Published in final edited form as:

J Proteome Res. 2013 February 1; 12(2): 704–718. doi:10.1021/pr3007842.

Proteomic analysis of an unculturable bacterial endosymbiont (*Blochmannia*) reveals high abundance of chaperonins and biosynthetic enzymes

Yongliang Fan[†], J. Will Thompson^{†,‡}, Laura G. Dubois^{†,‡}, M. Arthur Moseley^{†,‡}, and Jennifer J. Wernegreen^{†,§,*}

[†]Institute for Genome Sciences and Policy, Duke University, Box 3382, 101 Science Dr., Durham, NC27708

[‡]Duke Proteomics Core Facility, Duke University, B02 LSRC, Box 91009 450 Research Dr., Durham, NC 27708

[§]Nicholas School of the Environment, Duke University, Durham, NC 27708

Abstract

Many insect groups have coevolved with bacterial endosymbionts that live within specialized host cells. As a salient example, ants in the tribe Camponotini rely on *Blochmannia*, an intracellular bacterial mutualist that synthesizes amino acids and recycles nitrogen for the host. We performed a shotgun, label-free, LC/MS/MS quantitative proteomic analysis to investigate the proteome of *Blochmannia* associated with *Camponotus chromaiodes*. We identified more than 330 *Blochmannia* proteins, or 54% coverage of the predicted proteome, as well as 244 *Camponotus* proteins. Using the average intensity of the top 3 “best flier” peptides along with spiking of a surrogate standard at a known concentration, we estimated the concentration (fmol/μg) of those proteins with confident identification. The estimated dynamic range of *Blochmannia* protein abundance spanned three orders of magnitude and covered diverse functional categories, with particularly high representation of metabolism, information transfer, and chaperones. GroEL, the most abundant protein, totaled 6% of *Blochmannia* protein abundance. Biosynthesis of essential amino acids, fatty acids, and nucleotides, and sulfate assimilation had disproportionately high coverage in the proteome, further supporting a nutritional role of the symbiosis. This first quantitative proteomic analysis of an ant endosymbiont illustrates a promising approach to study the functional basis of intimate symbioses.

*Corresponding Author: CIEMAS rm. 2175, Box 3382, 101 Science Dr., Duke University, Durham, NC 27708. Tel: 1-919-681-0331; Fax: 1-919-668-0795; j.wernegreen@duke.edu.

Downloadable Files:

Several files are available on Tranche for this manuscript. They have been encrypted with the password “Blochmannia”. The .fasta sequence database used for this study is available at the following link: <https://proteomecommons.org/dataset.jsp?i=Ydtx60JloE60Cmqwq0IPgCFaZdZbsYDQfAVvshJUB1SCD%2Bgk2m6sWSxaXeWRgRFD0IdeDdRWK9TYt8d2VQunSfWbgssAAAAAADIADIA%3D%3D>. The MS/MS spectra and associated peptide identifications are available at the following link: <https://proteomecommons.org/dataset.jsp?i=WGSxTPy2oImldKjnw%2F01qeWZYu2%2BbcD3PvnHN0IKxI6ZeTNB2hdnq4yWoe2II5C9JTN1FD7%2F%2BYeTs4XNgK2zPy1j2PYAAAAAADAoA%3D%3D>

Author Contributions: YF and JJW collected ant specimen, YF prepared bacterial cells, JWT and LGD performed the proteomic analysis, all authors contributed to data analysis and interpretation. All authors contributed to the writing of the manuscript and have given approval to the final version.

The Supporting Information below is available free of charge via the Internet at <http://pubs.acs.org>.

Keywords

quantitative proteomics; high definition mass spectrometry; MS^E; HDMS^E; endosymbiosis; uncultivable bacteria; *Camponotus*; *Blochmannia*; GroEL

Introduction

The ubiquity and significance of microbes across diverse environments has become increasingly clear over the past 20 years. We now understand that microbes drive biogeochemical processes, perform key functions such as decomposition and nutrient synthesis, and are required for the survival of countless host species. Studying these critical functions has proved to be challenging, because >90% of microbes in a given environment cannot be cultured with standard methods.¹ Less than half of known bacterial phyla have cultured representatives.² By avoiding the need to culture individual species, DNA sequence-based sampling approaches (e.g., 16S rRNA tag sequencing and metagenomics) have catapulted studies of the genetic diversity, structure, and potential functions of unculturable microbial communities.³ Building on these techniques, proteomic approaches that do not rely on cultivation offer powerful tools to understand the metabolic functioning of microbial consortia across diverse ecosystems.^{4–7}

Many groups of unculturable bacteria perform key ecological functions within the context of close associations with animal hosts. Among animals, insects are particularly prone to establishing long-term relationships with intracellular bacterial symbionts that live exclusively within specialized host cells (called bacteriocytes) and perform key nutritional functions.⁸ These intracellular microbial associates have become models to study the evolution and functioning of intimate bacterial-animal mutualisms. Relatively well-studied examples include *Buchnera* in aphids, *Wigglesworthia* in tsetse flies, and the focus of this study, *Blochmannia* in the ant tribe Camponotini. In each case, the bacteria provide nutritional support and thus allow hosts to survive in ecological niches that otherwise would be inaccessible to them. Conversely, the insect hosts are thought to provide persistent, stable and nutrient-rich habitats to the endosymbionts. In ants, the endosymbionts live directly within the cytoplasm of host bacteriocytes that are intercalated among midgut epithelial cells, as well as within female ovaries, consistent with maternal transmission of bacteria from ant queens to developing eggs. This stable bacterial-ant relationship appears to be mutually obligate for both partners, as the bacteria cannot live outside of host cells, and host fitness is compromised if the endosymbionts are removed.^{9,10}

Availability of whole-genome sequences has allowed predictions of the biological functions such endosymbionts perform. Bacteriocyte-associated mutualists have very small genomes (<1 Mb, and often much smaller) as a consequence of their long-term intracellular lifestyle.¹¹ The genes they do retain typically highlight specific nutritional contributions to their respective host groups. For example, the small genomes of *Buchnera* of aphids maintain the ability to synthesize amino acids that are essential to the host but deficient in phloem sap diet.¹² The tsetse endosymbiont, *Wigglesworthia*, has lost most genes for amino acid biosynthesis but retains cofactor and vitamin biosynthesis genes.¹³ These functions complement the tsetse's mammalian blood diet, which is rich in amino acids but poor in certain cofactors and vitamins.

