJ. Probes were ordered from Eurogentec (Seraing, Belgium). All the *Periphyllus* species that we were able to collect from the field at the time of the FISH experiment are shown in Figure 4. Two of those species (*Periphyllus* sp. and *P. coracinus*) were not previously collected by us, which is why they were not included in the other experiments.

## QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses were performed using R 3.6.0 [56]. Lifetime fecundity was analyzed using GLMs with a quasi-Poisson distribution. Symbiont density data were analyzed using GLMs with a Gamma distribution.

## DATA AND CODE AVAILABILITY

The accession number for the data reported in this paper is Genbank: PRJNA605335. No novel software was created for this study.

**Current Biology, Volume *30***

## Supplemental Information

# Parallel Evolution in the Integration of a Co-obligate Aphid Symbiosis

**David Monnin, Raphaella Jackson, E. Toby Kiers, Marie Bunker, Jacintha Ellers, and Lee M. Henry**

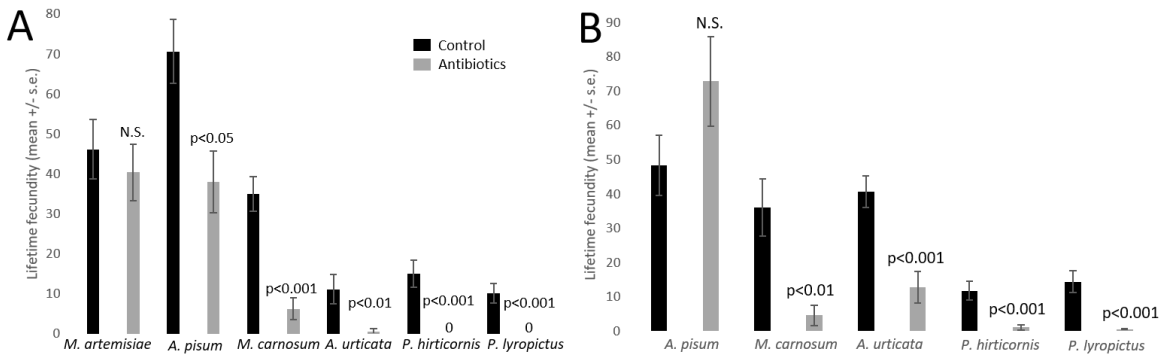


**Figure S1. Effect of antibiotic curing on aphid lifetime fecundity. Related to Figure 1A.** (A) Experiment 1 (United Kingdom) aphids treated for 5-days with antibiotics, and (B) Experiment 2 (Netherlands) aphids treated for 3-days with antibiotics. In both experiments, *A. pisum* carried *Serratia*. In the UK experiment, *M. artemisiae*, was included as an uninfected control.

