

Supplementary Information for

**Obligate endosymbionts limit thermal tolerance of
host species: *Buchnera* and aphids**

Bo Zhang, Sean P. Leonard, Yiyuan Li, Nancy A. Moran

corresponding author: Nancy Moran

Email: nancy.moran@austin.utexas.edu

This PDF file includes:

Supplementary Text

Figure S1

Tables S1 to S5

SI References

Supplementary Information Text

Supplementary Methods

Real-time qPCR of Buchnera titer: Genomic DNA was extracted from individual aphids using Qiagen DNeasy Blood & Tissue Kits. Copy numbers of genes encoding aphid elongation factor-1 alpha (Ef1 α) and *Buchnera* heat-shock protein (GroES) were determined by respective linear standard curves covering the range from 10² to 10⁸ copies. Due to unexpectedly low numbers of *Buchnera* in *Ap. fabae*, we repeated assays using *dnaK* as the *Buchnera* gene, and confirmed these low numbers. All qPCR samples were performed on Eppendorf Realplex Mastercycler using BioRad iTaq SYBR Green master mix. The primers and cycling conditions used in the study are provided in SI Appendix (Table S2).

Fluorescent in situ hybridization: We followed published protocols for visualizing aphid symbionts in whole aphid bodies (Koga et al. 2009, 2012). Newly matured adults were sampled and fixed in Carnoy's Solution overnight before bleaching in 6% hydrogen peroxide in 80% ethanol to quench autofluorescence. After legs, antennae and cornicles were removed, samples were stored in absolute ethanol at -20°C until use. The FISH probe Cy5-*ApisP2a* (Koga et al. 2012) and DAPI were hybridized overnight. Samples were washed, mounted in SlowFade Diamond Antifade Mountant (ThermoFisher Scientific), and observed under a confocal laser scanning microscope (Zeiss 710). For each treatment and species, we inspected 4-5 aphids and picked a single representative aphid to perform a z-stack capture (15-30 slices at 0.54 microns per slice) using constant laser intensity and aperture. Images were acquired using ZEISS ZEN microscope software and further analyzed using FIJI image analysis software (<https://fiji.sc>). Images were cropped and rotated to show only the aphid body, with areas outside the dotted outlines arbitrarily set to black. Full z-stacks were transformed by maximum intensity projection and displayed using identical scaling. Insets in the figure were from a single z-slice and show only the TRITC channel (red), scaled linearly to improve visualization.

Buchnera genome sequencing: We extracted genomic DNA from single aphids to construct DNA libraries for *Aphis fabae* and *Aphis gossypii*. DNeasy Blood & Tissue Kits (Qiagen) were used to extract total DNA; DNA concentration was determined by Qubit DNA BR Assay Kit (ThermoFisher Scientific), 100 ng DNA per sample was used to construct DNA libraries using the Swift 2S Turbo Flexible DNA Library Kit (Swift Biosciences) pairing with full-length indexed adapters of Illumina TruSeq DNA Single Indexes set A. Library quantity and size distribution was assessed using a Bioanalyzer before sequencing libraries from both ends (2 × 150 bp) on an Illumina iSeq100 instrument.

We trimmed adaptors and removed low quality base pairs using Trimmomatic version 0.38 with the same parameters as described above. We used SPAdes for genomic assembly with `gkmers (--k=21,33,55,77,99,127)` and “`--careful`” mode to reduce the number of mismatches and short indels (Bankevich *et al.* 2012).

Buchnera genomes were annotated by the RAST online server (<http://rast.nmpdr.org/rast.cgi>).