Likewise, *Blochmannia* in ants has a reduced genome size (0.7 to 0.79 Mb) compared to the moderate (4–5 Mb) genomes of related enterobacteria that retain a free-living phase in their lifecycle. *Blochmannia* strains encode in the range of 590 to 614 protein-coding genes.^{14–16} This genome reduction included the massive loss of genes that are no longer needed inside

the specialized bacteriocytes. Most notably, very few regulatory genes remain, and even the replication initiation gene, *dnaA*, has been lost.¹⁴ Despite severe genome reduction, *Blochmannia* retains genes for basic cellular functions, biosynthesis of nine essential amino acids (except arginine) and cofactors, and urease enzymes that allow this symbiont to recycle nitrogen from urea. By virtue of these nutritional functions, the symbiont plays a key role in the ability of worker ants to rear developing brood, and in the establishment of new ant colonies from single queens.^{9,10}

To date, functional genomic studies of *Blochmannia* indicate only modest fluctuations in transcript abundances. Recent studies have compared the transcription profile of *Blochmannia floridanus* across different developmental stages of its ant host (*Camponotus floridanus*), using genome-wide macroarrays and RT-qPCR of particular genes.^{17,18} These studies showed that the endosymbiont transcriptome varies, albeit modestly, across developmental stages of the host. While genes encoding basic metabolism have high expression in the early larval stages, so-called “symbiotic genes” (e.g., amino acid and fatty acid biosynthesis, nitrogen recycling, and other functions) show higher expression near eclosion, i.e., the late pupal stage.^{17,18} However, transcript differences rarely exceeded a factor of three, with a median across genes of 1.8-fold difference between maximal and minimal expression.¹⁷ Likewise, earlier microarray studies in *Buchnera* showed minimal fluctuations in transcript abundance of this aphid symbiont, as heat stress and host diet shifts produced only modest changes in the symbiont transcriptome.^{19,20} In *Buchnera*, genes showing a transcriptional response are limited to those retaining ancestral regulators. In general, the loss of many regulation functions in small endosymbiont genomes may constrain their ability to alter gene expression in response to host-level changes. Low transcriptional responses might also reflect relatively stable, nutrient-rich conditions in an intracellular microenvironment, which could buffer fluctuations that endosymbionts experience, e.g., across distinct host stages or nutritional profiles.

While investigations of transcript abundance have provided important insights into endosymbiont functions, such studies have inherent limitations. First, mRNA abundance does not always correlate well to the abundance of the corresponding protein. In particular, variation in translational efficiency, or changes in activity due to post-translational modifications can decouple transcript abundance from active protein levels. Studies in yeast have shown that the correlation between mRNA and protein levels was insufficient to reliably predict protein abundance from transcript levels.^{21,22} For a given gene, steady state levels of the protein were sometimes associated with >20-fold change in transcript levels, and vice versa.²¹ More fundamentally, cellular functions are typically performed by proteins rather than transcripts.

Second, the instability and rapid degradation of bacterial mRNA presents serious technical obstacles. Challenges of isolating bacterial mRNA are particularly severe for unculturable bacterial endosymbionts, for which nucleic acids are often scarce and mixed with insect host material. RNA-seq approaches offer alternatives to microarrays,²³ but face the distinct challenge of removing high-abundance rRNA, thereby requiring quantities of total RNA that are often unfeasible for unculturable bacteria. Therefore, studies of transcriptome dynamics have offered valuable insights, but also face technical obstacles and, by quantifying mRNA only, will necessarily offer indirect view of functional capabilities.

Given the above limitations, proteomics is a promising strategy to explore metabolic variation and plasticity in endosymbiotic species. In particular, bottom-up, label-free proteomic approaches offer valuable tools to explore the functional basis of symbiotic interactions, as these approaches can quantify hundreds or thousands of proteins in an unculturable organism simultaneously and have evolved rapidly to overcome technical

limitations such as proteome complexity and dynamic range.^{24–26} Applying label-free quantitative proteomics to study the aphid endosymbiont *Buchnera*, Poliakov et al.²⁴ revealed a high (68%) coverage of the *Buchnera* proteome and confirmed the key role of nutritional support in aphid/*Buchnera* symbiosis.

As in the *Buchnera* study, proteomic analysis of *Blochmannia* is expected to clarify functional mechanisms that underlie the symbiosis. Here, we identify and estimate relative abundance of proteins expressed by *Blochmannia chromaiodes*, the endosymbiont of the carpenter ant, *Camponotus chromaiodes*. This unculturable bacterium lives exclusively within the cytoplasm of ant cells. The endosymbiont is amenable for analysis for several reasons: genome sequence data are available for the specific *Blochmannia* strain used here (*B. chromaiodes*) as well as other *Blochmannia* genomes for comparison (the extremely closely related *B. pennsylvanicus*¹⁵, along with more distantly related *B. floridanus*¹⁴ and *B. vafer*¹⁶), genomic data is available for a congeneric host species *Camponotus floridanus*²⁷, and a wealth of information exists on the physiological and ecological significance of this ant-bacterial interaction, including the transcriptome analysis of *B. floridanus*.¹⁷ While simpler than multi-species microbial consortia living in an external environment, this clonal endosymbiont offers a valuable model to develop proteomics in an ecologically relevant yet very tractable microbial system.