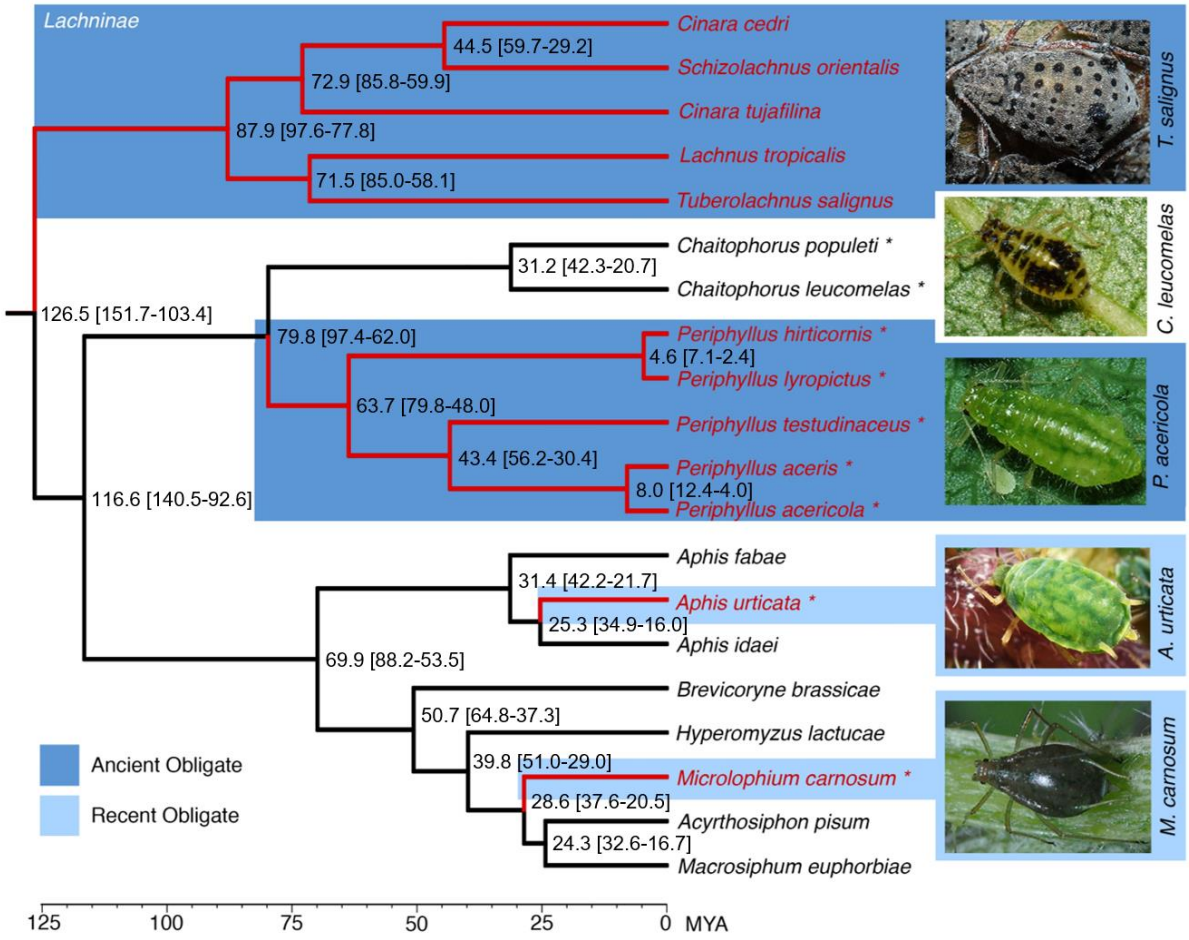


**Figure S2. Aphid phylogeny with ages based on a molecular clock. Related to Figure 2.** Age estimates and 95 % confidence intervals are displayed at nodes.