Transcriptome sequencing and assembly: RNA extraction was performed using Qiagen RNeasy Mini kit following the manufacturer’s protocol. The gDNA was digested using DNase from the Ambion TURBO DNA-free Kit following the manufacturer’s protocol (ThermoFisher Scientific). Measured by Qubit RNA BR Assay Kit (ThermoFisher Scientific), 10 µg of total RNA per sample was purified by MICROBExpress Bacterial mRNA Enrichment Kit (ThermoFisher Scientific) to reduce bacterial rRNA. The concentration of purified RNA was measured by Qubit again before removing eukaryotic ribosomal RNA, using NEBNext rRNA Depletion Kit (New England Biolabs). The twice-purified rRNA-depleted RNA was directly used to construct libraries by the Ultra II Directional RNA Library Kit according to the manufacturer’s instructions (New England Biolabs). Library read size distribution was assessed using a Bioanalyzer, indicating sufficient quantity and quality of all samples. Libraries were sequenced on an Illumina NextSeq500 instrument using

single 75 bp reads, at the University of Texas Genome Sequencing and Analysis Facility.

Trimmomatic version 0.38 (Bolger et al. 2014) was set to search for seed matches (16-bp) in sequencing reads allowing maximally 2 mismatches to the adaptor sequences, and then the seeds were extended and clipped when a score of 30 was reached (about 50-bp), or in the case of unpaired reads when a score of 10 was reached (about 17-bp). In particular, the program removed raw sequence reads from subsequent analysis if the read matched the Illumina sequencing adaptors with ≥ 8 -bp. Trimmomatic clipped the ends of reads if the quality of the base pairs at the start or the end of the read was below three. The program also clipped interior portions of the reads if the average quality was below 15 using a sliding window of 4-bp. After filtering, any reads shorter than 40 bp were removed. If the paired-ended reads were palindromic, both reads were kept for assembly.

For transcriptome assembly, we used Trinity version 2.8.4 with default parameters (Grabherr et al. 2011). For transcriptome assembly, we downloaded available transcriptome sequences for *Ap. fabae* on NCBI (SRR5831386-SRR5831424). We used Trimmomatic version 0.38 to trim adaptors and remove low quality base pairs with the same parameters as described above.

Functional annotation and ontology: To evaluate the completeness of our assembled transcriptome reference for *Ap. fabae*, we searched for the Benchmarking Universal Single-Copy Orthologs (BUSCOs) based on the 1,658 single-copy orthologs found in Insecta using BUSCO version 3 (Simão et al. 2015) and OrthoDB version 9 (Kriventseva et al. 2014). We searched for protein-coding sequences in the assembled transcriptomes using TransDecoder version 5.5.0 (Haas et al. 2013). We annotated the gene functions based on KEGG pathways and GO terms using KEGG GhostKOALA (Kanehisa et al. 2016) and PANNZER2 (Toronen et al. 2018), respectively.

Processing of RNA-Seq reads and gene expression differentiation: Raw fastq data files were concatenated by pooling four lanes of sequencing reads before determining read quality in FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc). All reads were then trimmed and filtered to remove poor-quality reads using the Trimmomatic as described above. For *Ap. gossypii* and *Ac. pisum* samples, reads were mapped to the aphid and *Buchnera* genomes independently, using HISAT2 (Kim et al. 2015) and bwa (Li and Durbin 2009), respectively. In each case, the proportion of mapped reads was determined from the resulting alignments using samtools (Li et al. 2009). Mapped reads were assigned to genes and counted using the HTSeq toolkit (Ander et al. 2015). For *Ap. fabae*, Kallisto was used to quantify aphid transcript abundances based on the transcriptome reference (Bray et al. 2016) while *Buchnera* reads were mapped to genome references. All subsequent analyses were performed using DESeq2 (Love et al. 2014) in R statistical language in RStudio Version 1.2.1335 (RStudio Team 2015). Log-transformation of DESeq2-normalized data (rld) was obtained using the “rlog” function. The top upregulated genes from both aphid and *Buchnera* are listed in SI Appendix (Table S5). The orthologous gene clusters among three aphid species were explored using the OrthoVenn2 server (Xu et al. 2019). In each aphid line, we selected genes with support for gene differentiation at $p < 0.05$ and absolute transcript fold-change > 2 (SI Appendix, Table S5) to analyze the gene ontology term (GO term) enrichment by the agriGO v2.0 online service (Tian et al. 2017).