We take advantage of data-independent analysis (DIA) techniques, which have emerged as powerful alternatives to overcome duty-cycle limitations of traditional data-dependent acquisition (DDA) approaches.^{28–33} Additional orthogonal separation strategies such as ion mobility spectrometry have been shown to improve the specificity of data independent acquisition fragment ion spectra.³⁴ In the MS^E implementation with ion mobility spectrometry, analysis is performed without precursor peptide isolation, and simultaneous qualitative/quantitative analysis is achieved by high sampling rate of both precursor peptide and fragment ions in a single analysis. High sampling frequency of DIA approaches allow accurate determination of peptide abundance by peak area measurement, thus enabling the use of the best ionizing peptides for individual proteins to estimate relative mole quantities of proteins in a sample using the so-called “Top3” or “best flier” method. This method has recently been extended to allow normalization to total protein (fmol/μg), for relative quantitation of proteins in mixed species matrices such as in obligate pathogens.^{35,36}

Specifically, we use DDA (MS/MS) and DIA (MS^E and HDMS^E) acquisition approaches for global qualitative protein identification and DIA (HDMS^E) for quantitative analysis. We report high coverage of the *Blochmannia chromaiodes* proteome (330 proteins detected, or 54% of the predicted proteome) and relative quantitation (fmol/μg) for 309 of these *B. chromaiodes* proteins. In addition, we identified 244 and quantified 161 ant host *Camponotus* proteins, presumably in the ant midgut. Our results demonstrate robust, highly reproducible quantitative results. This development of proteomics in *Blochmannia* sets a platform to investigate functional plasticity in this ant-bacterial association and, more broadly, to clarify symbiotic interactions between animals and unculturable microbes.

Experimental Section

Experimental material and processing for proteomics

A colony of the carpenter ant *Camponotus chromaiodes* was collected in April, 2011 from a wooded area south of Lake Raleigh, Raleigh, North Carolina, USA, GPS: N 35° 45' 40.6", W 78° 40' 53.5". The colony was maintained in the lab for just 48 hours before sample preservation below. During that short time in the lab, the colony was reared on a modified Bhatkar's diet supplemented with German cockroaches. The incubator was kept at a 16 h:8 h

light:dark photoperiod with temperatures of 22°C during the light period and 17°C during the dark period.

Three independent replicate experiments were conducted on minor workers taken from the collected colony. For each replicate, the midgut without any attachment of the Malpighian tubes and fat bodies was dissected out and pooled from six minor workers. Briefly, the midgut sac was cut open, gut content was emptied out, and the empty midgut was rinsed with ice cold PBS (Roche, Germany) containing complete protease inhibitor (Roche, Germany) at a concentration of 1 tablet per 50 mL PBS. Six cleaned midguts were pooled into a 1.5 mL microcentrifuge tube kept on ice during dissection. The resulting midguts sample was homogenized lightly using a blue pestle Eppendorf homogenizer in 100 μ L PBS inside a 1.5 mL microcentrifuge tube. Volume of homogenate was adjusted to 2 mL with PBS and filtered through 20 μ M nylon filter (Steriflip, Millipore, USA) with a vacuum pump. This filtration removed most material larger than *Blochmannia* cells, which are rod-shaped and ~10 μ m long. The filtrate (containing *Blochmannia* cells) was washed twice with 10 mL PBS by centrifuging at 1500 $\times g$ for 4 min at 4°C. The supernatant was discarded. The bacterial cell pellet was collected and diluted in 200 μ L PBS and stored at -80°C for protein preparation and proteomics analysis.

Proteome analysis by LC-MS/MS

The three bacterial cell samples (each from a pool of six ant midguts) were removed from -80°C and thawed on ice. After thawing, samples were centrifuged at 21,000 rcf at 4°C for 5 minutes. Supernatant was removed and pellets were washed once with 200 μ L of 50 mM ammonium bicarbonate. Wash layer was discarded, 100 μ L of 0.25% w/v RapiGest surfactant (Waters Corporation, USA) in 50 mM ammonium bicarbonate was added to pellet layer, and samples were heated at 50°C for 5 minutes. Samples were probe sonicated at medium power, 3 bursts each for 5 seconds per burst, storing on ice between bursts. Pre-digestion concentrations were measured by bicinchoninic acid (Pierce micro-BCA, USA) total protein assay, and total protein recovery was 22–29 μ g across the replicates. Samples were normalized to 19 μ g total protein in 86.4 μ L with 0.25% RapiGest, and were reduced, alkylated, and digested following a standard in-solution digestion protocol (<http://www.genome.duke.edu/cores/proteomics/sample-preparation/>). Briefly, the samples were reduced in 10 mM dithiothreitol (VWR, USA) alkylated in 20 mM iodoacetamide (Calbiochem, USA), and digested with 0.38 μ g sequencing grade modified trypsin (Promega, USA). After digestion at 37°C overnight, the samples were spiked with the surrogate standard, 25 fmol of yeast alcohol dehydrogenase digest (MassPrep, Waters Corporation, USA) per μ g of total protein. They were then acidified to 1% trifluoroacetic acid/2% acetonitrile, and heated for 2 hours at 60°C to hydrolyze Rapigest. The protein digests were then taken to dryness under vacuum and reconstituted to 1 mg/mL in 200 mM ammonium formate, pH 10. Samples were then pipetted into total recovery LC vials (Waters Corporation, USA).

LC-MS operation

Three microliters (3 μ g) of peptide sample digests were analyzed using a nanoAcquity UPLC system with 2D Technology coupled to a Synapt G2 HDMS mass spectrometer (Waters Corp, Milford, MA, USA), using a method which generated five fractions at pH 10 with approximately equal loading in each fraction, similar to previously described.^{35,37–39} The sample was first trapped at 2 μ L/min at 97/3 v/v water/acetonitrile on a 5 μ m XBridge BEH130 C18 300 μ m $\times$ 50 mm column. Two-minute steps at 2 μ L/min to the following % acetonitrile were used to generate the 5 fractions at pH 10, with the composition returning to 97/3/0.1 v/v/v water/acetonitrile/formic acid during the analytical second dimension analysis: 10.8%, 14.0%, 16.7%, 20.4%, and 65.0%. At each fraction step, the flow-through