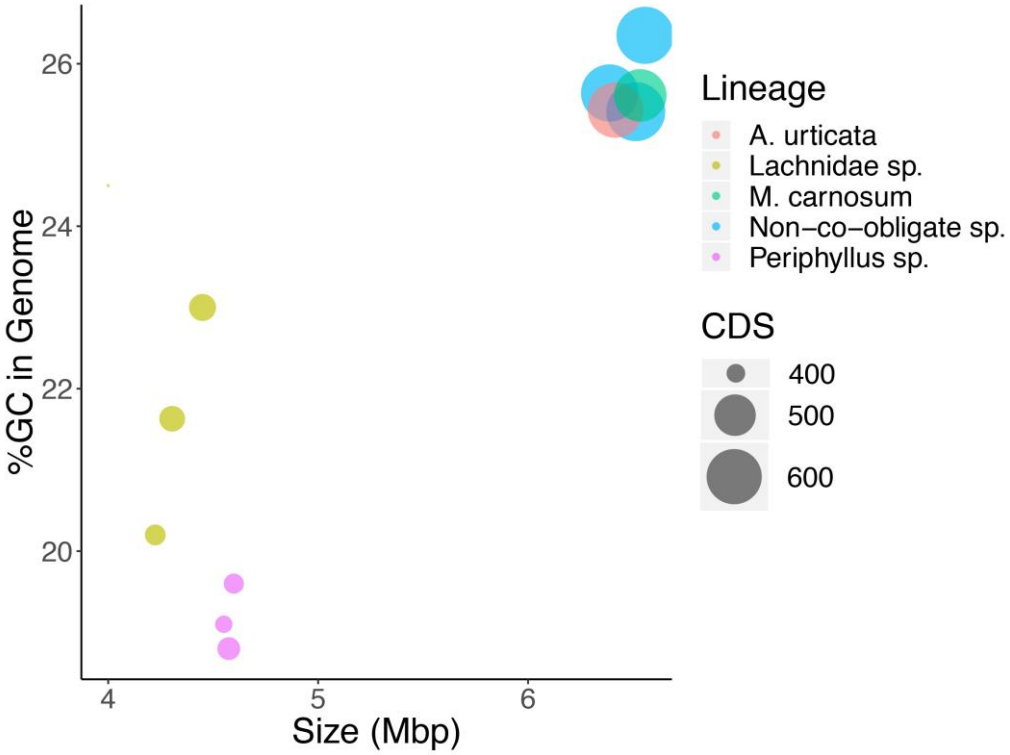


**Figure S3. Genome size, GC content and number of protein coding sequences (CDS) in *Buchnera* genomes in aphid lineages that have evolved dependency on *Serratia* compared to those that have not. Related to Figure 3.** Aphid species in the Lachninae subfamily and *Periphyllus* genus have relatively small genomes, with low CDS and GC content indicative of additional genome erosion that is a product of an ancient co-obligate association with *Serratia*. Conversely, *Buchnera* from *A. urticata* and *M. carnosum* have more similar genomic features to species where *Buchnera* is the sole obligate symbionts (represented here by non-co-obligate species *A. pisum*, *M. persicae* and *A. glycines*) despite having evolved dependency on the symbiont, suggesting a more recent co-obligate relationship.