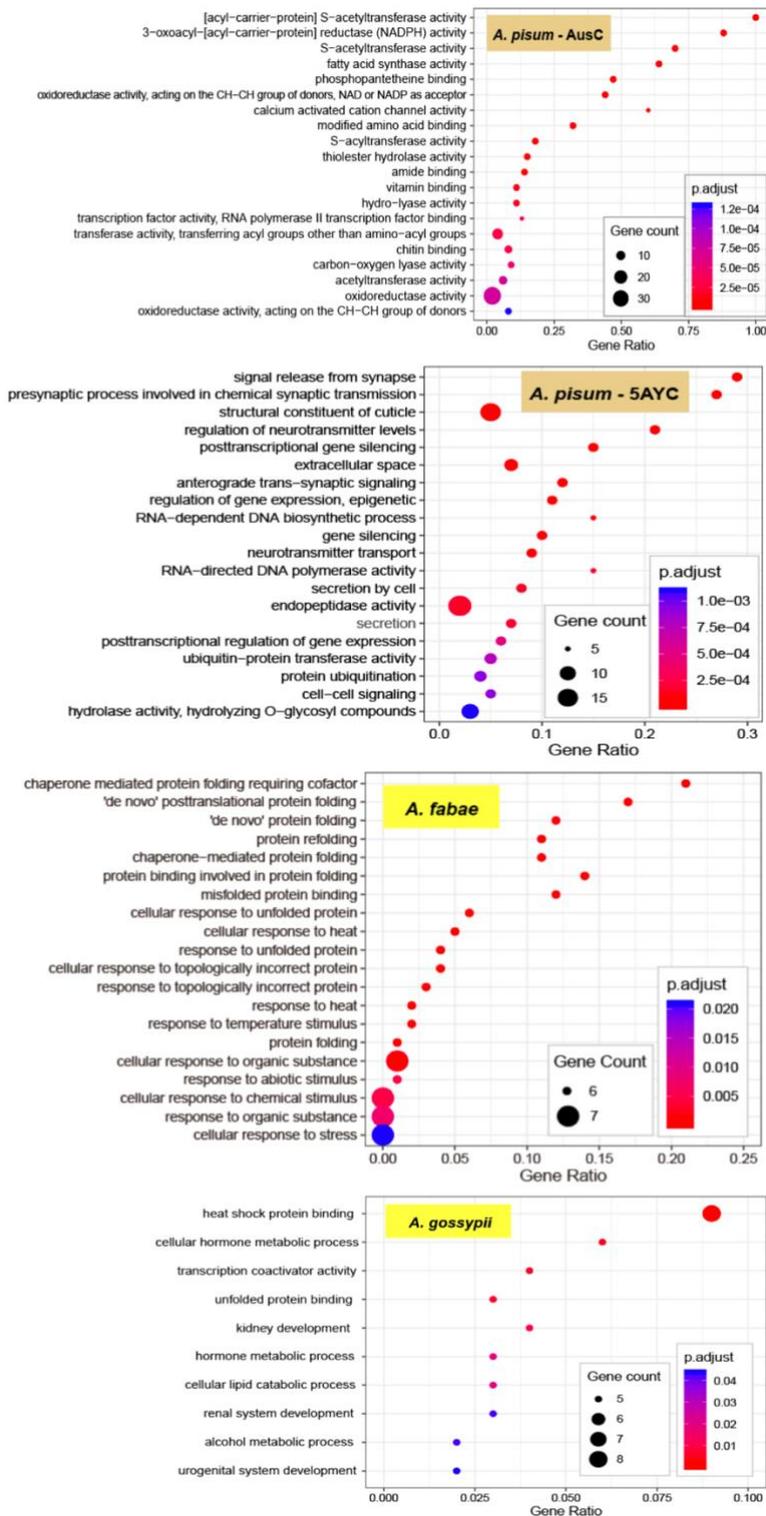


Fig. S1. Top GO term enrichment terms for genes upregulated following heat exposure in four aphid lines. Size of disc indicates number of genes in category and color indicates statistical support for category.

