Parallel Histories of Horizontal Gene Transfer Facilitated Extreme Reduction of Endosymbiont Genomes in Sap-Feeding Insects

Daniel B. Sloan,*[1] Atsushi Nakabachi,[2] Stephen Richards,[3] Jiaxin Qu,[3] Shwetha Canchi Murali,[3] Richard A. Gibbs,[3] and Nancy A. Moran[4]

[1]Department of Biology, Colorado State University

[2]Electronics-Inspired Interdisciplinary Research Institute, Toyohashi University of Technology, Toyohashi, Japan

[3]Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas

[4]Section of Integrative Biology, University of Texas

***Corresponding author:** E-mail: dan.sloan@colostate.edu.

Associate editor: Andrew Roger

Abstract

Bacteria confined to intracellular environments experience extensive genome reduction. In extreme cases, insect endosymbionts have evolved genomes that are so gene-poor that they blur the distinction between bacteria and endosymbiotically derived organelles such as mitochondria and plastids. To understand the host's role in this extreme gene loss, we analyzed gene content and expression in the nuclear genome of the psyllid *Pachypsylla venusta*, a sap-feeding insect that harbors an ancient endosymbiont (*Carsonella*) with one of the most reduced bacterial genomes ever identified. *Carsonella* retains many genes required for synthesis of essential amino acids that are scarce in plant sap, but most of these biosynthetic pathways have been disrupted by gene loss. Host genes that are upregulated in psyllid cells housing *Carsonella* appear to compensate for endosymbiont gene losses, resulting in highly integrated metabolic pathways that mirror those observed in other sap-feeding insects. The host contribution to these pathways is mediated by a combination of native eukaryotic genes and bacterial genes that were horizontally transferred from multiple donor lineages early in the evolution of psyllids, including one gene that appears to have been directly acquired from *Carsonella*. By comparing the psyllid genome to a recent analysis of mealybugs, we found that a remarkably similar set of functional pathways have been shaped by independent transfers of bacterial genes to the two hosts. These results show that horizontal gene transfer is an important and recurring mechanism driving coevolution between insects and their bacterial endosymbionts and highlight interesting similarities and contrasts with the evolutionary history of mitochondria and plastids.

Key words: **amino acid biosynthesis, endosymbionts, lateral gene transfer, *Pachypsylla venusta*, psyllids.**

Introduction

One of the most well-documented patterns in genome evolution is the reduction in genome size and gene content exhibited by obligately intracellular bacteria (Andersson and Kurland 1998; Moran and Wernegreen 2000; McCutcheon and Moran 2012). This pattern can largely be explained by a combination of relaxed functional constraint on redundant or superfluous genes (Andersson et al. 1998), inefficient purifying selection in a host-restricted environment (Moran 1996), and a mutational bias favoring deletions over insertions (Mira et al. 2001; Kuo and Ochman 2009). However, extreme cases of genome reduction involving the loss of seemingly essential genes remain puzzling. In particular, many nutritional endosymbionts in sap-feeding insects maintain genomes that are less than 200 kb in size and that have lost genes thought to be essential for basic cellular processes, including the primary symbiotic role of these bacteria—the biosynthesis of amino acids that are lacking in their hosts' diets and that cannot be made by animals (Nakabachi et al. 2006; McCutcheon et al. 2009b; McCutcheon and Moran 2010; McCutcheon and von Dohlen 2011; Bennett and Moran 2013). The first example of such extreme genome reduction was found in *Candidatus* Carsonella ruddii (hereafter referred to as *Carsonella*), a vertically transmitted gammaproteobacterial endosymbiont that is present in all psyllids (Hemiptera: Sternorrhyncha) examined to date (Thao et al. 2000). Gene loss appears to be an ongoing process in this endosymbiont with observed *Carsonella* genome sizes ranging in size from 158 to 174 kb across different host species (Nakabachi et al. 2006, 2013; Sloan and Moran 2012). These tiny genomes lack many widely conserved genes involved in DNA replication, transcription, and translation, and most of their essential amino acid biosynthesis pathways are incomplete (Tamames et al. 2007). Similar to other nutritional endosymbionts in sap-feeding insects, *Carsonella* is sequestered within bacteriocytes, specialized host cells found in an abdominal organ known as the bacteriome (Buchner 1965; Fukatsu and Nikoh 1998).

Article

The exceptional gene loss observed in *Carsonella* suggests three possible mechanisms (none of which are mutually exclusive). First, modification of highly conserved cellular processes or selection for multifunctional proteins could have allowed the endosymbiont to dispense with otherwise "essential" genes (McCutcheon 2010; Kelkar and Ochman 2013). Second, as commonly observed in sap-feeding insects, the presence of additional endosymbionts may compensate for gene losses (Wu et al. 2006; McCutcheon et al. 2009a; McCutcheon and Moran 2010; Lamelas et al. 2011; McCutcheon and von Dohlen 2011; Nakabachi et al. 2013). There is evidence supporting a role for other endosymbionts in some psyllids. For example, the loss of *Carsonella* genes necessary for arginine and tryptophan biosynthesis in the *Eucalyptus*-feeding psyllid *Ctenarytaina eucalypti* appears to be offset by a second bacterial endosymbiont that retains these genes (Sloan and Moran 2012). However, some psyllids, including hackberry-feeding species in the genus *Pachypsylla*, lack additional symbiotic partners (Thao et al. 2000) but still maintain *Carsonella* strains with highly reduced genomes that have lost genes required for synthesis of some essential amino acids (Nakabachi et al. 2006). Finally, missing *Carsonella* genes may have been functionally replaced by host-encoded proteins, including pre-existing eukaryotic homologs (Wilson et al. 2010; Price et al. 2011) and/or genes of bacterial origin acquired by horizontal gene transfer (HGT) (e.g., Nowack and Grossman 2012).

Here, we focus on the last of these three potential mechanisms to address the question of whether *Carsonella* has followed a similar trajectory as the evolution of mitochondria and plastids from bacterial progenitors, in which genetic control was largely shifted from the endosymbiont (organelle) to the host (nuclear) genome through a combination of direct gene transfer and functional gene replacement. Recent functional genomic analyses in two sap-feeding insects, aphids and mealybugs, have supported the prediction that host genomes play a central role in endosymbiont metabolism and provided evidence that both insects have acquired multiple genes of bacterial origin, many of which are now preferentially expressed in bacteriocytes (Nakabachi et al. 2005; Nikoh and Nakabachi 2009; Nikoh et al. 2010; Hansen and Moran 2011; Poliakov et al. 2011; Macdonald et al. 2012; Husnik et al. 2013). The pea aphid endosymbiont *Buchnera aphidicola* has maintained a (relatively) intact genome with a 641 kb chromosome containing almost 600 protein-coding genes (Shigenobu et al. 2000), and none of the horizontally acquired bacterial genes in the pea aphid genome appear to be involved in core metabolic pathways in *Buchnera* (Nikoh et al. 2010). In contrast, the betaproteobacterial endosymbiont *Ca.* Tremblaya princeps from the citrus mealybug has a tiny genome (139 kb) even by the standards of obligately intracellular bacteria, and it has the notable distinction of harboring another bacterial endosymbiont nested within its own cells (von Dohlen et al. 2001; McCutcheon and von Dohlen 2011). Most of the inferred metabolic pathways in the citrus mealybug bacteriome involve a patchwork of gene products encoded in different symbiotic compartments, including more than 20 genes of bacterial origin in the host genome, many of which appear to directly complement *Tremblaya* gene losses (Husnik et al. 2013). Interestingly, there is no evidence that any of the HGT events in aphids or mealybugs involve the direct transfer of functional genes from their respective obligate endosymbionts, contrasting with the massive gene movement from mitochondria and plastids to the nucleus during the history of eukaryotic genome evolution (Timmis et al. 2004). Instead, genes appear to have been acquired from other insect-associated bacteria, such as *Arsenophonus*, *Cardinium*, *Rickettsia*, *Sodalis*, *Serratia*, and *Wolbachia* (Nikoh and Nakabachi 2009; Nikoh et al. 2010; Husnik et al. 2013). However, the pea aphid genome does contain two small, nonfunctional fragments of *Buchnera* DNA, indicating that genetic transfer to insects from their obligate endosymbionts is possible (Nikoh et al. 2010).

By analyzing genome-wide patterns of host gene expression in the hackberry petiole gall psyllid *Pachypsylla venusta*, we show that extreme genome reduction in *Carsonella* and *Tremblaya* has involved a number of striking evolutionary parallels, including independent HGT events affecting the same functional pathways. Although multiple bacterial lineages have acted as HGT donors, we find that psyllids have acquired at least one gene directly from *Carsonella*.

Results

Identification of Preferentially Expressed Genes in the Psyllid Bacteriome by mRNA-Seq

Deep strand-specific sequencing of polyadenylated transcripts (mRNA-seq) in *P. venusta* nymphs isolated from hackberry galls was used to broadly identify and characterize psyllid genes. Furthermore, by comparing expression levels between the bacteriome and the rest of the insect body, we identified dramatic differences in host gene expression (fig. 1). After quality filtering, we obtained between 14.3 and 20.9 million Illumina read pairs for each of three replicate mRNA-seq libraries derived from isolated *P. venusta* bacteriomes and three corresponding libraries derived from the remaining body tissues (table 1). Despite the polyA selection step employed during library construction to isolate eukaryotic mRNAs, we also recovered some *Carsonella* sequences (an average of 7.75% and 0.05% of reads in bacteriome and body libraries, respectively), reflecting the extremely AT-rich nucleotide composition of the *Carsonella* genome (Nakabachi et al. 2006). After excluding reads with identical or near-identical matches to the *Carsonella* genome (see Materials and Methods), de novo assembly of the combined sequence data from all six libraries generated 142,684 transcripts that were grouped into 80,904 "subcomponents." In principle, these subcomponents can be interpreted as distinct genetic loci, but transcriptome assembly artifacts and assembly fragmentation can inflate the identified number of subcomponents. Gene expression estimates (measured in transcripts per million) were highly repeatable among biological replicates, with sets of bacteriome and body libraries both exhibiting intraclass correlation coefficients of 0.98. A total of 11,947 transcripts (8.4%) were identified as differentially expressed between the bacteriome and the rest of the body

with a false discovery rate of less than 0.001 (of which 6,848 were upregulated in the bacteriome and 5,099 were downregulated). More than half of these transcripts exhibited a difference in expression of greater than 100-fold between the two tissue categories.