from the first dimension was diluted online 10-fold with 99.8/0.1/0.1 v/v/v water/acetonitrile/formic acid and trapped on a 5 μm Symmetry C18 300 $\mu\text{m} \times 180$ mm trapping column. Separations for each fraction were then performed on a 1.7 μm Acquity BEH130 C18 75 $\mu\text{m} \times 150$ mm column (Waters Corp, Milford, MA, USA) using a 37-min gradient of 7 to 35% acetonitrile with 0.1% formic acid at a flow rate of 0.5 ml/min and 35°C column temperature. To compare MS data acquisition methods, for each sample we conducted one five-fraction data-independent analysis using ion mobility (HDMS^E) analysis, one five-fraction data-independent analysis without ion mobility (MS^E), and one five-fraction traditional data-dependent analysis (DDA) analysis. The HDMS^E runs used 0.6 sec cycle time alternating between low collision energy (6 V) and high collision energy ramp in the MS Triwave transfer region (27 to 50 V). The MS^E runs used 0.6 sec cycle time alternating between low collision energy (6 V) and high collision energy ramp in the MS Triwave trap region (15 to 40 V). The DDA mode utilized a 0.6 sec MS scan followed by MS/MS acquisition on the top 3 ions with charge greater than 1 using a collision-energy derived from a m/z and charge-state lookup table. MS/MS scans for each ion used an isolation window of approximately 2.3 Da, 0.6 second scans with a maximum of 3 seconds per precursor, and dynamic exclusion for 120 seconds within 1.2 Da of the selected precursor m/z.

LC-MS data processing

Prior to this study, we fully sequenced and annotated the 791-kb genome of *Blochmannia chromaiodes* (CP003903) and 13 coding sequences of the *Camponotus chromaiodes* mitochondrial genome (JX966368). A custom database was created by combining NCBI's *Camponotus floridanus* database with these newly obtained data for predicted *B. chromaiodes* (endosymbiont) proteins and *C. chromaiodes* mitochondrial proteins (15415 total entries). The sequence for ADH1_YEAST from www.uniprot.org was also added to the database, since this protein was used as standard surrogate standard. The database was appended with a 1X reverse database for peptide false discovery rate determination. The combined database in .fasta format, containing 30832 nonredundant entries, has been uploaded to the Tranche database (www.proteomecommons.org/tranche). Common contaminants were not included in the database; however, an initial search of the DDA data the Swissprot database (www.uniprot.org) with all taxonomies showed very few matches to exogenous proteins. MS/MS spectra generated by DDA were compiled into .mgf format using Mascot Distiller v2.2 and submitted to the Mascot v2.2 (Matrix Sciences, Inc) search engine. For HDMS^E and MS^E data, ProteinLynx Global Server (PLGS) 2.5 (Waters Corporation, USA) was used to generate searchable files, and were then submitted to the IdentityE search engine (Waters Corporation, Milford, MA, USA). The automated merge function within PLGS 2.5 was used to combine spectral information from each of the five fractions prior to search in IdentityE. Precursor ion mass tolerance was 10 ppm and 5 ppm for Mascot and PLGS searches respectively, and product ion tolerance was 0.04 Da for Mascot and 13 ppm for PLGS. For both search engines, enzyme specificity was set to tryptic and a maximum of 2 missed cleavages were allowed. Carbamidomethyl cysteine was included as a fixed modification, and variable modifications included oxidized methionine and deamidated asparagine and glutamine. Specific search parameters for the PLGS searches were as follows: requirement of at least 3 products per precursor, 7 product ions per protein, and 1 peptide per protein, at 100% False Positive Rate. The MS/MS, MS^E, and HDMS^E data were then imported in Scaffold v3.1 (Proteome Software, Inc) for curation. Using a target-decoy database search strategy for false discovery rate (FDR) determination, the overall FDR for the aggregate result set (DDA, MS^E, and HDMS^E) was determined to be 1.2% at the peptide level (99% confidence interval), selecting a minimum protein confidence of 90%, 2 peptides to match, and a minimum peptide confidence of 50% in Scaffold v3.1.^{40,41} This threshold was set initially such that as many peptides to each protein

could be used as possible for quantitation, to improve quantitative accuracy. Peptide-to-protein ambiguity was assessed within the Scaffold software. Only 5 out of 330 *B. chromaiodes* proteins were flagged as having protein grouping ambiguity within Scaffold, with a total of 7 peptides which could not be confidently assigned to one species or the other (0.1%). The full set of tandem mass spectra for each biological replicate has been made available as a Scaffold 3.1 file at https://discovery.genome.duke.edu/express/resources/2698/Wernegreen_Supplemental_Spectra.sf3, and additionally at www.proteomecommons.org/tranche; this file can be utilized by the user to additionally curate the data to either lower or higher peptide and protein FDR. The summarized number of peptides for each of the 633 proteins, per sample and analysis approach, is also summarized in Supplemental Table 1.

Estimation of species-specific protein abundance

For the HDMS^E analysis, the previously described method of Silva et al.³⁰ was utilized for protein-level quantitation, with specific modifications previously described and validated in a complex model system in order to correct for the sample containing the proteomes of multiple species.³⁵ Area-under-the-curve measurements for the high resolution accurate mass selected ion chromatograms were generated in PLGS 2.5 by processing the data with a minimum low energy threshold of 100, minimum high energy threshold of 10. Lockmass correction with Glu-Fibrinopeptide B (785.8426 (M+2H)²⁺) was enabled. After database search as described above, femtomol (fmol) values for each protein were determined in PLGS 2.5 (Waters Corporation, USA) by assigning the internal standard, ADH1_YEAST, as having 75 fmol on-column per injection. Utilizing the PLGS output from each file, additional minor calculations described here were performed in Excel 2010 (Microsoft). In PLGS 2.5, the method uses the intensity average of the top 3 peptides to ADH1_YEAST to calculate an intensity/fmol scaling factor, and then applies this factor to the Top3 average for the other proteins to calculate a fmol value for every protein confidently identified in the sample. PLGS also provides nanogram quantities for each protein, which is simply the product of calculated fmol measurement and the molecular weight of each protein in the database. We have recently introduced a simple modification to this approach that allows it to be applied to mixed-species proteomes, in order to measure relative protein quantity only in the species of interest, in the units of fmol/ μ g.³⁵ Without this correction, samples which have mixed proteomes (such as endosymbionts or obligate pathogens) may display wide variability in apparent abundance of some proteins, simply due to variation in the interference from the contaminating species. We applied this approach to the *Blochmannia* and *Camponotus* dataset. For quantitation, results were restricted to the HDMS^E datasets (n=3) only, for those proteins having two or more peptides in at least two of the three biological replicates according to the Scaffold-curated results. Whereas the initial protein FDR including all results (DDA, MSE and HDMS^E) as shown in our Scaffold file was approximately 10% (1% peptide FDR), utilizing this approach we reduced the protein FDR to 0.6% for the quantitative dataset (3 reverse protein identifications out of 477 total), with all proteins having at least one peptide of very high quality (0.95 PeptideProphet probability). For each replicate, the fmol of each bacterial protein from PLGS (309 total) was divided by the sum ng for all bacterial proteins, and multiplied by a scaling factor of 1000 to convert ng to μ g. Raw protein quantitation outputs from PLGS 2.5 are reported in Supplemental Tables 9–11, and the average and standard deviation for each of the proteins the Scaffold-curated protein list are reported in Supplemental Table 1. This simple ratio calculation results in relative protein abundance expressed in fmol/ μ g for each protein, specific to the bacterial endosymbiont proteome (Supplemental Table 2). The same analysis was performed for ant host proteins, except results were reported simply in fmol on-column (Supplemental Table 3).