Table S1. Sources of aphid lines and samples.

Species/Strain	Origin location	Origin host plant	Origin date and collector	References and notes
<i>Acyrtosiphon kondoi</i>	Tucson AZ USA	<i>Medicago arborea</i> in greenhouse	March 2007, K. Hammond	Line used in experiments (Same as Degnan <i>et al.</i> 2011 <i>PLoS Genetics</i>)
<i>Acyrtosiphon pisum</i> Austin-Cured	Austin TX USA	<i>Vicia faba</i> in greenhouse	2014, K. Hammond	Line used in experiments (Cured of <i>Serratia symbiotica</i>)
<i>Acyrtosiphon pisum</i> 5AY-Cured	Madison WI USA	<i>Medicago</i> sp.	June 1999, N. Moran	Line used in experiments (Same as Moran & Jarvik 2010 <i>Science</i> , cured of facultative symbionts)
<i>Aphis craccivora</i>	Austin TX USA	<i>Phaseolus vulgaris</i>	2017, N. Moran	Line used in experiments
<i>Aphis fabae</i>	Belgium	unknown	May 2009	Line used in experiments. (Same line as Sabri <i>et al.</i> 2011 <i>Int J Syst Evol Microbiol</i> , cured of <i>Serratia symbiotica</i>)
<i>Aphis gossypii</i>	Austin TX USA	<i>Cucumis melo</i>	2018, J. Perreau	Line used in experiments
<i>Aphis fabae</i>	Berlin Germany	<i>Capsella</i> sp.	2019, N. Moran	DNA used to verify lack of <i>ibpA</i> in <i>Buchnera</i> genome
<i>Aphis fabae</i>	Berlin Germany	<i>Rumex</i> sp.	2019, N. Moran	DNA used to verify lack of <i>ibpA</i> in <i>Buchnera</i> genome
<i>Aphis fabae</i>	Berlin Germany	<i>Viburnum opulus</i>	2019, N. Moran	DNA used to verify lack of <i>ibpA</i> in <i>Buchnera</i> genome

Table S2. Statistical analyses of effect of heat exposure on *Buchnera* titer according to aphid line, heat treatment, day and interactions.

	Term	df	F value	p value
All lines	Line (L)	5	86.243	< 2e-16 ***
	Treatment (T)	1	164.929	< 2e-16 ***
	Day (D)	3	64.421	< 2e-16 ***
	L*T	5	14.063	2.61e-10 ***
	L*D	15	13.884	< 2e-16 ***
	T*D	3	5.438	0.0017 **
	L*T*D	15	6.491	2.46e-09 ***
<i>Ac. pisum</i> - AusC	Treatment (T)	1	94.509	4.06e-08 ***
	Day (D)	3	3.418	0.042902 *
	T*D	3	10.688	0.422 ***
<i>Ac. pisum</i> - 5AYC	Treatment (T)	1	13.028	0.002352 **
	Day (D)	3	10.793	0.000401 ***
	T*D	3	4.114	0.024266 *
<i>Ac. kondoi</i>	Treatment (T)	1	35.958	1.86e-05 ***
	Day (D)	3	12.225	0.000207 ***
	T*D	3	5.979	0.006206 **
<i>Ap. fabae</i>	Treatment (T)	1	24.925	0.000133 ***
	Day (D)	3	52.911	1.58e-08 ***
	T*D	3	3.537	0.038832 *
<i>Ap. gossypii</i>	Treatment (T)	1	10.632	0.00491 **
	Day (D)	3	24.195	0.00000346 ***
	T*D	3	0.583	0.63495
<i>Ap.</i> <i>craccivora</i>	Treatment (T)	1	47.55	3.60e-06 ***
	Day (D)	3	215.62	3.81e-13 ***
	T*D	3	24.92	2.86e-06 ***