Integration between Host and Endosymbiont Metabolic Pathways Facilitates Amino Acid Biosynthesis in Psyllids

Previous analyses have summarized gene content in the *Carsonella* genome from *P. venusta* (Nakabachi et al. 2006; Tamames et al. 2007). Although these studies found that the *Carsonella* genome is highly enriched for genes involved in amino acid biosynthesis, many pathways appear to be incomplete or absent altogether (summarized in fig. 2).

Host genes that were upregulated in the bacteriome appear to complement many of the pathways that are missing or incomplete as a result of gene loss in *Carsonella* (Nakabachi et al. 2006), confirming that the bacteriome is highly active in the synthesis of both essential and nonessential amino acids (fig. 2 and table 2). The overall pattern of gene expression in the psyllid bacteriome exhibited numerous similarities to that in bacteriomes of related sap-feeding insects (Hansen and Moran 2011; Macdonald et al. 2012; Husnik et al. 2013). Similar to aphids and mealybugs, the psyllid genome contains genes encoding glutamine synthetase (GS, EC 6.3.1.2) and glutamine oxoglutarate aminotransferase (GOGAT, EC 1.4.1.13), both of which were upregulated in the bacteriome, with GS being expressed at exceptionally high levels. The GS/GOGAT cycle provides a means of recycling free ammonia (NH_3) into glutamine and glutamate, which are used as amino group donors in many of the amino acid biosynthesis pathways. The ability to recycle ammonia is presumably important for sap-feeding insects, given the limits on nitrogen quantity and quality in their diets (Lamb 1959; Douglas 2006). A few nonessential amino acids including asparagine often represent the main source of nitrogen in phloem sap (Fukumorita and Chino 1982; Weibull et al. 1986; Sandström and Pettersson 1994). Therefore, the activity of asparaginase (EC 3.5.1.1), which was highly upregulated in the psyllid bacteriome, is likely a primary source of ammonia for the GS/GOGAT cycle (Macdonald et al. 2012; Husnik et al. 2013).

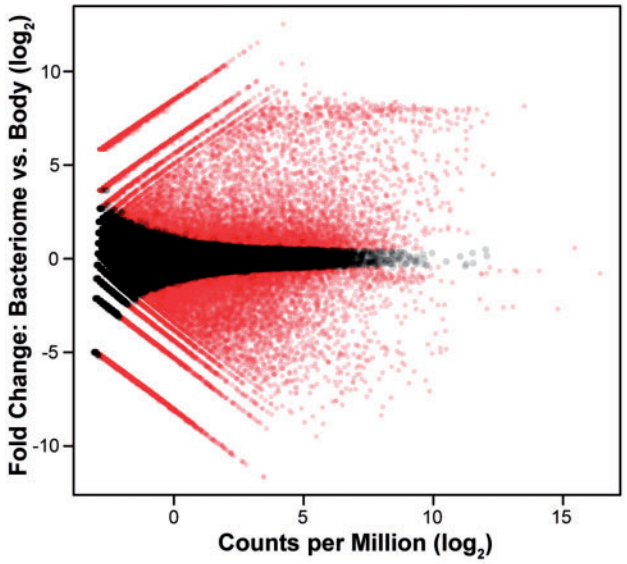

FIG. 1. Differential gene expression between psyllid bacteriome and remaining body tissues. Each point represents a Trinity subcomponent ("gene"), with the *x* axis indicating overall gene expression and the *y* axis indicating differential expression between tissue types. Genes identified by edgeR as being significantly upregulated or downregulated in the bacteriome are in red.

Similar to the endosymbionts of many phloem-feeding insects (Hansen and Moran 2013), *Carsonella* has lost the genes necessary for production of ornithine, a substrate in the arginine biosynthesis pathway, and for catalyzing the terminal transamination reaction in the synthesis of phenylalanine (*aspC*) (Nakabachi et al. 2006). Two psyllid genes that were highly expressed in the bacteriome, encoding delta-1-pyrroline-5-carboxylate synthase (P5CS, EC 2.7.2.11/1.2.1.41) and ornithine aminotransferase (OAT, EC 2.6.1.13), potentially provide an alternative pathway to synthesize ornithine from glutamate. Similarly, the psyllid genome contains an aspartate aminotransferase (AAT, EC 2.6.1.1) gene that was highly upregulated in the bacteriome and likely acts as a replacement for *aspC*. In both cases, these observations mirror expression patterns in mealybugs and/or aphids (Hansen and Moran 2011; Macdonald et al. 2012; Husnik et al. 2013). Psyllid genes involved in biosynthesis of the nonessential amino acids proline, serine, and tyrosine were also highly expressed and upregulated in the bacteriome (fig. 2 and table 2).

Although, in most cases, the apparent complementarity between the psyllid and *Carsonella* genomes provided a clear hypothesis for reconstructing complete amino acid biosynthesis pathways (fig. 2), a handful of *Carsonella* gene losses

Table 1. Summary Statistics for mRNA-Seq Libraries.

Replicate	Tissue Type	Illumina Read Pairs (Millions)			*Carsonella* Library Contamination (%)
		Raw Sequence	Quality Filtered	*Carsonella* Filtered	
1	Bacteriome	22.3	14.7	13.5	7.64
	Body	31.8	20.9	20.9	0.08
2	Bacteriome	21.5	14.3	13.2	7.66
	Body	28.5	18.5	18.5	0.03
3	Bacteriome	29.3	19.3	17.8	7.94
	Body	31.6	20.9	20.9	0.04

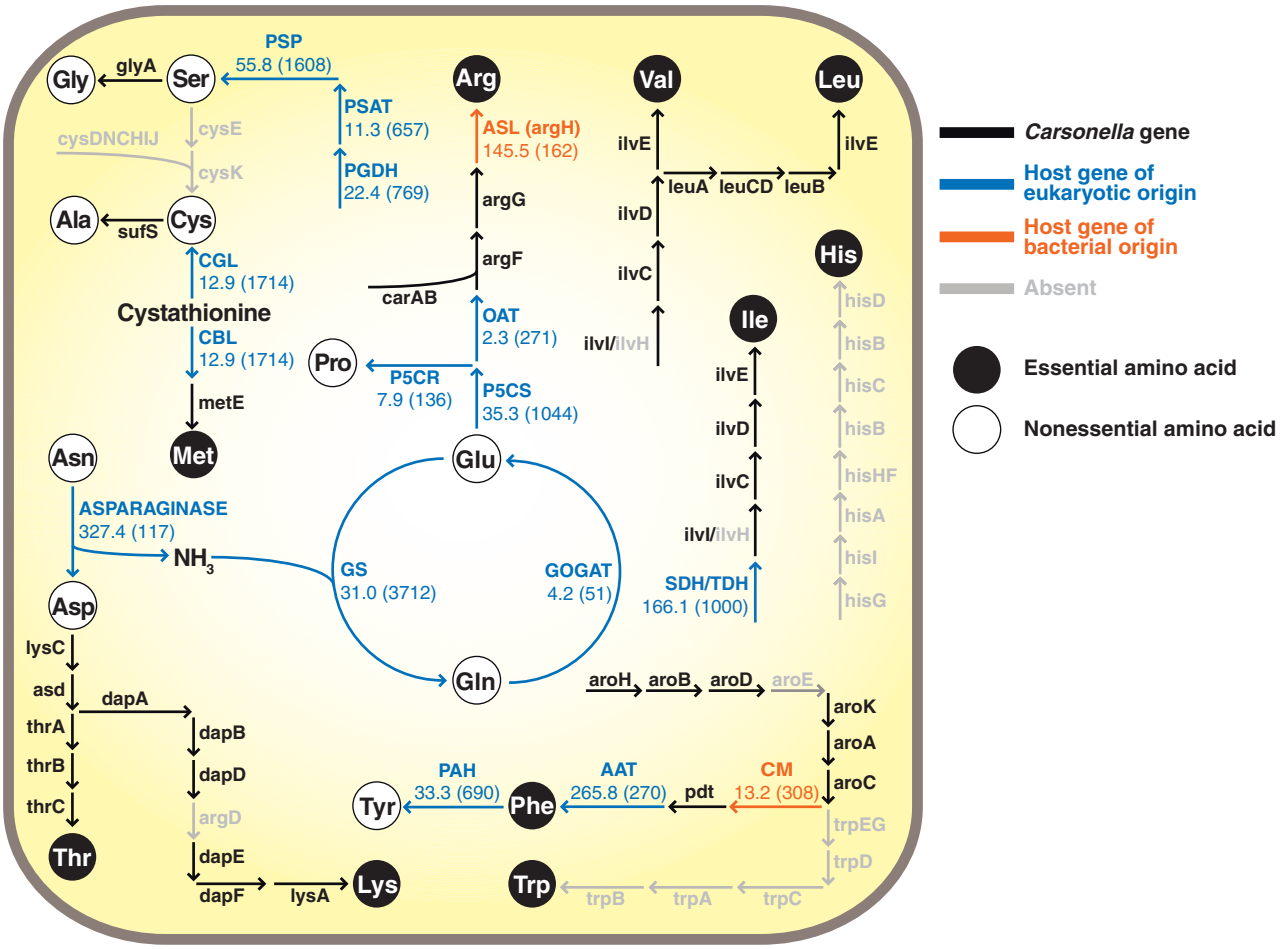

FIG. 2. Inferred amino acid biosynthesis pathways in the *Pachypsylla venusta* bacteriome. For each host-encoded gene, the fold increase in expression relative to the rest of the body is shown with absolute bacteriome expression in transcripts per million (TPM) shown in parentheses.

remain difficult to explain. The most conspicuous involve the absence of the entire histidine and tryptophan pathways. These losses must have occurred since the divergence of *Pachypsylla* from other psyllid lineages because both pathways are retained in other hosts (Sloan and Moran 2012; Nakabachi et al. 2013). We found no evidence that any of the corresponding genes have been transferred to the host genome or functionally replaced by insect genes, suggesting that the gall-forming *Pachypsylla* species must acquire sufficient levels of histidine and tryptophan from their diet. *Carsonella* also lacks scattered genes in other amino acid pathways, including *argD* (lysine), *aroE* (phenylalanine/tyrosine), and *ilvH* (branched-chain amino acids). However, in each of these cases, the rest of the pathway is widely conserved in *Carsonella* strains from other psyllids, suggesting that gene loss may have been offset by other enzymes—either from the host or the endosymbiont—that have yet to be identified. This interpretation is supported by the occurrence of similar or identical gene losses in endosymbionts from related sap-feeding insects (Hansen and Moran 2013).