Protein ortholog detection

For comparisons across species, orthologs of *B. chromaiodes* proteins were identified in *B. floridanus* (NC_005061), *Buchnera* APS chromosome (NC_002528) and plasmids (NC_002252, NC_002253), and *E. coli* (NC_000913). Specifically, the Reciprocal Sequence Distance algorithm⁴² was used to identify the reciprocal best BLAST hits between translated open reading frames. This program uses blastp to identify potential matches of a given translated gene, aligns all potential matches using ClustalW⁴³, and calculates a maximum likelihood estimation of amino acid substitutions between proteins by invoking PAML 4.2.⁴⁴ The protein with the lowest divergence was then compared to the first genome using blastp, followed by the alignment and divergence calculations. If the protein match with the lowest divergence was the same as the original query sequence, the pair was considered orthologous and the divergence was retained in the output. We used stringent criteria of retaining only blast matches of $E < 10^{-10}$, and requiring that the alignable region of two sequences exceed 80% of the alignment's total length. These stringent criteria may generate false negatives, by excluding true orthologs that are highly divergent. The orthologs identified here are included in Supplemental Table 4.

Functional classification of quantified proteins

MultiFun classification, developed for *E. coli*⁴⁵, was slightly modified (see below) and used to classify proteins in *B. chromaiodes* based on their corresponding ortholog in *E. coli*. We applied the same scheme to *Buchnera*, in order to compare the functional composition of the *Blochmannia* proteome with that of the recently-published *Buchnera* proteome (i.e., normalized spectral abundance factor, NSAF, results in Bu-2 sample)²⁴. Given that chaperonins are known to be highly expressed in many endosymbionts, we followed the precedent of Poliakov et al.²⁴ and established a distinct category for Chaperones/Adaptation to Stress (MultiFun 2.3.4 and 5.5, respectively). The remaining broad categories included Metabolism (MultiFun category 1), Information transfer (MultiFun category 2, excluding 2.3.4), Regulation (MultiFun category 3), Transport (MultiFun category 4), Cell processes (MultiFun category 5), Cell structure (MultiFun category 6), Other (no *E. coli* homolog or unknown molecular function). We included Extrachromosomal origin (MultiFun category 8) within "Other." (Three *Blochmannia* proteins were assigned to the 'Extrachromosomal origin' based on their orthology to *E. coli* proteins; however, these genes are chromosomally encoded in *Blochmannia*, which lacks any plasmids.) Many proteins have multiple cellular roles and are assigned to more than one category. However, some comparisons required that we assign each protein to just one of the broad categories above. Following the order in which functional categories are listed above (Chaperones/Adaptation to Stress, Metabolism, Information transfer, etc.), proteins were assigned to the first relevant category listed, and were not assigned to any subsequent category (Supplemental Table 4).

In addition to the broad functional categories above, we also considered subcategories of Metabolism (MultiFun category 1) that are particularly relevant to symbiont nutritional functions. These subcategories included: amino acid biosynthesis (1.5.1), fatty acid biosynthesis (1.5.4), nucleotide biosynthesis (1.5.2), cofactor and small molecule carrier biosynthesis (1.5.3), sulfur metabolism (1.8.2), and nitrogen metabolism (1.8.3).

Proteome-transcriptome comparisons

We compared the quantitative data for the *B. chromaiodes* proteome obtained here to published data for highly expressed *B. floridanus* transcripts.¹⁷ The publicly available transcriptome data only includes the 53 genes representing the 10% most abundant transcripts, based on pooled data across various life stages of the ant host *C. floridanus*, with transcript abundances expressed as the median-based trimmed mean density value of each macroarray spot (sMTM). Despite the use of different measurement units (sMTM for

transcripts vs. fmol/ μ g for proteins) and sampling of distinct *Blochmannia* species, broad comparisons are possible. Among the 53 genes with high transcript levels in *B. floridanus*, we quantified 45 corresponding proteins in the *B. chromaiodes* proteome. We then tested whether, in our *B. chromaiodes* dataset, these 45 proteins had atypically high fmol/ μ g quantities using a Wilcoxon test (JMP Pro 9.0).

Safety Considerations

Proper sample handling precautions in a BSL2 environment were utilized in the preparation and analysis of these samples.