Table S3. Summary of RNA-seq read mapped to aphid and *Buchnera* genomes.

library ID	Species	Treatment	Total data (base pairs)	# clean reads	Aphid mapped reads	Mapping %	Buchnera mapped reads	Mapping %	Total mapped reads	Total mapped %
FabaeC1	<i>Aphis fabae</i>	Control	2,672,771,100	35,636,948	32,959,830	92.49	2010902	5.64	34,970,732	98.13
FabaeC2			1,492,889,775	19,905,197	18,566,081	93.27	1247470	6.27	19,813,551	99.54
FabaeC3			1,859,096,925	24,787,959	22,862,286	92.23	1858268	7.50	24,720,554	99.73
FabaeH1		Heat	1,406,507,475	18,753,433	17,272,090	92.10	1357845	7.24	18,629,935	99.34
FabaeH2			1,220,453,025	16,272,707	15,296,305	94.00	836925	5.14	16,133,230	99.14
FabaeH3			1,337,869,875	17,838,265	16,477,307	92.37	1282420	7.19	17,759,727	99.56
GossyC1	<i>Aphis gossypii</i>	Control	1,650,228,375	22,003,045	3,827,634	17.40	1,475,557	6.71	5,303,191	24.10
GossyC2			1,594,793,325	21,263,911	3,515,616	16.53	1,045,978	4.92	4,561,594	21.45
GossyC3			1,661,213,850	22,149,518	3,478,081	15.70	1,041,562	4.70	4,519,643	20.41
GossyH1		Heat	1,394,695,950	18,595,946	2,921,389	15.71	1,288,809	6.93	4,210,198	22.64
GossyH2			1,399,530,375	18,660,405	2,989,011	16.02	1,250,941	6.70	4,239,952	22.72
GossyH3			1,191,573,825	15,887,651	2,458,601	15.47	898,777	5.66	3,357,378	21.13
AusC1	<i>Acyrtosiphon pisum -AusC</i>	Control	1,488,120,450	19,841,606	3,089,863	15.57	957,811	4.83	4,047,674	20.40
AusC2			1,349,408,625	17,992,115	2,474,517	13.75	661,791	3.68	3,136,308	17.43
AusC3			2,385,915,600	31,812,208	4,313,317	13.56	1,343,174	4.22	5,656,491	17.78
AusH1		Heat	1,673,271,375	22,310,285	3,906,968	17.51	1,289,497	5.78	5,196,465	23.29
AusH2			1,705,020,600	22,733,608	3,374,165	14.84	1,016,529	4.47	4,390,694	19.31
AusH3			1,587,057,150	21,160,762	2,739,119	12.94	776,509	3.67	3,515,628	16.61
5AYC1	<i>Acyrtosiphon pisum -5AYC</i>	Control	3,810,987,600	50,813,168	2,868,336	5.64	2111448	4.16	4,979,784	9.80
5AYC2			2,089,226,325	27,856,351	1,469,754	5.28	1079018	3.87	2,548,772	9.15
5AYC3			2,742,708,375	36,569,445	1,716,702	4.69	1181830	3.23	2,898,532	7.93
5AYH1		Heat	1,619,122,200	21,588,296	1,036,654	4.80	988148	4.58	2,024,802	9.38
5AYH2			2,932,856,850	39,104,758	1,758,739	4.50	1755500	4.49	3,514,239	8.99
5AYH3			2,549,378,250	33,991,710	1,847,095	5.43	1888371	5.56	3,735,466	10.99

Table S4. Top ten upregulated genes from aphids and *Buchnera* following heat exposure.