The metabolic pathways related to cysteine and methionine also remain uncertain. *Carsonella* retains an enzyme that catalyzes the conversion of homocysteine into methionine (*metE*) but lacks the rest of the methionine biosynthesis pathway. Similar to aphids and mealybugs, the psyllid genome contains a gene related to cystathionine gamma- and beta-lyases (CGL and CBL, EC 4.4.1.1/4.4.1.8) that was very highly expressed in the bacteriome (fig. 2 and table 2). It has been hypothesized that this gene may catalyze consecutive steps in converting cysteine to homocysteine via a cystathionine intermediate, providing the necessary substrate for endosymbiont-mediated synthesis of methionine (Wilson et al. 2010; Hansen and Moran 2011; Husnik et al. 2013). However, although recent metabolic analysis confirmed that aphid host cells can generate homocysteine from cystathionine, there was no evidence that they convert cysteine into cystathionine (Russell et al. 2013). Psyllids also differ from aphids and mealybugs in an important respect. Although the enzymatic machinery necessary to reduce sulfate and synthesize cysteine from serine is still present (in symbiont or host) in both aphids and mealybugs, *Carsonella* has lost all genes in this pathway, and we found no evidence that they have been functionally replaced by host genes. For these reasons, it is unlikely that methionine is synthesized from cysteine in psyllid bacteriomes. Instead, methionine biosynthesis likely depends on the availability of cystathionine. Interestingly, this compound may also be used as a substrate for CGL/CBL to produce cysteine

Table 2. Host-Encoded Amino Acid Biosynthesis Genes that Are Upregulated in the Bacteriome.

Enzyme	Name	EC	Trinity Subcomponent	Bacteriome Expression		
				TPM (FPKM)[a]	Fold Change[b]	FDR
Asparaginase		3.5.1.1	comp113172_c2	118 (93)	327.4	2E-151
Aspartate aminotransferase	AAT	2.6.1.1	comp115799_c6	270 (165)	265.8	2E-124
Serine/threonine dehydratase	SDH/TDH	4.3.1.17/4.3.1.19	comp108647_c2	1000 (788)	166.1	1E-144
Argininosuccinate lyase	ASL-1[c]	4.3.2.1	comp116240_c1	162 (127)	145.5	6E-95
Argininosuccinate lyase	ASL-2	4.3.2.1	comp113612_c8	38 (30)	141.8	4E-87
Phosphoserine phosphatase	PSP	3.1.3.3	comp107672_c1	1608 (1266)	55.8	6E-108
Delta-1-pyrroline-5-carboxylate synthase	P5CS	2.7.2.11/1.2.1.41	comp116236_c0	1044 (822)	35.3	1E-77
Phenylalanine 4-monooxygenase	PAH	1.14.16.1	comp115765_c0	690 (542)	33.3	5E-75
Glutamine synthetase	GS	6.3.1.2	comp104503_c0	3712 (2923)	31.0	3E-102
Phosphoglycerate dehydrogenase	PGDH	1.1.1.95	comp106096_c1	769 (605)	22.4	7E-63
Chorismate mutase	CM	5.4.99.5	comp113157_c6	308 (242)	13.2	2E-64
Cystathionine gamma/beta lyase	CGL/CBL	4.4.1.1/4.4.1.8	comp111121_c10	1714 (1350)	12.9	2E-64
Phosphoserine aminotransferase	PSAT	2.6.1.52	comp111308_c2	657 (517)	11.3	1E-57
Pyrroline-5-carboxylate reductase[d]	P5CR	1.5.1.2	comp115632_c0	135 (107)	7.9	1E-35
Glutamine oxoglutarate aminotransferase	GOGAT	1.4.1.13	comp114372_c4	51 (40)	4.2	6E-19
Ornithine aminotransferase	OAT	2.6.1.13	comp110483_c0	271 (213)	2.3	1E-07

NOTE.—FDR, false discovery rate; TPM, transcripts per million.

[a]TPM is used throughout as a preferred measure of gene expression (Li et al. 2010; Wagner et al. 2012), but, for comparative purposes, the commonly used FPKM statistic (fragments per kb per million fragments mapped) is reported in parentheses.

[b]Fold change is measured relative to expression in the rest of the body.

[c]The ASL-1 locus appears to have been broken into two contigs by the Trinity assembler. Most of the gene is found in comp116240_c1, but the 3′-end is in comp106522_c3.

[d]The *P. venusta* genome contains a second copy of P5CR (comp111126_c4: sequences 7, 8, 9, and 14), which was very highly expressed in the bacteriome. However, edgeR did not identify it as upregulated because of the expression data from additional sequences that were included within the same Trinity subcomponent.

(fig. 2), but the ultimate source of cystathionine is not readily apparent. There are also alternative hypotheses to consider, for example, that sap-feeding insects obtain cysteine and/or methionine by processing dietary glutathione and S-methylmethionine, respectively, which are often found in plant phloem (Akman Gündüz and Douglas 2009; Wilson et al. 2010).

The activity of CGL/CBL may also play a role in isoleucine biosynthesis by providing a source of the precursor 2-oxobutanoate from homoserine and/or cystathionine (Poliakov et al. 2011; Husnik et al. 2013; Russell et al. 2013). If this is the case, it could compensate for the fact that the endosymbionts of mealybugs, aphids, and psyllids have all lost the threonine dehydratase (TDH, EC 4.3.1.19) encoded by *ilvA*, which catalyzes the production of 2-oxobutanoate from threonine. Previous analyses had suggested that the loss of *ilvA* could be offset by the presence of a host-encoded TDH, but this hypothesis was undermined by findings that the native insect gene is not upregulated in the bacteriome in either aphids or mealybugs (Hansen and Moran 2011; Husnik et al. 2013). We found that the orthologous TDH gene in *P. venusta* was modestly downregulated (6-fold) in the bacteriome relative to the rest of the body, but we also identified a second gene with strong homology to serine/threonine dehydratases (SDH/TDH, EC 4.3.1.17/4.3.1.19) that is not present in aphids or mealybugs. Although this gene has not been previously identified in insects, it is present in other animal lineages (supplementary fig. S1, Supplementary Material online), suggesting that it has been independently lost from multiple insect lineages. This SDH/TDH gene was highly expressed and upregulated (116-fold) in the psyllid bacteriome, suggesting that it may functionally replace *ilvA* in the isoleucine biosynthesis pathway (fig. 2 and table 2). It is also possible that the enzyme uses serine as a substrate, which could provide an additional source of pyruvate as input for the leucine/valine pathway. Either reaction would contribute to the pool of free ammonia available for the GS/GOGAT cycle.

Although the inferred metabolic pathways in the psyllid bacteriome resemble those in related sap-feeding insects, there are some clear contrasts, such as the loss of cysteine biosynthesis genes and the presence of an additional SDH/TDH gene (see earlier). In addition, *ilvE*, which encodes a branched-chain aminotransferase (BCAT, EC 2.6.1.42), has been lost from the endosymbiont genomes in aphids and mealybugs. This enzyme catalyzes the terminal step in the synthesis of isoleucine, leucine, and valine. In both aphids and mealybugs, a host BCAT gene is upregulated in the bacteriome, likely compensating for the loss of *ilvE* in the endosymbionts. This functional replacement may enable host regulatory control of the branched-chain amino acid biosynthesis pathways (Wilson et al. 2010). In *Carsonella*, however, *ilvE* is retained, and the overall expression level of the host-encoded BCAT gene was essentially unchanged in the bacteriome relative to the rest of the psyllid body (<2-fold difference). Therefore, control over the terminal step in branched-chain amino acid synthesis appears to be retained by *Carsonella*. Interestingly, however, the length distributions of host BCAT transcripts differed between the bacteriome and the rest of the body. We found that antisense expression

at the 5′-end of the host BCAT coding sequence was upregulated more than 200-fold in the bacteriome (supplementary fig. S2, Supplementary Material online). The amount of sense-strand transcripts from the corresponding region was highly reduced in the bacteriome, whereas the rest of the gene was expressed at a slightly higher level in the bacteriome than the insect body (supplementary fig. S2, Supplementary Material online). The functional consequences of antisense expression and variable transcript lengths in BCAT are not clear, but this example illustrates the value of strand-specific expression data. Conventional RNA-seq data would likely have led to the misinterpretation that the BCAT gene is highly upregulated in the psyllid bacteriome.

The *P. venusta* Genome Contains Many Functional Genes of Bacterial Origin

Most of the host genes predicted to compensate for genome reduction in *Carsonella* are native animal genes that appear to have been co-opted to mediate this symbiotic relationship. However, two genes in the inferred amino acid biosynthesis pathways do not have identifiable orthologs in other insects and appear to have been acquired by HGT (fig. 2). The first of these is the gene coding for chorismate mutase (CM, EC 5.4.99.5), which catalyzes a key step in the phenylalanine biosynthesis pathway. Bacteria typically contain a bifunctional CM-prephenate dehydratase enzyme (PheA, EC 5.4.99.5/4.2.1.51) that catalyzes the two successive reactions that convert chorismate into the immediate precursor of phenylalanine. The *pheA* gene is retained in *Carsonella*, but it has lost the CM domain. We identified a CM gene that is encoded in the *P. venusta* genome and highly expressed in the bacteriome—a clear case of HGT because CM is not normally found in animals. Interestingly, this psyllid gene does not appear to be derived from *pheA* and instead belongs to a distinct family of monofunctional CMs with a scattered distribution in bacteria (TIGR01806, CM_mono2). The only other documented cases of CMs in animals (found in phytoparasitic nematodes) also involve bacterial HGT from this monofunctional group (Lambert et al. 1999; Jones et al. 2003). Second, the *P. venusta* genome contains two divergent copies of a gene encoding argininosuccinate lyase (ASL, EC 4.3.2.1), which is responsible for the terminal step in the arginine biosynthesis pathway. Although this enzyme is present in many insects, it has not been identified previously in Hemiptera, and Blast searches found that the *P. venusta* ASLs are most similar to the bacterial ArgH. Quantitative polymerase chain reaction (qPCR) using cDNA from multiple tissues confirmed that the expression of both ASL gene copies is highly upregulated in the *P. venusta* bacteriome (fig. 3). In contrast, qPCR amplification of genomic DNA found that the genomic copy number was essentially constant across tissues (fig. 3), as expected for genes in the insect genome.