Results

HDMS^E improved depth of proteome coverage compared to DDA or MS^E

To assess the relative performance of DDA, MS^E (DIA without ion mobility) and HDMS^E (DIA with ion mobility), the three replicate bacterial samples were subjected to analysis with each method. A total of 633 proteins and 6837 peptides were identified with an overall 1.2% FDR for both ant and bacterial peptides (Supplemental Table 1). The protein-level FDR with this first pass curation of all data types was 9% (Supplemental Table 1), but is reduced to 3% (514 proteins) when one considers proteins observed in at least two of the 9 runs. Among these identified proteins, 330 proteins were from the bacterial proteome. The three modes of acquisition showed substantial differences in bacterial proteome coverage (Figure 1). Overall, HDMS^E showed the highest proteome coverage with 329 bacterial proteins (54% of the proteins annotated in the genome), while DDA and MS^E identified only 213 and 212 bacterial proteins, respectively (Figure 1a). Only one unique protein (30S ribosomal protein S14) was confidently identified in all three biological samples solely by DDA. A relatively small number of peptides were uniquely identified by DDA (153 peptides) or MS^E (474 peptides), whereas HDMS^E provided a high degree of additional coverage (2373 peptides) (Figure 1b). Most of these unique peptides were identified consistently for each method, suggesting that differences in fragmentation characteristics between the analysis methods are likely a play, rather than simple undersampling. Combination of DDA, MS^E, and HDMS^E using Scaffold software resulted in a total of 330 detected *B. chromaiodes* proteins (Supplemental Table 1) or 54% coverage of the predicted proteins in the proteome, however the mobility-assisted DIA approach (HDMS^E) obtained all but one of these proteins on its own, suggesting the DDA and MSE approaches are only of limited complementarity at the protein-level. Improved proteome coverage observed when adding an ion mobility separation to a data-independent workflow is consistent with previous observations of improved identification specificity with this additional separation dimension.³⁴

Relative quantitation of proteins within the *B. chromaiodes* proteome

Among the 329 *B. chromaiodes* proteins identified by HDMS^E, 309 were identified at least in two of the three biological replicates and with more than two peptides. For these 309 proteins, the quantity was estimated as fmol/ μ g *B. chromaiodes* protein, allowing comparison between individual proteins and classes of proteins. This value spanned nearly three orders of magnitude, from 1442.2 ± 65.8 to 3.9 ± 3.5 fmol per μ g (Figure 2; Supplemental Table 2). The most abundant *B. chromaiodes* protein was the chaperone GroEL (1442.2 ± 65.8 fmol/ μ g of *B. chromaiodes* protein, or 6% of the total *B. chromaiodes* protein abundance). GroES, the co-protein of GroEL, was also relatively abundant at 162.4 ± 17.2 fmol/ μ g (1%). Other abundant proteins classified as chaperonins or adaptation to stress were the heat shock protein IbpA (874 ± 73.9 fmol/ μ g, or 4%), protein chain elongation factor Tuf (373.1 ± 2.9 fmol/ μ g, or 2%), and heat shock protein 70 family members DnaK (179.6 ± 6.5 fmol/ μ g, or 1%) and DnaJ (188.6 ± 15 fmol/ μ g, or 1%).

Additional quantified *Blochmannia* proteins ranking in the top ten most abundant were lipoprotein Lpp1 (1305.1 ± 28.1 fmol/ μ g, or 5%), outer-membrane protein OmpC (1195.8 ± 130.4 fmol/ μ g, or 5%), the regulation protein HflC (467 ± 32.6 fmol/ μ g, or 2%), 30S ribosomal subunit proteins RpsC (349.7 ± 27.8 fmol/ μ g, or 1%) and RpsG (308 ± 20.6 fmol/ μ g, or 1%), RNA polymerase RpoA (344.4 ± 41.4 fmol/ μ g, or 1%), and Gnd involved in the pentose phosphate pathway (298.5 ± 9.6 fmol/ μ g, or 1%).

Reproducibility of sample preparation and analysis

The three, independent *Blochmannia* preparations yielded very similar results, indicating a high reproducibility of the ant midgut preparation, protein purification, and proteomic analysis. Across the three samples, the sum of *Blochmannia* proteins was 17030 ± 4998 fmol. *Blochmannia* protein comprised $76.6\% \pm 0.9\%$ of the total (ant and *Blochmannia*) protein (Figure 3a). This low ($<1\%$) RSD across the three biological samples illustrates that nearly identical ratios of host and endosymbiont protein were obtained for each preparation. In addition, the three preparations yielded very similar quantitative values for individual endosymbiont proteins. Of the 309 quantified *Blochmannia* proteins, 34% had less than 10% RSD, 80% had less than 30% RSD, and 93.2% of proteins had less than 50% RSD (Figure 3b), showing excellent biological and technical reproducibility of the approach. Individual host proteins also showed reproducible relative quantitation across the three independent samples, though with overall higher %RSD than *Blochmannia* proteins (Supplemental Table 3). This is consistent with previous our observation that the proteome making up the minor component of the mixed proteome matrix will show slightly higher variability in the quantitative values for individual proteins.³⁵ Overall, the excellent observed analytical reproducibility of the purification and relative quantitation method is critical to future studies looking into other aspects of endosymbiosis in this model system.

Transcriptome vs. proteome

The *B. floridanus* transcriptome analysis made publicly available the sMTM data for the 10% most highly expressed endosymbiont genes (53 genes).¹⁷ Among the *B. chromaiodes* proteins corresponding to these genes, we identified and quantified 85% (45 out of 53 proteins) in the *B. chromaiodes* proteome (Figure 2, Supplemental Figure 1). Among the eight not quantified in the proteome, Rpe and YcbL were identified. The remaining six are encoded by the *B. chromaiodes* genome but were not identified in the proteome, possibly expressed but falling below the detection limit. The median abundance of these 45 quantified proteins was high, at 107 ± 4.1 fmol/ μ g, compared to a median value of 34.3 ± 5.3 fmol/ μ g for the remaining proteins. A Wilcoxon test indicates this difference in abundance is highly significant ($p < 0.000001$). Consistent with this result, the distribution of protein abundances illustrates that these 45 proteins are indeed skewed toward high quantities and include some of the most abundant proteins (Figure 2). Notably, GroEL and GroES ranked #1 and #2 most abundant transcripts in *B. floridanus*¹⁷, and their corresponding proteins were also abundant in the *B. chromaiodes* proteome (ranked #1 and #37, Supplemental Table 2).