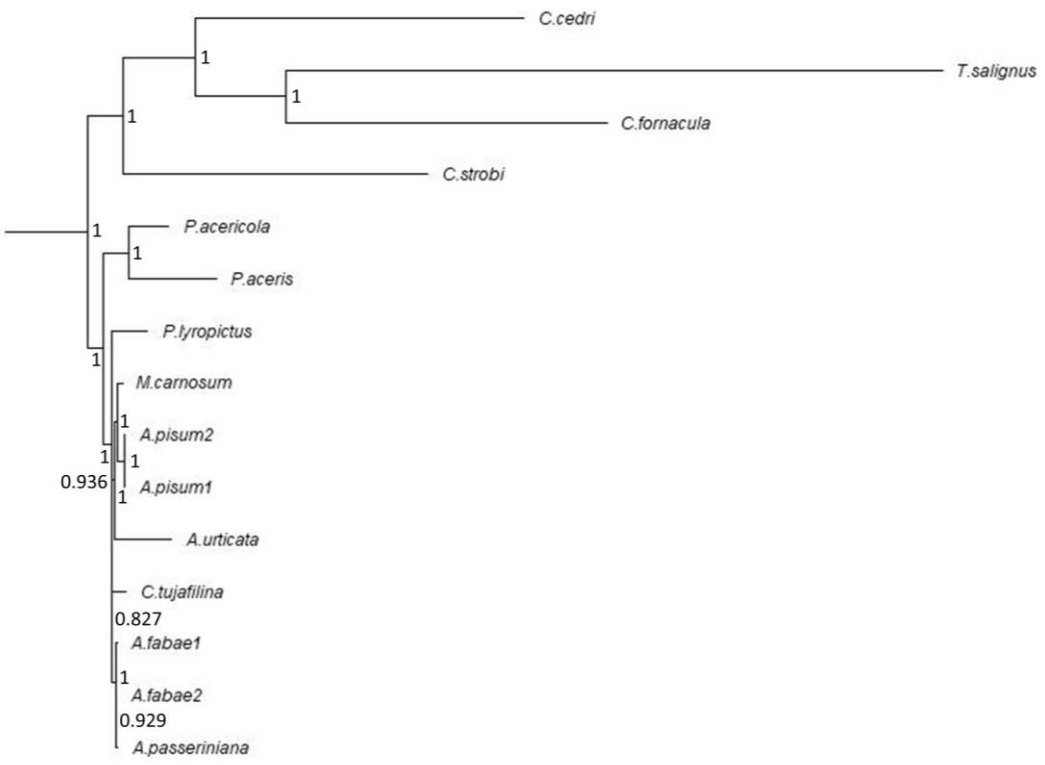


**Figure S4. Phylogeny of *Serratia symbiotica* strains. Related to Figure 2.** Number at the nodes indicate local support values obtained using the Shimodaira-Hasegawa test and based on 1,000 resamples. See Data S1H for the accession numbers of the sequences used.

| **Aphids infected by *Serratia* / Aphids screened** | | |
|---|---|---|
| **Species** | UK aphids | Dutch aphids |
| ***A. urticata*** | 9/9 | 6/6 |
| ***M. carnosum*** | 20/20 | 7/7 |
| ***P. acericola*** | 5/5 | NA |
| ***P. aceris*** | 3/3 | NA |
| ***P. hirticornis*** | 11/11 | 20/20 |
| ***P. lyropictus*** | 9/9 | NA |
| ***P. testudinaceus*** | 70/70 | 16/16 |

**Table S1. Prevalence of *Serratia* in seven focal aphid species. Related to Figure 2.** PCR screening for *Serratia* to confirm ubiquitous infections in the UK and Netherlands.

| Gene Name | Protein Name | **Function** (as reported by UniProt for *Escherichia coli* strain K12) |
|---|---|---|
| **FliN** | Flagellar Motor Switch Protein FliN | FliN is one of three proteins (FliG, FliN, FliM) that form a switch complex that is proposed to be located at the base of the basal body. This complex interacts with the CheY and CheZ chemotaxis proteins, in addition to contacting components of the motor that determine the direction of flagellar rotation. |
| **GrpE** | Protein GrpE | Participates actively in the response to hyperosmotic and heat shock by preventing the aggregation of stress-denatured proteins, in association with DnaK and GrpE. It is the nucleotide exchange factor for DnaK and may function as a thermosensor [S1]. |
| **RsxA** | Ion-translocating oxidoreductase complex subunit A | Part of a membrane-bound complex that couples electron transfer with translocation of ions across the membrane (By similarity) [S2]. |
| **trpS** | Tryptophan--tRNA ligase | Catalyzes the attachment of tryptophan to tRNA(Trp). Amino acylates tRNA(Trp) with both L- and D-tryptophan, although D-tryptophan is a poor substrate [S3]. |
| **YbeY** | Endoribonuclease YbeY | Single strand-specific metallo-endoribonuclease involved in late-stage 70S ribosome quality control and in maturation of the 3' terminus of the 16S rRNA [S4–6]. |
| **YeeX** | UPF0265 protein YeeX | Function unknown. |

**Table S2. Identity and function of all genes that were absent from the *Buchnera* in *A. urticata* that were present in *Buchnera* from *A. pisum*, *M. persicae* and *A. glycines*, where it is the sole obligate symbiont. Related to Figure 3.**