To determine whether these cases were representative of a broader history of HGT from bacteria to psyllids, we screened all assembled *P. venusta* transcripts for evidence of a bacterial origin. After excluding cases of probable bacterial contamination in our sequencing libraries (see Materials and Methods),

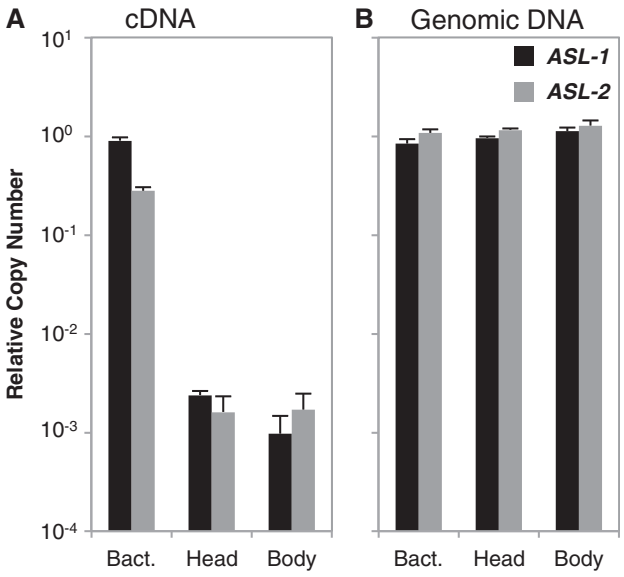

Fig. 3. Quantitative PCR analysis of ASL gene expression (cDNA) and genomic copy number (genomic DNA) in bacteriome, head, and body tissues from *Pachypsylla venusta*. Abundance values were normalized to the psyllid nuclear gene encoding the ribosomal protein (Rp)L18e. Error bars represent one standard error.

we identified seven additional examples of bacterial HGT (table 3). The functional pathways affected by HGT in *P. venusta* mirror earlier observations in the citrus mealybug (Husnik et al. 2013), but the specific genes that were transferred differ between the two insects (table 4). The parallel history of these HGT events suggests that the transferred genes have an important function, but, in some cases, their specific roles are not immediately obvious. For example, *P. venusta* has acquired the gene coding for riboflavin synthase (*ribC*), but we found it to be expressed at very low levels. Furthermore, the genomes of both the psyllid and *Carsonella* appear to lack the rest of the genes in the riboflavin biosynthesis pathway. The findings from a recent genomic analysis of endosymbionts in the Asian citrus psyllid *Diaphorina citri* may provide some insight into the history of the *ribC* transfer (Nakabachi et al. 2013). In addition to *Carsonella*, *D. citri* harbors a second bacteriome-associated endosymbiont (*Ca.* Profftella armatura). The highly reduced *Profftella* genome retains all the genes in the riboflavin biosynthesis pathway with the exception of *ribC*. We found that *ribC* was transferred to the psyllid genome prior to the divergence between *Pachypsylla* and *Diaphorina* and that it is still present in *D. citri* (see later). Therefore, the *ribC* transfer may reflect an ancient complementarity in riboflavin biosynthesis with an endosymbiont that has since been lost in the *Pachyspylla* lineage.

Most identified genes of bacterial origin in the *P. venusta* genome were preferentially expressed in the bacteriome, often by more than an order of magnitude (table 3), suggesting that they play important roles in mediating the symbiosis with *Carsonella*. However, two genes exhibited modestly higher expression in the body. One of these, *ydcJ*, is a conserved bacterial gene, but its function remains unclear even in

Table 3. Bacterial Genes in the Psyllid Genome.

Name	Description	Trinity Subcomponent	Bacteriome Expression			Top Blast Hit
			TPM (FPKM)	Fold Change	FDR	
ORF	AAA-ATPase-like	comp111207_c2	235 (185)	168.5	5.E-124	ZP_04700017: *Rickettsia* (Alphaproteobacteria)
RSMJ	16S rRNA methyltransferase	comp107791_c0	23 (18)	152.8	2.E-66	YP_003019747: *Pectobacterium carotovorum* (Gammaproteobacteria)
ASL-1	Argininosuccinate lyase, *argH*-like	comp116240_c1	162 (128)	145.5	6.E-95	ZP_09288859: *Halomonas sp. GFAJ-1* (Gammaproteobacteria)
ASL-2	Argininosuccinate lyase, *argH*-like	comp113612_c8	38 (30)	141.8	4.E-87	ZP_09187732: *Halomonas boliviensis* (Gammaproteobacteria)
ORF	AAA-ATPase-like	comp115390_c0	127 (100)	113.0	1.E-113	ZP_04700042: *Rickettsia* (Alphaproteobacteria)
CM	Chorismate mutase	comp113157_c6	308 (242)	13.2	2.E-64	ZP_11190527: *Pseudomonas sp. R81* (Gammaproteobacteria)
RIBC	Riboflavin synthase	comp97442_c1	2 (2)	4.2	4.E-04	YP_001453245: *Citrobacter koseri* (Gammaproteobacteria)
ORF	Ankyrin repeat domain protein	comp113863_c0	10 (8)	2.7	2.E-07	ZP_03335207: *Wolbachia* (Alphaproteobacteria)
MUTY	A/G-specific adenine glycosylase	comp106533_c0	2 (2)	−2.2	1.E-03	ZP_11310983: *Bacillus bataviensis* (Firmicutes)
YDCJ	Conserved hypothetical protein	comp111380_c3	14 (11)	−3.0	1.E-11	YP_001237396: *Bradyrhizobium sp. BTAi1* (Alphaproteobacteria)

NOTE.—FDR, false discovery rate; TPM, transcripts per million.

Table 4. Bacterial Genes Transferred to Psyllid and Mealybug Genomes.

Functional Classification	*Pachypsylla venusta* (Psyllid)	*Planococcus citri* (Mealybug)
Amino acid synthesis/metabolism	*argH, cm*	*cysK, dapF, lysA, tms1*
rRNA methyltransferase	*rsmJ*	*rlmI*
Riboflavin synthesis	*ribC*	*ribA, ribD*
AAA-ATPase	Two genes	One gene
Ankyrin repeat domain protein	One gene	One gene
Peptidoglycan synthesis/metabolism	—	*amiD, ddlB, mltB, mraY, murABCDEFG*
Biotin synthesis	—	*bioA, bioB, bioD*
Other	*ydcJ*	Glutamate-cysteine ligase-like protein, type III effector, urea amidolyase

well-studied bacterial model systems. The other, *mutY*, encodes a DNA mismatch repair enzyme that prevents G:C to T:A transversions caused by mispairings involving the oxidatively damaged base 8-oxo-G. Although *mutY* homologs are widespread in eukaryotes, they have been lost in insects (Jansson et al. 2010). Therefore, *P. venusta* may have reacquired the ability to perform *mutY*-dependent repair of 8-oxo-G damage in its genome by means of HGT, although we found that this gene was only expressed at low levels (table 3).

Draft Genome Assembly of *P. venusta* Confirms Presence of Bacterial Genes

Data from the *P. venusta* genome project were used to verify that genes of interest identified by mRNA-seq are actually located in the psyllid genome. Sequencing of *P. venusta* genomic DNA produced an initial draft assembly with a total contig length of 371.9 Mb and a total scaffold length of 701.8 Mb (including estimated gaps), which is consistent with an earlier genome size estimate of 724 Mb for this species

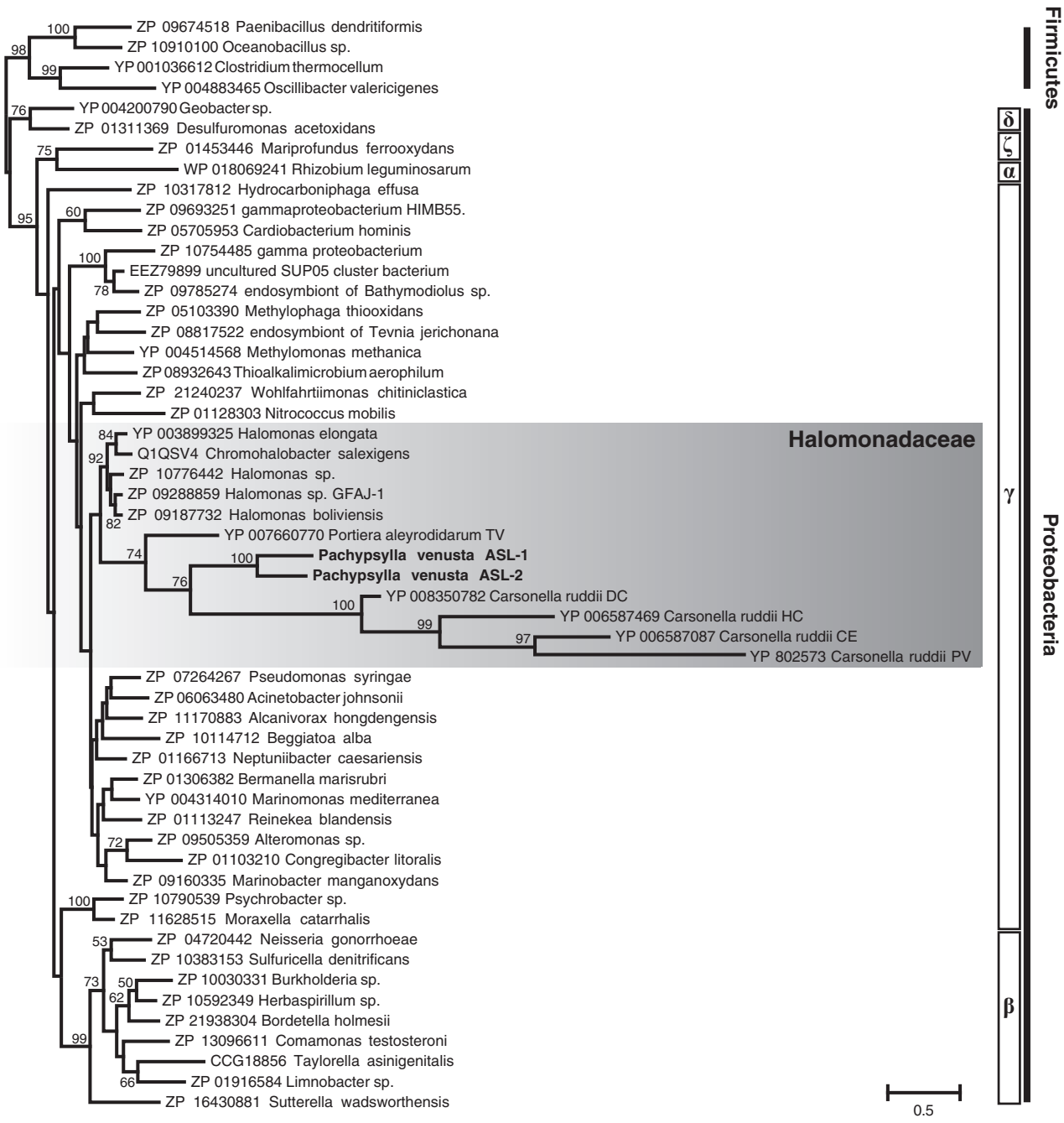

Fig. 4. Phylogenetic analysis of *argH*-derived ASL genes in *Pachypsylla venusta*. Bootstrap values are indicated for nodes with ≥50% support. The terminal branch for *Carsonella ruddii* PV was scaled to half its length to improve readability.