However, among these 45 genes, rank order of abundances of *B. floridanus* transcripts and *B. chromaiodes* proteins showed no significant association (Supplemental Figure 1). Moreover, some of the most abundant proteins (e.g., Lpp1, IbpA, HflC) were not among the top 10% most abundant *B. floridanus* transcripts. These differences may reflect discrepancies between transcript versus protein abundances in *Blochmannia*. However, they may also reflect significant differences in the types of samples used, including differences in ant host species (*C. chromaiodes* vs. *C. floridanus*), colony environment (field-collected vs. lab-reared), or life stages sampled (workers only in this study, versus averaged across brood and worker stages).

Contribution of functional groups to genome and proteome

All identified and quantified *B. chromaiodes* proteins were assigned to one broad functional category based on a slight modification of the MultiFun classification schema⁴⁵ (Figure 4, Supplemental Table 4). For each category, the proportion of proteins identified in the proteome was quite similar to the proportion of annotated proteins in the *B. chromaiodes* genome (Figure 4a, b). In this sense, the functions detected in the proteome roughly reflect those encoded by the genome. Metabolism and information transfer together accounted for the majority in both datasets (75% of the annotated proteins in the genome and 82% of the proteins identified in the proteome). However, particular functional groups had disproportionately high protein abundance, based on quantitative analysis. Most notably, chaperones and other stress-related proteins accounted for only 5–6% of the number of identified proteins (Figure 4b), but contributed more than twice that to protein abundance (15%; Figure 4c), reflecting the high levels of many chaperonin proteins. Table 1 shows comparisons for chaperonins and stress adaptation proteins, as well as subcategories of Metabolism that are particularly relevant to symbiont nutritional functions.

Our re-analysis of the *Buchnera* proteome data²⁴ revealed similar trends in the contribution of functional categories to protein abundance (albeit measured by a different quantitative approach, spectral-counting by NSAF) (Figure 4d). Metabolism and information transfer accounted for a large fraction (42% and 24%, respectively) of the total protein abundance in *Buchnera* (Figure 4d), a value comparable to that in *Blochmannia* (40% and 30%, respectively) (Figure 4c). Chaperonins accounted for 22% of protein abundance in *Buchnera* and (as noted above) 15% in *Blochmannia*. In *Blochmannia*, transport proteins constitute a greater proportion of protein abundance (8%) compared to the proportion of proteins detected (3%) (Figure 4c vs. 4b). By contrast, in *Buchnera*, transport proteins constitute a similar proportion of protein abundance (2%; Figure 4d) and the proportion of proteins detected (3%; Supplemental Table 4). The high abundance of transport proteins in *Blochmannia* reflects the unusually high levels of OmpC, which accounted for about 5% of total protein abundance. OmpC is an outer-membrane porin protein that is involved in the passive transport of small water-soluble molecules across the outer membrane.

In addition to broad similarities in functional composition, the *Blochmannia* and *Buchnera* proteomes show a significant correlation in abundances of individual proteins. Of the 402 proteins we detected as orthologs shared by *Buchnera*-APS and *Blochmannia*-C. *chromaiodes*, 213 were quantified in both proteomes (as NSAF and fmol/μg, respectively). While the two methods of quantification differ, the rank order of protein abundance was significantly associated (Spearman Rho; $r_s=0.415$; $p<0.0000000003$).

Symbiont nutritional functions

Previous experimental analysis of the *Blochmannia*-*Camponotus* symbiosis has shown the bacteria supply nutrients such as essential amino acids to their host insects, with important implications for host fitness.⁹ Of the 56 *B. chromaiodes* annotated genes involved in amino acid biosynthesis, we identified 46 of the corresponding proteins in the proteome (Table 1, Table 2, Supplemental Table 5). This 82% coverage of amino acid biosynthetic proteins exceeds the overall 54% proteome coverage. Though *Blochmannia* lacks the ability to synthesize arginine, two enzymes in the initial steps of arginine synthesis pathway, carbamoyl-phosphate synthetase (CarAB) and ornithine carbamoyl-transferase (ArgI), are retained in the genome and were identified in the proteome. This data further supports the previously proposed hypothesis based on genome sequence data, that enzymes of both *Blochmannia* and the host contribute to arginine synthesis.¹⁴

In addition, quantitative analysis showed that proteins involved in amino acid biosynthesis proteins totaled 12% of *Blochmannia* total protein (Table 1), a significant component within the broader category of Metabolism. Proteins contributing to amino acid synthesis ranged from 295.8 ± 20.6 (MetE for methionine synthesis) to 10.6 ± 2.2 fmol/ μ g (LysA for lysine synthesis) (Table 2, Supplemental Table 5). Other high-abundance amino acid biosynthesis enzymes included ThrA for threonine biosynthesis and TyrA for tyrosine biosynthesis.

Almost all *Blochmannia* proteins in the sulfate assimilation pathway were identified (Table 1). The sole exception, CysH, is encoded by the *B. chromaiodes* genome but was not detected in the proteome (see Discussion). High coverage of this pathway in the proteome is consistent with the view that *Blochmannia* can reduce sulfate *via* the adenosine-5'-phosphosulfate (APS) - 3'-phosphoadenosine-5'-phosphosulfate (PAPS) pathway (Figure 5).

B. chromaiodes encodes 13 genes with a predicted function for fatty acid biosynthesis. Eight of these proteins were quantified here, or 62% coverage of fatty acid biosynthesis proteins (Table 1, Supplemental Table 6). The estimated quantity of these proteins ranged from 93.3 ± 7.2 fmol/ μ g (FabI) to, 13.9 ± 2.1 fmol/ μ g (FabH), together accounting for ~2% of the total *Blochmannia* protein abundance. Of the 14 annotated proteins involved in nucleotide (purine and pyrimidine) biosynthesis, ten were detected and nine quantified, accounting for ~2% of the total *Blochmannia* protein abundance (Table 1, Supplemental Table 7).