<i>Ac. pisum</i> - 5AYC - aphid				
Gene_ID	baseMean	log2FoldChange	padj	Product (red indicates heat shock gene)
100167145	458.3128	8.079816	2.56E-70	heat shock protein 68-like
100168413	225.0099	6.760927	6.29E-36	heat shock protein 68-like
100160289	215.8739	5.628106	4.64E-63	heat shock protein 70 A1-like
100163372	13.43809	4.951963	4.48E-06	uncharacterized protein LOC100163372
100160157	223.3622	4.362143	1.31E-75	SPRY domain-containing SOCS box protein 3
100168754	65.55485	3.443916	1.44E-20	uncharacterized protein LOC100168754
100166712	178.4332	3.163589	4.32E-08	adhesive plaque matrix protein
100161897	103.1355	2.829181	1.27E-26	uncharacterized protein LOC100161897
100164113	147.4312	2.781567	1.84E-26	cuticle protein-like isoform X1
100168430	66.66299	2.368003	4.54E-13	venom protease
<i>Ac. pisum</i> - 5AYC - <i>Buchnera</i>				
Gene_ID	baseMean	log2FoldChange	padj	Product
1109597	44308.37	6.427417	0	molecular chaperone DnaK
1109920	798.0937	5.302976	1.24E-218	ferredoxin-NADP reductase
1109919	847.6644	4.769763	5.85E-159	16 kDa heat shock protein A
1109527	73810.87	4.76124	1.52E-220	molecular chaperone GroEL
1109922	341.459	4.723024	5.49E-68	phosphopantetheine adenylyltransferase
1109526	11351.02	4.710015	6.04E-219	co-chaperonin GroES
1109596	4007.862	4.66036	4.13E-214	molecular chaperone DnaJ
1109925	96.29774	4.14275	9.88E-56	YhiQ hypothetical protein
1109744	489.1774	3.666575	6.97E-93	50S ribosomal protein L13
109743	976.3295	3.474754	2.13E-113	30S ribosomal protein S9
<i>Ac. pisum</i> - AusC - aphid				
Gene_ID	baseMean	log2FoldChange	padj	Product
100168413	267.3859494	5.883857	1.00E-67	heat shock protein 68-like
100167145	278.3466004	5.438237	1.14E-67	heat shock protein 68-like
100166890	4.872289697	4.726808	0.007903	angiotensin-converting enzyme-like isoform X3
100164459	4.630958666	4.657228	0.009878	cytochrome P450 6k1-like isoform X3
100165009	12.39832584	4.209613	2.36E-05	uncharacterized protein LOC100165009
100160289	323.5582772	4.105719	2.20E-65	heat shock protein 70 A1-like
100166605	8.594564879	4.003526	0.001499	glycine-rich cell wall structural protein-like
100160157	150.5241087	3.470803	1.66E-42	SPRY domain-containing SOCS box protein 3
100169526	30.96467505	3.395304	4.76E-11	arylsulfatase B-like
100164918	5.275089266	3.37332	0.023349	dynein beta chain ciliary-like
<i>Ac. pisum</i> - AusC - <i>Buchnera</i>				
Gene_ID	baseMean	log2FoldChange	padj	Product
1109528	724.7738	6.30712	1.35E-250	elongation factor P(efp)
1109597	70596.43	5.932531	3.06E-255	molecular chaperone DnaK(dnaK)
1109526	24521.59	4.329766	7.14E-106	co-chaperonin GroES(groES)

1109527	186173.1	4.201377	1.60E-107	molecular chaperone GroEL(groEL)
1109733	26.07614	4.131724	2.47E-15	preprotein translocase subunit SecG(secG)
1109596	5336.084	3.828675	3.45E-132	molecular chaperone DnaJ(dnaJ)
1109594	324.1869	3.625845	1.69E-50	FMN adenylyltransferase(ribF)
1109686	56.42018	3.448706	5.82E-29	hydroxyacylglutathione hydrolase(gloB)
1109963	86.97923	3.236355	3.84E-25	tRNA(BU624)
1109945	146778.9	2.806126	2.28E-33	16S ribosomal RNA(rrs)