(Nakabachi et al. 2010). Contig and scaffold N50s for the assembly were 1.85 and 157.7 kb, respectively. The assembly confirmed that the putative genes of bacterial origin are encoded in the insect genome. Despite the fact that insects were sampled from geographically disparate locations for genome and transcriptome sequencing (Tucson, AZ, and Austin, TX, respectively), the HGT candidates from the transcriptome (table 3) were all verified to be present in the draft genome assembly. They also contained introns with typical insect sequence motifs at splice sites (Mount et al. 1992) and were found in large scaffolds (21–1,353 kb) that contained characteristic insect genes and/or repetitive elements.

Genes of Bacterial Origin Were Acquired from Diverse Donor Lineages Including *Carsonella* Prior to the Major Diversification of Psyllids

Blast and phylogenetic analyses both indicated that diverse bacterial lineages acted as sources of HGT into the psyllid genome (table 3 and fig. 3; supplementary fig. S3, Supplementary Material online). As observed in aphids and mealybugs (Nikoh and Nakabachi 2009; Nikoh et al. 2010; Husnik et al. 2013), many genes appear to be derived from groups with well-known insect-associated bacteria, including *Rickettsia* and *Wolbachia*. Despite substantial amino acid

sequence divergence (35%), the two *argH*-derived ASL genes in the *P. venusta* genome appear to be closely related and, therefore, the result of repeated transfers from the same source or a single transfer and subsequent duplication within the psyllid genome (fig. 4). Blast searches against the National Center for Biotechnology Information (NCBI) nr database found that these genes are most similar to homologs within the Halomonadaceae, which includes *Carsonella* and its sister lineage *Ca.* Portiera aleyrodidarum, the obligate endosymbiont of whiteflies (Spaulding and von Dohlen 1998; Sloan and Moran 2013). Unlike most bacterial genes in the *P.venusta* genome, a copy of *argH* is still present in the *Carsonella* genome (Nakabachi et al. 2006), and phylogenetic analysis placed the host-encoded genes in a monophyletic group with *Carsonella*, providing strong evidence that they were acquired directly from *Carsonella* (fig. 4). Although the branch lengths are long for the psyllid/endosymbiont sequences and the statistical support for some of the phylogenetic relationships is weak, it is unlikely that the placement of these sequences within the Halomonadaceae is a long-branch artifact given the divergent nucleotide compositions of the genomes involved. Although the psyllid and endosymbiont sequences are very AT-rich, the rest of the genomes in this family are GC biased. None of the other examples of HGT were clearly derived from *Carsonella* or the Halomonadaceae, although the high level of sequence divergence in transferred genes precluded unambiguous taxonomic assignment in some cases.

To assess the timing of bacterial HGT events in psyllids, we searched available genomic/transcriptomic data from two other psyllids, the Asian citrus psyllid (*D. citri*) and the potato psyllid (*Bactericera cockerelli*). With a single exception, the identified genes of bacterial origin in the *P. venusta* genome are also present in both *D. citri* and *B. cockerelli*. Only the *Wolbachia*-derived ankyrin repeat domain protein does not appear to be shared with the other psyllids. Because the divergences between these species span some of the deepest splits in the psyllid phylogeny (Thao et al. 2000), the identified HGT events must have occurred before the radiation of the major psyllid lineages. However, none of the transfers appear to be shared with aphids, mealybugs, or whiteflies, indicating that they took place after the divergence of the superfamilies within the suborder Sternorrhyncha. Therefore, the establishment of *Carsonella* in a common ancestor of psyllids roughly 150 to 200 Mya (Rasnitsyn and Quicke 2002) seems to have coincided with a burst of bacterial HGT events.

Discussion

Parallelism in the Evolution of Metabolic Systems in Sap-Feeding Insects

The metabolic interactions between the psyllid *P. venusta* and its obligate endosymbiont *Carsonella* exhibit remarkable similarities to those in related sap-feeding insects. The parallels are particularly strong with the citrus mealybug, which harbors an endosymbiont that, like *Carsonella*, maintains one of the most reduced bacterial genomes ever identified. The host genomes in both insects have independently acquired bacterial genes in a largely identical set of functional categories, including amino acid biosynthesis/metabolism, riboflavin biosynthesis, and rRNA methylation (table 4). The parallelism between aphids and mealybugs also extends to many of the inferred phylogenetic sources of bacterial HGT and the upregulation of existing eukaryotic genes in the bacteriome. However, there are also clear differences between the insect groups. Notably, we did not find any evidence that the psyllid genome had acquired peptidoglycan biosynthesis genes, supporting the hypothesis that the abundant transfer of peptidoglycan genes to the citrus mealybug genome is related to the regulation of the unique nested endosymbiosis in that host (Husnik et al. 2013). Overall, the striking similarities in the histories of HGT in psyllids and mealybugs outweigh the differences and suggest that HGT may be an important force in facilitating the extreme genome reduction observed in many endosymbionts of sap-feeding insects. Recent evidence that HGT has also shaped amino acid biosynthesis pathways in trypanosomatids and their endosymbiotic bacteria suggests that this may be a more widespread phenomenon in eukaryotes (Alves et al. 2013). The independent origins of nutritional endosymbionts in numerous insects with highly specialized diets (Moran et al. 2008), combined with the impending flood of data from insect genome sequencing initiatives (Robinson et al. 2011), should provide an exciting opportunity to identify the basic biological mechanisms that promote or constrain HGT in these symbiotic consortia.

The Functional Significance of Host Acquisition of ASL from *Carsonella*

In some cases (e.g., CM), the function of host genes of bacterial origin can be readily inferred because no other genes with equivalent activity can be identified in either the host or endosymbiont genomes. However, the function of the *argH*-derived ASL genes found in the *P. venusta* genome is more ambiguous. We identified two divergent copies in the host genome in addition to the copy that is still retained in the *Carsonella* genome (Nakabachi et al. 2006). There are various potential explanations involving subfunctionalization and/or neofunctionalization for the maintenance of multiple gene copies. For example, it is possible that the terminal enzymatic step in the arginine biosynthesis pathway occurs in parallel in the distinct cellular compartments within psyllid bacteriomes for reasons related to arginine transport or regulation. However, multiple lines of evidence suggest that one or both of the host-encoded copies have taken over the ancestral role of *Carsonella argH* in arginine biosynthesis. First, both host copies are highly upregulated in the bacteriome relative to the rest of the body (fig. 3 and table 3). Second, an earlier analysis of the amino acid sequence of the *Carsonella argH* gene concluded that it is no longer functional as an ASL because of the loss of key catalytic residues (Tamames et al. 2007). Third, *argH* is retained and under purifying selection in all sequenced *Carsonella* genomes, even those in which the arginine pathway is otherwise disrupted and replaced by the presence of a second endosymbiont (Sloan and Moran 2012).

This strongly suggests that the gene has evolved a function that is independent of arginine biosynthesis. Fourth, the loss and functional replacement of endosymbiont enzymes catalyzing the terminal steps of amino acid biosynthesis pathways has been a recurring theme in sternorrhynchan insects (Hansen and Moran 2013). All these points suggest that the host-encoded genes now control the final step in arginine biosynthesis and that the *Carsonella* ArgH has adopted a new function, as observed for ASLs that have evolved a structural role as crystallins in the lenses of bird eyes (Piatigorsky et al. 1988).

Endosymbiotic Gene Transfer and Comparisons to Organelle Genome Evolution

Regardless of the function of the host-encoded ASL genes in psyllids, they represent an exception to the emerging pattern of bacterial HGT in sap-feeding insects. These genes appear to represent the first documented case in which an insect directly acquired a functional gene from its primary endosymbiont. In contrast, dozens of identified transfers in aphids, mealybugs, and now psyllids originated from other bacterial sources (Nikoh et al. 2010; Husnik et al. 2013). Given that gene transfer from organelle genomes to the nucleus has been a major mode of evolution in the history of eukaryotes (Adams and Palmer 2003; Timmis et al. 2004) and that functional transfer from primary endosymbionts can occur, why are such transfers so rare in sap-feeding insects? Below, we consider multiple potential explanations, all of which may play some role.

First, it is possible that there is an ascertainment bias and direct transfers are actually more common than currently recognized. Extreme nucleotide compositions and rapid evolution of protein sequences can make it difficult to identify genes in obligate endosymbionts. For example, despite the reduction of the *Carsonella* genome to a small set of the most essential and widely conserved genes, some *Carsonella* genes still lack detectable homology to any known protein sequences (Tamames et al. 2007). Therefore, the level of divergence in genes transferred from obligate endosymbionts may make it difficult to confidently assign their phylogenetic origin. We found that long branches made it difficult to pinpoint a specific HGT donor lineage for a number of the bacterial genes in the psyllid genome (supplementary fig. S3, Supplementary Material online). Notably, many of the most highly upregulated genes in the bacteriomes of sap-feeding insects lack detectable homology with any known protein sequences (supplementary table S1, Supplementary Material online; Hansen and Moran 2011; Shigenobu and Stern 2013). For these reasons, we believe the identified genes of bacterial origin likely represent a conservative estimate of the amount of HGT that has occurred in psyllids.