| Compound | Opportunity Cost (High energy-phosphate bonds) ❖ | Cost (µmol ATP per µmol AA)✢ | Yield (µmol AA per µmol Glucose)✢ | Yield (µmol AA per µmol Glycerol)✢ | Yield (µmol AA per µmol Acetate)✢ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|---|---|---|
| Threonine | 18.7 | 2 | 1.368 | 0.778 | 0.37 | ■ | ■ | ■ | ■ |
| Valine | 23.3 | 0 | 1 | 0.5 | 0.25 | ■ | ■ | ■ | |
| Leucine | 27.3 | 0 | 0.67 | 0.33 | 0.2 | ■ | ■ | ■ | ■ |
| Arginine | 27.3 | 5 | 0.929 | 0.5 | 0.27 | ■ | ■ | ■ | ■ |
| Lysine | 30.3 | 2 | 0.839 | 0.484 | 0.22 | ■ | ■ | ■ | ■ |
| Isoleucine | 32.3 | 2 | 0.839 | 0.484 | 0.21 | ■ | ■ | ■ | |
| Methionine | 34.3 | 10 | 0.689 | 0.394 | 0.19 | ■ | ■ | ■ | |
| Histidine | 38.3 | 6 | 0.82 | 0.483 | 0.19 | ■ | ■ | ■ | |
| Phenylalanine | 52 | 1 | 0.571 | 0.3 | 0.14 | ■ | ■ | ■ | |
| Tryptophan | 74.3 | 5 | 0.444 | 0.25 | 0.11 | ■ | ■ | | |

**Table S3. Metabolic cost of essential amino acid synthesis. Related to Figure 3.** ❖Data Source: [S7] Opportunity Cost: Number of high-energy phosphate bonds (contained in ATP, GTP, NADH, NADPH, and FADH2 with an equivalence of 2 P per H) that would have been gained if the metabolite had remained in energy-producing pathways minus the numbers of these molecules gained before diversion. ✢Data Source: [S8] Yield is expressed as µmol of amino acid produced per µmol of carbon source. Cost is expressed as µmol of ATP require to make a µmol of the amino acid. 1= *A. pisum* 2= *M. carnosum* 3= Lachninae subfamily species 4= *Periphyllus* genus species.

## SUPPLEMENTAL REFERENCES


S1. Wu, B., Wawrzynow, A., Zylicz, M., and Georgopoulos, C. (1996). Structure-function analysis of the Escherichia coli GrpE heat shock protein. EMBO J. *15*, 4806–4816.

S2. Koo, M.S., Lee, J.H., Rah, S.Y., Yeo, W.S., Lee, J.W., Lee, K.L., Koh, Y.S., Kang, S.O., and Roe, J.H. (2003). A reducing system of the superoxide sensor SoxR in Escherichia coli. EMBO J. *22*, 2614–2622.

S3. Soutourina, J., Plateau, P., and Blanquet, S. (2000). Metabolism of D-aminoacyl-tRNAs in Escherichia coli and Saccharomyces cerevisiae cells. J. Biol. Chem. *275*, 32535–32542.

S4. Rasouly, A., Davidovich, C., and Ron, E.Z. (2010). The heat shock protein YbeY is required for optimal activity of the 30S ribosomal subunit. J. Bacteriol. *192*, 4592–4596.

S5. Davies, B.W., Köhrer, C., Jacob, A.I., Simmons, L.A., Zhu, J., Aleman, L.M., RajBhandary, U.L., and Walker, G.C. (2010). Role of escherichia coli YbeY, a highly conserved protein, in rRNA processing. Mol. Microbiol. *78*, 506–518.

S6. Jacob, A.I., Köhrer, C., Davies, B.W., RajBhandary, U.L., and Walker, G.C. (2013). Conserved Bacterial RNase YbeY Plays Key Roles in 70S Ribosome Quality Control and 16S rRNA Maturation. Mol. Cell *49*, 427–438.

S7. Akashi, H., and Gojobori, T. (2002). Metabolic efficiency and amino acid composition in the proteomes of Escherichia coli and Bacillus subtilis. Proc. Natl. Acad. Sci. U. S. A. *99*, 3695–3700.

S8. Kaleta, C., Schäuble, S., Rinas, U., and Schuster, S. (2013). Metabolic costs of amino acid and protein production in Escherichia coli. Biotechnol. J. *8*, 1105–1114.