Ap. fabae - aphid

Gene_ID	baseMean	log2FoldChange	padj	Product
DN1912_c0_g4	66.83902	7.109736	1.54E-17	Heat shock protein 68
DN1912_c0_g5	340.2075	6.8401	3.98E-69	Heat shock protein 68 HSP70
DN1912_c0_g2	202.0387	5.115627	8.31E-59	Heat shock protein 70
DN1912_c0_g3	51.4477	4.710199	1.82E-14	Heat shock protein 68
DN1912_c0_g1	93.25364	4.649184	4.36E-28	Heat shock protein 68
DN67_c0_g1	1378.925	4.617608	5.31E-13	Heat shock protein 68 HSP70
DN1456_c0_g1	94.85145	1.9187	0.0059	Ubiquitin carboxyl-terminal hydrolase
DN1349_c0_g1	104.143	1.688521	5.98E-10	Lipid storage droplets surface-binding protein 1C
DN1454_c0_g1	261.9811	1.184126	0.016458	Putative Clip-domain serine protease
DN9811_c0_g1	240.6797	1.016449	0.046037	Fatty acid synthase

Ap. fabae - Buchnera

Gene_ID	baseMean	log2FoldChange	padj	Product
410	385.6919	6.112146	2.38E-172	Translation elongation factor P
536	107592.8	5.859571	3.42E-193	Chaperone protein DnaK
409	323999.4	4.925892	1.22E-205	Heat shock protein 60 family chaperone GroEL
535	11990.59	4.858418	1.53E-183	Chaperone protein DnaJ
408	31432.78	4.770082	1.58E-194	Heat shock protein 60 family co-chaperone GroES
414	423.6683	3.403657	1.06E-89	Signal recognition particle receptor FtsY
281	207.4942	2.861173	4.57E-37	Shikimate 5-dehydrogenase I alpha
277	84.21392	2.751089	4.97E-16	Methenyltetrahydrofolate cyclohydrolase
571	634.462	2.734013	7.01E-58	Monothiol glutaredoxin GrxD
41	59.68869	2.587107	9.21E-15	DNA polymerase III epsilon subunit

Ap. gossypii - aphid

Gene_ID	baseMean	log2FoldChange	padj	Product
114124251	178.8841	6.907387	2.73E-23	heat shock protein 70 A1-like
114131704	298.8421	6.83628	2.03E-56	heat shock protein 68-like
114119569	278.6587	6.082816	2.14E-25	heat shock protein 70 A1-like
114122099	150.2895	5.909863	7.30E-42	heat shock protein 70 A1-like
114133077	43.3545	5.908767	1.57E-13	heat shock protein 70 A1-like
114132056	103.0518	5.67257	1.52E-13	heat shock protein 70 A1-like
114129492	212.138	1.898825	9.53E-22	protein-glutamate O-methyltransferase-like
114127182	61.05246	1.896211	1.73E-09	protein ABHD18
114129944	31.72557	1.833076	0.000274	guanine nucleotide-binding protein-like 1

114132933	35.28907	1.783734	0.000115	SPRY domain-containing SOCS box protein 3-like
<i>Ap. gossypii</i> - Buchnera				
Gene_ID	baseMean	log2FoldChange	padj	Product
13	2358.669369	6.6751	7.34E-73	16 kDa heat shock protein A (ibpA)
184	124417.0988	6.251196	8.18E-178	Chaperone protein DnaK
462	369.6630541	4.807173	5.40E-47	tRNA-modifying protein YgfZ
58	274525.6122	4.76337	1.18E-70	Heat shock protein 60 family GroEL
57	38588.00295	4.609548	2.46E-53	Heat shock protein 60 family GroES
183	5756.723782	4.488861	4.75E-76	Chaperone protein DnaJ
223	2797.950925	4.418786	1.69E-61	GTP-binding and nucleic acid-binding protein YchF
222	555.0127641	4.296327	1.24E-45	Peptidyl-tRNA hydrolase
461	523.6199011	3.980994	1.42E-34	Guanylate kinase
254	173.1862898	3.710269	3.33E-25	Phospho-N-acetylmuramoyl-pentapeptide-transferase

Table S5. Primers and PCR conditions* used for determining *Buchnera* titers.