There may also be legitimate barriers to transfer that restrict gene movement from primary endosymbionts to their hosts. Borrowing from the literature on reproductive isolation, we can loosely classify these barriers as either "pre-zygotic" or "post-zygotic." Barriers that prevent the physical movement of DNA to the host genome would be analogous to prezygotic isolation. For example, the "limited transfer window" hypothesis holds that rates of plastid DNA transfer to the nucleus are lower in species with a single plastid per cell because plastid lysis and the release of its DNA into the cytoplasm would result in cell death in these species (Barbrook et al. 2006). The obvious example in sap-feeding insects is that, unlike organelles and bacteria that act as reproductive manipulators (e.g., *Wolbachia*), primary endosymbionts are only present in germline cells for a small fraction of the host life cycle (Buchner 1965; Nikoh et al. 2010). Because DNA transfers must occur in germline cells to be evolutionarily meaningful, there may be limited opportunities for movement of DNA from primary endosymbionts to the host genome. This may explain why only two small pseudogene fragments of *Buchnera* DNA were found in the pea aphid genome (Nikoh et al. 2010) in contrast to the veritable bombardment of organelle and bacterial DNA that occurs in the nuclear genomes of some eukaryotes (Huang et al. 2003; Dunning Hotopp et al. 2007).

Genes from primary endosymbionts may also face unique "post-zygotic" barriers that prevent them from being retained and becoming functional even after DNA is integrated into the host genome. Transferred DNA sequences must clear a number of hurdles to be functionally expressed, such as acquiring eukaryotic-specific promoters and regulatory elements (Rand et al. 2004). Although these hurdles apply to HGTs from any bacterial donor, genes from primary endosymbionts may face additional challenges related to their extreme sequence divergence. For example, their highly biased codon-usage may be poorly suited for translation by host machinery. Some insect endosymbionts even have a modified genetic code (McCutcheon et al. 2009b; McCutcheon and Moran 2010; Bennett and Moran 2013), posing a further barrier to transfer, but this is not the case in aphids, mealybugs, or psyllids. In addition, protein sequences in primary endosymbionts are so degenerate that they often require massive coexpression of chaperones for proper folding (Moran 1996; Poliakov et al. 2011), which may not be available in the host environment. Therefore, gene transfers from ancient endosymbionts with long histories of mutation accumulation may be less likely to be tolerated by selection.

The inferred timing of HGT events in psyllids exhibits interesting similarities and contrasts with the history of organelle genome evolution in eukaryotes. For example, almost all the identified transfer events occurred before the divergence of the major psyllid lineages. This is reminiscent of the massive gene transfer early in the evolution of mitochondria and plastids, which subsequently slowed or stopped entirely in some eukaryotic lineages (Gray et al. 1999). However, in groups with ongoing gene loss in organelle genomes, functional transfer to the nucleus continues to be the dominant mode of evolution (Adams and Palmer 2003). This does not appear to be the case in the psyllid-*Carsonella* system, as we found no evidence that recent losses of endosymbiont genes (e.g., in the histidine and tryptophan biosynthesis pathways) have been functionally replaced by bacterial HGT to the host. Regardless, it is clear that genes distributed across two or more genomic compartments mediate many of the metabolic pathways at

the heart of the relationship between sap-feeding insects and their nutritional endosymbionts. This raises fundamental biological questions about how these pathways are coordinated. In particular, even in the best-studied insect-endosymbiont systems, we still know remarkably little about which compounds are being moved between endosymbionts and the host cytoplasm (e.g., metabolites, proteins, or RNAs) and how they are being moved.

Our understanding of the early stages of eukaryotic genome evolution may merit additional scrutiny in light of the evidence that HGT from numerous bacterial lineages was involved in establishing ancient endosymbiotic relationships in sap-feeding insects. Although it is easy to imagine that the thousands of genes of bacterial origin in eukaryotes are overwhelmingly derived from the bacterial progenitors of mitochondria and plastids (Weeden 1981; Marcotte et al. 2000; Timmis et al. 2004), phylogenetic analyses have painted a more complex picture. Many genes of apparent bacterial origin in eukaryotic nuclear genomes do not specifically group with the Alphaproteobacteria or Cyanobacteria as would be expected if they were derived from organelle genomes (Karlberg et al. 2000; Gabaldón and Huynen 2007; Suzuki and Miyagishima 2010). Phylogenomic analysis of endosymbiotic gene transfer involving secondary acquisition of plastids from endosymbiotic algae has produced similar evidence for multiple gene donors (Curtis et al. 2012). It is, therefore, conceivable that a significant fraction of early HGT events associated with the establishment of endosymbiotic organelles in eukaryotes originated in bacterial lineages other than the progenitors of mitochondria and plastids (Huang and Gogarten 2007; Szklarczyk and Huynen 2010; Gray and Archibald 2012). For example, recent evidence supports the controversial hypothesis that the establishment of plastids from a cyanobacterial ancestor also involved intracellular bacteria from the Chlamydiales (Ball et al. 2013). Our findings lend plausibility to the notion that genes from multiple bacterial lineages may have been responsible for forging eukaryotic chromosomes.

Materials and Methods

Psyllid Collection and Dissection

Petiole galls containing *P. venusta* nymphs (third and fourth instar) were collected with leaves still attached from a single hackberry tree in Austin, TX, on September 20, 2012, and stored at 4 °C for approximately 2 weeks. Prior to dissection, the galls were left out overnight inside plastic bags to equilibrate to room temperature. Individual psyllids were dissected with insect mounting pins in 100 µl of Qiagen RNAprotect Bacteria Reagent to isolate the bacteriome from the rest of the insect body. Both male and female nymphs were used. In total, 72 nymphs were dissected and divided into three pools of 24 individuals each, which were used for RNA extraction and transcriptome sequencing. Separate collections of hackberry petiole galls containing fifth-instar nymphs were also made from a population in Tucson, AZ, in October 2006 and 2011 and used for qPCR analysis and whole-genome sequencing, respectively.

RNA Extraction and Illumina Sequencing

Pooled samples of dissected bacteriomes and corresponding psyllid bodies were centrifuged for 5 min at 13,000 × g to pellet tissue, and the RNAprotect supernatant was removed. RNA was extracted from pelleted tissue using TRI Reagent Solution (Invitrogen) following manufacturer's instructions. Each RNA sample was run on an Agilent 2100 Bioanalyzer with an RNA Nano 6000 Chip to verify RNA quality and estimate quantity. Each pool of 24 psyllids yielded an average of 3.3 and 58.2 µg of total RNA from the bacteriomes and bodies, respectively, and 1 µg from each sample was used for constructing Illumina mRNA-seq libraries. Polyadenylated RNA transcripts were enriched from total RNA with two rounds of selection on Dynabeads Oligo$(dT)_{25}$ (Invitrogen). RNA was fragmented to a mean size of approximately 120 nt, and first-strand cDNA synthesis was performed with random hexamers and SuperScript III Reverse Transcriptase (Invitrogen). Strand specificity was achieved by replacing dTTP with dUTP in the nucleotide pool for second-strand cDNA synthesis, which was performed with DNA polymerase I (New England Biolabs). Following ligation of library adapters (Parhomchuk et al. 2009), cDNA fragments were treated with uracil-DNA glycosylase (New England Biolabs) to remove uracils introduced during second-strand synthesis so that library fragments produced in the subsequent amplification step were all derived from first-strand cDNA. Library amplification was performed with eight cycles of PCR using the KAPA Bio HiFi polymerase and barcoded primers. The six resulting libraries were multiplexed and sequenced on a single lane of a 2 × 76 bp paired-end run on an Illumina HiSeq 2000.

Transcriptome Assembly, Annotation, and Differential Expression Analysis

Paired-end Illumina reads were trimmed to remove low-quality regions and adapter sequences using Trimmomatic v0.22 (Lohse et al. 2012) with the following parameters: ILLUMINACLIP:2:40:15 LEADING:20 TRAILING:20 SLIDINGWINDOW:4:20 MINLEN:40. As a result of a mechanical failure that occurred during cycle 40 of the second read in the Illumina sequencing run, there was an increased error rate at the ends of the second reads. Therefore, all second-read sequences were truncated to a length of 39 bp. To identify sequences derived from contaminating *Carsonella* transcripts, all reads (prior to truncation) were mapped to the *Carsonella* genome using SOAP v2.21 (Li et al. 2009), allowing for a maximum of three mismatches and a single indel of up to 3 bp. Read pairs were excluded if one or both sequences mapped to *Carsonella*. The resulting data sets from each of the 6 libraries were combined in a single de novo transcriptome assembly performed with Trinity r2013-02-25 (Grabherr et al. 2011). To identify homologous genes in other species, each assembled transcript was searched against the NCBI nr database (downloaded March 3, 2013) using NCBI-BlastX v2.2.27 + (Altschul et al. 1997) with a maximum *E* value of 1e-6 and a maximum of 25 target sequences. In addition, specific proteins of interest from the pea aphid genome (International Aphid Genomics Consortium 2010) were searched against the

assembled transcriptome with NCBI-TBlastN + . Amino acid biosynthesis pathways were reconstructed manually, using the EcoCyc database (Keseler et al. 2013) and published analyses in related insects as a guide (Wilson et al. 2010; Hansen and Moran 2011; Poliakov et al. 2011; Husnik et al. 2013).

To identify genes that were differentially expressed between the bacteriome and the rest of the psyllid body, reads from each library were mapped on to the Trinity assembly with bowtie v0.12.8 (Langmead et al. 2009) and analyzed with RSEM v1.2.3 (Li and Dewey 2011) and edgeR v2.4.6 (Robinson et al. 2010), using the run_RSEM_align_n_estimate.pl and run_DE_analysis.pl scripts distributed with Trinity (Grabherr et al. 2011; Haas et al. 2013). The three bacteriome libraries and three body libraries were treated as biological replicates in this analysis. Intraclass correlation coefficients among replicate libraries were calculated with the irr package in R v2.15.2.

Identification of Genes of Bacterial Origin

To identify psyllid genes of bacterial origin, the output from BlastX searches against the NCBI nr database for each transcript was analyzed with MEGAN v4.70.4 (Huson et al. 2011) with the following parameters: min support = 5; min score = 50; top percent = 10; win score = 0; and min complexity = 0.3. After the exclusion of duplicate sequences from the same Trinity subcomponent, MEGAN assigned 152 assembled transcripts to one of the nodes within the Bacteria clade in the NCBI taxonomy. The Blast hits from each of these transcripts were individually inspected to remove spurious assignments and likely contaminants. Many sequences were excluded because they exhibited essentially perfect identity with *Carsonella* or other common bacteria. The filtering of bacterial-like sequences from eukaryotic genome and transcriptome assemblies may be an important source of bias in underestimating the frequency of HGT (Dunning Hotopp et al. 2007). However, our exclusion of identical and near-identical hits should only preclude identification of extremely recent transfers. Furthermore, sequences matching *Carsonella* were almost completely restricted to bacteriome samples, providing evidence that these excluded sequences were contaminants rather than the result of very recent transfers. Sequences derived from rRNA genes were also excluded because they were expected to be expressed at high levels by contaminating bacteria. After manual filtering, this analysis produced a list of genes putatively acquired by horizontal transfer from bacteria (table 3). Further phylogenetic analysis of candidate genes was performed with RAxML v7.4.4 (Stamatakis 2006), using an LG + gamma substitution model, as identified by ProtTest v2.4 (Abascal et al. 2005). For each analysis, 100 bootstrap replicate searches were performed.