Species	Gene	Sequence	Length (bp)	T _m (°C)
<i>Ac. kondoi</i>	GroES	F: CCATTGCATGATCGTGTGCTT R: TCCGACCGCTGTTACTGTTC	120	60
	EF1 a	F: GACGTGGTTCAAGGGATGGA R: GAGACGGAGAGCCTTGTCAG	121	60
<i>Ap. gossypii</i>	GroES	F: ATCTGCCGGTGGTATTGTTC R: AACTGCTGTAAGTGTGCCTC	70	58
	EF1 a	F: CGCACCTGGTCACAGAGATT R: TGCTCACGGGTTTGTCCATT	135	58
<i>Ap. craccivora</i>	GroES	F: AATCAAATCTGCTGGCGGT R: ACACGACCATTACCAACTGCT	91	60
	EF1 a	F: TGCACCACGAAGCTTTGGTA R: TGTCTCCAGCAACGAAACCA	105	60
<i>Ap. fabae</i>	GroES	F: TGGCGGTATTGTTCTTACAGG R: AGAACACGACCGTTACCCAC	81	59
	EF1 a	F: CGCACCTGGTCACAGAGATT R: TCACGGGTTTGTCCGTTCTT	132	59
	dnaK	F: ATGCAGGTAGAATCGCTGGT R: ACAGCTATAGTCCGGTCCCT	109	59

*Touchdown PCR cycling conditions: 95°C for 2min, [95 °C for 10 s, 65(-1) °C for 15 s, 72 °C for 20 s] × 5-7 cycles, and (95 °C for 10 s, T_m °C for 15 s, 72 °C for 20 s) × 33-35 cycles.

References:

- S. Anders, P. T. Pyl, W. Huber. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169 (2015).
- A. Bankevich, *et al.*, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455-477 (2012).
- A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* **30**, 2114-2120 (2014).
- N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525-527 (2016).
- M. G. Grabherr, *et al.*, Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* **29**, 644-652 (2011).
- B. J. Haas, *et al.*, De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc.* **8**, 1494-1512 (2013).
- M. Kanehisa, Y. Sato, K. Morishima, BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428**, 726-731 (2016).
- D. Kim, B. Langmead, S. L. Salzberg, HISAT: a fast spliced aligner with low memory requirements. *Nat. Meth.* **12**, 357-360 (2015).
- R. Koga, T. Tsuchida, T. Fukatsu, Quenching autofluorescence of insect tissues for in situ detection of endosymbionts. *Appl. Entomol. Zool.* **44**, 281-291 (2009).
- R. Koga, X. Y. Meng, T. Tsuchida, T. Fukatsu. Cellular mechanism for selective vertical transmission of an obligate insect symbiont at the bacteriocyte–embryo interface. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 1230-1237 (2012).
- E. V. Kriventseva, *et al.*, OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. *Nucl. Acids Res.* **43**, D250-D256 (2014).
- H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
- H. Li, *et al.*, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
- M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
- RStudio Team, 2015. RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL <http://www.rstudio.com/>. (accessed 1 August 2019)

- F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212 (2015).
- T. Tian, *et al.*, agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update, *Nucl. Acids Res.* **45**, W122-W129 (2017)
- P. Toronen, A. Medlar, L. Holm, PANNZER2: A rapid functional annotation webserver. *Nucl. Acids Res.* **46**, W84-W88 (2018).
- L. Xu, *et al.*, OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucl. Acids Res.* **47**, W52-W58 (2019).