Pachypsyllya venusta Genome Sequencing and Assembly

As part of the pilot phase of the i5k arthropods genomics initiative, we generated Illumina sequence data for an initial draft assembly of the *P. venusta* genome. Genomic DNA was isolated with the Qiagen Blood and Tissue DNeasy Kit from fifth-instar *P. venusta* nymphs (mixed sexes) collected from Tucson, AZ, on November 21, 2011. Extracted DNA was used to generate Illumina sequencing libraries with a range of different insert sizes according to standard protocols. Paired-end libraries with target insert sizes of 180 bp and 500 bp were generated with DNA isolated from a single insect each. Mate-pair libraries with target insert sizes of 3 kb and 8–10 kb were generated from pooled DNA samples derived from 6 and 20 individual insects, respectively. All four libraries were sequenced on an Illumina HiSeq 2000 (2 × 100 bp runs). The three smaller insert libraries were sequenced to 40× coverage and the 8–10 kb mate-pair library was sequenced to 20× coverage. The resulting sequence data were assembled with ALLPATHS-LG v35218 (Gnerre et al. 2011) with the minimum contig length set to 300 and the "haploidify" option enabled. The assembly was incrementally improved with Atlas-Link and Atlas-Gapfill (https://www.hgsc.bcm.edu/software/, last accessed January 16, 2014). Select transcript sequences derived from mRNA-Seq data were mapped to the draft genome assembly with NCBI-BlastN + .

Identification of Bacterial HGTs Shared with Other Psyllids and Sternorrhynchan Insects

To determine whether observed genes of bacterial origin were the results of ancient HGT events, we searched for the presence of these genes in two distantly related psyllids (*Bactericera cockerelli* and *D. citri*). The Diaci1.1 genome assembly for *D. citri* and the 0074-L04-B02 transcriptome assembly for *B. cockerelli* were obtained from http://psyllid.org/ (last accessed January 16, 2014). Related copies of *P. venusta* genes in *B. cockerelli* and *D. citri* were identified with NCBI-TBlastN searches, and orthology was assessed by comparing *B. cockerelli* and *D. citri* hits with the top hits in the NCBI nr database. We also searched sequenced genomes and transcriptomes from related lineages of sap-feeding insects, including the pea aphid *Acyrthosiphon pisum* (International Aphid Genomics Consortium 2010), the citrus mealybug *Planococcus citri* (Husnik et al. 2013), and the whitefly *Bemisia tabaci* (Wang et al. 2010, 2012).

Quantitative PCR

Real-time qPCR was performed using the Roche LightCycler system as described previously (Nakabachi et al. 2005) with both cDNA and genomic DNA templates. Primers were designed to amplify the *P. venusta* genes encoding ASL-1 and ASL-2 as well as the ribosomal protein (Rp)L18e, which was used for normalization (supplementary table S2, Supplementary Material online). *Pachypsylla venusta* nymphs were dissected in phosphate-buffered saline (pH 7.4), and three different tissue types were used for amplification: 1) bacteriomes, 2) heads, and 3) whole bodies (without bacteriomes). The three tissue types were derived from pools of six, five and one individual(s), respectively. Analyses were performed with five biological replicates and three technical replicates.

Data Access

Raw Illumina reads from *P. venusta* are available from the NCBI SRA for both the RNA-seq (SRA099681) and genomic data sets (SRA093597, SRA093598), and assembled transcript sequences are available from the NCBI TSA database (GAOP00000000). Assembled genomic sequences are available from the NCBI (BioProject PRJNA167476) and the BCM-HGSC i5k pilot website (http://www.hgsc.bcm.tmc.edu/content/i5k-hackberry-petiole-gall-psyllid, last accessed January 16, 2014).

Supplementary Material

Supplementary figures S1–S3 and tables S1 and S2 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank Quinn McFrederick for kindly collecting hackberry galls and John Overton for helpful advice on Illumina sequencing. They acknowledge the Yale Center for Genome Analysis and the Baylor College of Medicine Human Genome Sequencing Center library construction and sequencing teams. This study was supported by Yale University, Colorado State University, a National Institutes of Health Postdoctoral Fellowship (1F32GM099334), and the Japan Society for the Promotion of Science (KAKENHI 21687020 and 2411751). The i5k genome sequencing efforts were supported by the National Institutes of Health (NHGRI U54 HG003273). Computational analyses were performed with resources from the Biomedical HPC Center at Yale University, which is supported by the National Institutes of Health (RR19895 and RR029676-01), and from the ISTeC Cray HPC System at Colorado State University, which is supported by the National Science Foundation (CNS-0923386).

References

Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21:2104–2105.

Adams KL, Palmer JD. 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Mol Phylogenet Evol.* 29: 380–395.

Akman Gündüz E, Douglas AE. 2009. Symbiotic bacteria enable insect to use a nutritionally inadequate diet. *Proc Biol Sci.* 276:987–991.

Altschul S, Madden T, Schäffer A, Zhang J, Zhang Z, Miller W, Lipman D. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402.

Alves JMP, Klein CC, da Silva FM, Costa-Martins AG, Serrano MG, Buck GA, Vasconcelos ATR, Sagot M-F, Teixeira MMG, Motta MCM, et al. 2013. Endosymbiosis in trypanosomatids: the genomic cooperation between bacterium and host in the synthesis of essential amino acids is heavily influenced by multiple horizontal gene transfers. *BMC Evol Biol.* 13:190.

Andersson SG, Kurland CG. 1998. Reductive evolution of resident genomes. *Trends Microbiol.* 6:263–268.

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

Ball SG, Subtil A, Bhattacharya D, Moustafa A, Weber AP, Gehre L, Colleoni C, Arias MC, Cenci U, Dauvillee D. 2013. Metabolic effectors secreted by bacterial pathogens: essential facilitators of plastid endosymbiosis? *Plant Cell* 25:7–21.

Barbrook AC, Howe CJ, Purton S. 2006. Why are plastid genomes retained in non-photosynthetic organisms? *Trends Plant Sci.* 11: 101–108.

Bennett GM, Moran NA. 2013. Small, smaller, smallest: the origins and evolution of ancient dual-obligate symbioses in a phloem-feeding insect. *Genome Biol Evol.* 5:1675–1688.

Buchner P. 1965. Endosymbiosis of animals with plant microorganisms. New York: John Wiley & Sons.

Curtis BA, Tanifuji G, Burki F, Gruber A, Irimia M, Maruyama S, Arias MC, Ball SG, Gile GH, Hirakawa Y. 2012. Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature* 492: 59–65.

Douglas AE. 2006. Phloem-sap feeding by animals: problems and solutions. *J Exp Bot.* 57:747–754.

Dunning Hotopp JC, Clark ME, Oliveira DC, Foster JM, Fischer P, Munoz Torres MC, Giebel JD, Kumar N, Ishmael N, Wang S, et al. 2007. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* 317:1753–1756.

Fukatsu T, Nikoh N. 1998. Two intracellular symbiotic bacteria from the mulberry psyllid *Anomoneura mori* (Insecta, Homoptera). *Appl Environ Microbiol.* 64:3599–3606.

Fukumorita T, Chino M. 1982. Sugar, amino acid and inorganic contents in rice phloem sap. *Plant Cell Physiol.* 23:273–283.

Gabaldón T, Huynen MA. 2007. From endosymbiont to host-controlled organelle: the hijacking of mitochondrial protein synthesis and metabolism. *PLoS Comput Biol.* 3:2209–2218.

Gnerre S, MacCallum I, Przybylski D, Ribeiro F, Burton J, Walker B, Sharpe T, Hall G, Shea T, Sykes S, et al. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci U S A.* 108:1513–1518.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 29:644–652.

Gray MW, Archibald JM. 2012. Origins of mitochondria and plastids. In: Bock R, Knoop V, editors. Genomics of chloroplasts and mitochondria. Dordrecht (The Netherlands): Springer. p. 1–30.

Gray MW, Burger G, Lang BF. 1999. Mitochondrial evolution. *Science* 283:1476–1481.

Haas BJ, Papanicolaou A, Moran Y, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc.* 8:1494–1512.

Hansen AK, Moran NA. 2011. Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. *Proc Natl Acad Sci U S A.* 108:2849–2854.

Hansen AK, Moran NA. 2013. The impact of microbial symbionts on host plant utilization by herbivorous insects. *Mol Ecol.* Advance Access published August 16, 2013, doi: 10.1111/mec.12421.

Huang CY, Ayliffe MA, Timmis JN. 2003. Direct measurement of the transfer rate of chloroplast DNA into the nucleus. *Nature* 422:72–76.

Huang J, Gogarten JP. 2007. Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids? *Genome Biol.* 8:R99.

Husnik F, Nikoh N, Koga R, Ross L, Duncan RP, Fujie M, Tanaka M, Satoh N, Bachtrog D, Wilson ACC. 2013. Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell* 153:1567–1578.

Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC. 2011. Integrative analysis of environmental sequences using MEGAN4. *Genome Res.* 21:1552–1560.

International Aphid Genomics Consortium. 2010. Genome sequence of the pea aphid *Acyrthosiphon pisum*. *PLoS Biol.* 8:e1000313.

Jansson K, Blomberg A, Sunnerhagen P, Alm Rosenblad M. 2010. Evolutionary loss of 8-oxo-G repair components among eukaryotes. *Genome Integr.* 1:12.

Jones JT, Furlanetto C, Bakker E, Banks B, Blok V, Chen Q, Phillips M, Prior A. 2003. Characterization of a chorismate mutase from the potato cyst nematode *Globodera pallida*. *Mol Plant Pathol.* 4: 43–50.

Karlberg O, Canbaück B, Kurland CG, Andersson SGE. 2000. The dual origin of the yeast mitochondrial proteome. *Yeast* 17:170–187.
Kelkar YD, Ochman H. 2013. Genome reduction promotes increase in protein functional complexity in bacteria. *Genetics* 193:303–307.
Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, Bonavides-Martinez C, Fulcher C, Huerta AM, Kothari A, Krummenacker M, et al. 2013. EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res.* 41:D605–D612.
Kuo CH, Ochman H. 2009. Deletional bias across the three domains of life. *Genome Biol Evol.* 1:145–152.
Lamb K. 1959. Composition of the honeydew of the aphid *Brevicoryne brassicae* (L.) feeding on swedes (*Brassica napobrassica* DC.). *J Insect Physiol.* 3:1–13.
Lambert KN, Allen KD, Sussex IM. 1999. Cloning and characterization of an esophageal-gland-specific chorismate mutase from the phytoparasitic nematode *Meloidogyne javanica*. *Mol Plant Microbe Interact.* 12:328–336.
Lamelas A, Gosalbes MJ, Manzano-Marin A, Pereto J, Moya A, Latorre A. 2011. *Serratia symbiotica* from the aphid *Cinara cedri*: a missing link from facultative to obligate insect endosymbiont. *PLoS Genet.* 7: e1002357.
Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25.
Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323.
Li B, Ruotti V, Stewart RM, Thomson JA, Dewey CN. 2010. RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics* 26:493–500.
Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J. 2009. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 25: 1966–1967.
Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M, Usadel B. 2012. RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Res.* 40:W622–W627.
Macdonald SJ, Lin GG, Russell CW, Thomas GH, Douglas AE. 2012. The central role of the host cell in symbiotic nitrogen metabolism. *Proc Biol Sci.* 279:2965–2973.
Marcotte EM, Xenarios I, van Der Bliek AM, Eisenberg D. 2000. Localizing proteins in the cell from their phylogenetic profiles. *Proc Natl Acad Sci U S A.* 97:12115–12120.
McCutcheon JP. 2010. The bacterial essence of tiny symbiont genomes. *Curr Opin Microbiol.* 13:73–78.
McCutcheon JP, McDonald BR, Moran NA. 2009a. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. *Proc Natl Acad Sci U S A.* 106:15394–15399.
McCutcheon JP, McDonald BR, Moran NA. 2009b. Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. *PLoS Genet.* 5:e1000565.
McCutcheon JP, Moran NA. 2010. Functional convergence in reduced genomes of bacterial symbionts spanning 200 my of evolution. *Genome Biol Evol.* 2:708–718.
McCutcheon JP, Moran NA. 2012. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol.* 10:13–26.
McCutcheon JP, von Dohlen CD. 2011. An interdependent metabolic patchwork in the nested symbiosis of mealybugs. *Curr Biol.* 21: 1366–1372.
Mira A, Ochman H, Moran NA. 2001. Deletional bias and the evolution of bacterial genomes. *Trends Genet.* 17:589–596.
Moran NA. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc Natl Acad Sci U S A.* 93:2873–2878.
Moran NA, McCutcheon JP, Nakabachi A. 2008. Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet.* 42:165–190.
Moran NA, Wernegreen JJ. 2000. Lifestyle evolution in symbiotic bacteria: insights from genomics. *Trends Ecol Evol.* 15:321–326.
Mount SM, Burks C, Hertz G, Stormo GD, White O, Fields C. 1992. Splicing signals in *Drosophila*: intron size, information content, and consensus sequences. *Nucleic Acids Res.* 20:4255–4262.
Nakabachi A, Koshikawa S, Miura T, Miyagishima S. 2010. Genome size of *Pachypsylla venusta* (Hemiptera: Psyllidae) and the ploidy of its bacteriocyte, the symbiotic host cell that harbors intracellular mutualistic bacteria with the smallest cellular genome. *Bull Entomol Res.* 100:27–33.
Nakabachi A, Shigenobu S, Sakazume N, Shiraki T, Hayashizaki Y, Carninci P, Ishikawa H, Kudo T, Fukatsu T. 2005. Transcriptome analysis of the aphid bacteriocyte, the symbiotic host cell that harbors an endocellular mutualistic bacterium, *Buchnera*. *Proc Natl Acad Sci U S A.* 102:5477–5482.
Nakabachi A, Ueoka R, Oshima K, Teta R, Mangoni A, Gurgui M, Oldham NJ, van Echten-Deckert G, Okamura K, Yamamoto K, et al. 2013. Defensive bacteriome symbiont with a drastically reduced genome. *Curr Biol.* 23:1478–1484.
Nakabachi A, Yamashita A, Toh H, Ishikawa H, Dunbar HE, Moran NA, Hattori M. 2006. The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. *Science* 314:267.
Nikoh N, McCutcheon JP, Kudo T, Miyagishima SY, Moran NA, Nakabachi A. 2010. Bacterial genes in the aphid genome: absence of functional gene transfer from *Buchnera* to its host. *PLoS Genet.* 6: e1000827.
Nikoh N, Nakabachi A. 2009. Aphids acquired symbiotic genes via lateral gene transfer. *BMC Biol.* 7:12.
Nowack ECM, Grossman AR. 2012. Trafficking of protein into the recently established photosynthetic organelles of *Paulinella chromatophora*. *Proc Natl Acad Sci U S A.* 109:5340.
Parhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, Krobitsch S, Lehrach H, Soldatov A. 2009. Transcriptome analysis by strand-specific sequencing of complimentary DNA. *Nucleic Acids Res.* 37:e123.
Piatigorsky J, O'Brien WE, Norman BL, Kalumuck K, Wistow GJ, Borras T, Nickerson JM, Wawrousek EF. 1988. Gene sharing by delta-crystallin and argininosuccinate lyase. *Proc Natl Acad Sci U S A.* 85:3479–3483.
Poliakov A, Russell CW, Ponnala L, Hoops HJ, Sun Q, Douglas AE, van Wijk KJ. 2011. Large-scale label-free quantitative proteomics of the pea aphid-*Buchnera* symbiosis. *Mol Cell Proteomics.* 10:M110.007039.
Price DRG, Duncan RP, Shigenobu S, Wilson AC. 2011. Genome expansion and differential expression of amino acid transporters at the aphid/*Buchnera* symbiotic interface. *Mol Biol Evol.* 28: 3113–3126.
Rand DM, Haney RA, Fry AJ. 2004. Cytonuclear coevolution: the genomics of cooperation. *Trends Ecol Evol.* 19:645–653.
Rasnitsyn AP, Quicke DLJ. 2002. History of insects. Dordrecht (The Netherlands): Kluwer.
Robinson GE, Hackett KJ, Purcell-Miramontes M, Brown SJ, Evans JD, Goldsmith MR, Lawson D, Okamuro J, Robertson HM, Schneider DJ. 2011. Creating a buzz about insect genomes. *Science* 331:1386.
Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140.
Russell CW, Bouvaine S, Newell PD, Douglas AE. 2013. Shared metabolic pathways in a coevolved insect-bacterial symbiosis. *Appl Environ Microbiol.* 79:6117–6123.
Sandström J, Pettersson J. 1994. Amino acid composition of phloem sap and the relation to intraspecific variation in pea aphid (*Acyrthosiphon pisum*) performance. *J Insect Physiol.* 40:947–955.
Shigenobu S, Stern DL. 2013. Aphids evolved novel secreted proteins for symbiosis with bacterial endosymbiont. *Proc Biol Sci.* 280: 20121952.
Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H. 2000. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. *APS. Nature* 407:81–86.
Sloan DB, Moran NA. 2012. Genome reduction and co-evolution between the primary and secondary bacterial symbionts of psyllids. *Mol Biol Evol.* 29:3781–3792.
Sloan DB, Moran NA. 2013. The evolution of genomic instability in the obligate endosymbionts of whiteflies. *Genome Biol Evol.* 5: 783–793.

Spaulding AW, von Dohlen CD. 1998. Phylogenetic characterization and molecular evolution of bacterial endosymbionts in psyllids (Hemiptera: Sternorrhyncha). *Mol Biol Evol.* 15: 1506–1513.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.

Suzuki K, Miyagishima S. 2010. Eukaryotic and eubacterial contributions to the establishment of plastid proteome estimated by large-scale phylogenetic analyses. *Mol Biol Evol.* 27:581–590.

Szklarczyk R, Huynen M. 2010. Mosaic origin of the mitochondrial proteome. *Proteomics* 10:4012–4024.

Tamames J, Gil R, Latorre A, Peretó J, Silva FJ, Moya A. 2007. The frontier between cell and organelle: genome analysis of *Candidatus* Carsonella ruddii. *BMC Evol Biol.* 7:181.

Thao ML, Moran NA, Abbot P, Brennan EB, Burckhardt DH, Baumann P. 2000. Cospeciation of psyllids and their primary prokaryotic endosymbionts. *Appl Environ Microbiol.* 66:2898–2905.

Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet.* 5:123–135.

von Dohlen CD, Kohler S, Alsop ST, McManus WR. 2001. Mealybug beta-proteobacterial endosymbionts contain gamma-proteobacterial symbionts. *Nature* 412:433–436.

Wagner GP, Kin K, Lynch VJ. 2012. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* 131:281–285.

Wang XW, Luan JB, Li JM, Bao YY, Zhang CX, Liu SS. 2010. De novo characterization of a whitefly transcriptome and analysis of its gene expression during development. *BMC Genomics* 11:400.

Wang XW, Zhao QY, Luan JB, Wang YJ, Yan GH, Liu SS. 2012. Analysis of a native whitefly transcriptome and its sequence divergence with two invasive whitefly species. *BMC Genomics* 13:529.

Weeden NF. 1981. Genetic and biochemical implications of the endosymbiotic origin of the chloroplast. *J Mol Evol.* 17:133–139.

Weibull J, Brishammar S, Pettersson J. 1986. Amino acid analysis of phloem sap from oats and barley: a combination of aphid stylet excision and high performance liquid chromatography. *Entomol Exp Appl.* 42:27–30.

Wilson AC, Ashton PD, Calevro F, Charles H, Colella S, Febvay G, Jander G, Kushlan PF, Macdonald SJ, Schwartz JF, 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.* 19:249–258.

Wu D, Daugherty SC, Van Aken SE, Pai GH, Watkins KL, Khouri H, Tallon LJ, Zaborsky JM, Dunbar HE, Tran PL, et al. 2006. Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PLoS Biol.* 4:1079–1092.