The discovery of a urease gene cluster (UreABCDGF) in the *Blochmannia* genome indicated that the bacteria contribute to nitrogen recycling.¹⁴ Urease activity leads to production of ammonium from hydrolysis of urea, which can be incorporated into amino acids by glutamine synthetase (GlnA). Within the urease gene cluster, we identified only the accessory protein UreG in the proteome, at 30.1 ± 5.1 fmol/ μ g (Supplemental Tables 2 and 4). (Likewise, *ureG* was the only *B. floridanus* urease gene represented within the top 10% most abundant transcripts.¹⁷) We identified and quantified the GlnA protein at a relatively high abundance of 65.8 ± 11.9 fmol/ μ g. SpeB, an enzyme involved in the urea production, was also identified and quantified at 26.2 ± 21.4 fmol/ μ g. This corroborates the prediction that *Blochmannia* assimilates nitrogen from ammonia to glutamine and participate in other amino acid biosynthesis.

Ant host proteins

A total of 244 proteins matching the ant *C. floridanus* database were detected and contained at least two identified peptides. Of these, 161 proteins were detected by HDMS^E in at least two biological replicates in addition to the more than two peptides criterion. The abundance of these ant proteins were measured directly as fmol on-column using the ADH1_YEAST surrogate standard, and they spanned two orders of magnitude, ranging from approximately 166 to 2 fmol (Supplemental Table 3). The most abundant host proteins were actin (166 ± 50.6 fmol), a subunit of cellular filaments, and transferrin (151 ± 58.0 fmol), a protein that is involved in iron homeostasis, increases upon bacterial challenge, and may play a role in insect immunity (see Discussion).^{46, 47} Other abundant *Camponotus* proteins included the ATP synthase alpha subunit (147 ± 47.3 fmol), an ADP, ATP carrier protein (105 ± 25.2 fmol), and pericentrin (105 ± 25.2 fmol), a component of the centrosome. Well-studied but lower-level (<25 fmol) host proteins quantified included juvenile hormone epoxide hydrolase, which participates in the degradation of juvenile hormone; apolipoproteins which transport lipids including hydrocarbons; ubiquitin, a small regulatory protein that is involved in protein recycling; and cytochrome P450, which catalyzes the oxidation of organic compounds including many insecticides.

Discussion

A major purpose of this study was to quantify the global proteome of the unculturable bacterium, *Blochmannia*. Using a gel-free and label-free proteomics approach, we confidently identified 330 of the 609 annotated proteins (54% proteome coverage). Other, unidentified bacterial proteins may either be absent from the minor ant workers sampled, or may have fallen below our limit of detection. As discussed below, this first insight into the *Blochmannia* proteome highlights points of congruence and departure from transcript-based studies, sheds light on the nutritional underpinning of the symbiosis, demonstrates exceptionally high GroEL expression, and allows direct comparisons among protein quantification methods. As a complement to existing genomic and transcriptomic tools, this development of a new proteomics approach offers a powerful strategy to investigate nutritional interactions and functional responses of a highly specialized, unculturable mutualist.

Correlations of genome, transcriptome and proteome

While gene expression is routinely quantified by measuring mRNA levels, experimental studies show that abundances of transcripts and their corresponding proteins are not always positively correlated.^{21, 22, 48, 49} The fraction of protein variation that can be explained by mRNA abundance was just 37% in a recent mouse experiment⁴⁸, and even lower in other studies.^{50, 51} Despite this potential decoupling of protein and transcript levels, we found shared patterns between the quantitative proteomics analysis here and the previous microarray-based transcriptome profile for *B. floridanus*.¹⁷ Among the 10% most highly expressed genes in *B. floridanus*, all but six (of the 53) genes were detected in the *B. chromaiodes* proteome and many had high relative protein abundance. This general congruence illustrates that, at least among very highly expressed genes, the approaches identified similar gene products.

Dynamic range is perhaps the most striking difference between the *Blochmannia* transcriptome and proteome. While the range of transcript abundance spanned less than one order of magnitude¹⁷, we found that levels of proteins ranged three orders of magnitude, nearly the maximum range possible for the mass spectrometer using the “best flier” protein quantity estimation approach. It has been reported in mammals that proteins are five times more stable than mRNAs, about 900 times more abundant than the mRNAs used to make them, and that cellular abundance of proteins is predominantly controlled at the level of translation.⁴⁹ Whether this applies to bacterial endosymbionts remains unknown, but if so, we should expect imperfect associations between transcript and protein abundances, as well as wider dynamic ranges for protein abundance compared to mRNA.

Insights into endosymbiont metabolism

Genomic and experimental data indicate that *Blochmannia* is a nutritional symbiont whose core functions include biosynthesis of essential amino acids and nitrogen recycling.^{9,14,15} The *Blochmannia* proteome included almost all enzymes annotated for essential amino acid biosynthesis, further confirming this key symbiont function. Complete pathways for nine out of ten (except for Arginine) essential amino acids were identified in the proteome and almost all enzymes have been quantified. Based on data collected here, the abundance of biosynthesis enzymes also did not vary substantially among different essential amino acids (Table 2), a pattern also found in the *Buchnera* proteome.²⁴

Our ability to detect and estimate quantity of most *Blochmannia* enzymes involved in the APS-PAPS pathway of sulfate assimilation underscores the significance of this symbiont function. CysH, or PAPS reductase, was the only protein in this assimilatory pathway that

we did not detect. This lack of detection indicates that CysH fell below our detection limit, due to low protein abundance or perhaps its absence entirely. One possible explanation for low abundance of CysH may be that the reduction of PAPS to sulfite is a rate-limiting step and highly regulated. Because *cysJIH* genes occur in the same operon in *Blochmannia* and other bacteria, regulation leading to different protein abundances likely occurs post-transcriptionally. Notably, studies in plants have shown that sulfite is toxic to the cell^{52,53}, and the enzyme that generates sulfite (APS reductase in the plants studied) is rate-limiting, highly regulated, and a proposed “control point” for cysteine synthesis.^{53,54} While speculative, it is possible that CysH may be the rate-limiting step of sulfate assimilation in *Blochmannia* and, like APS reductase in plants, may be highly regulated.

Our detection of proteins involved in urea production and hydrolysis (UreG, GlnA, and SpeB) further supports the view that *Blochmannia* contributes to nitrogen recycling in the ant host.^{9,14,15,18} The retention of urease genes yet surprising loss of GlnA from another *Blochmannia* lineage, *B. vafer*, suggests variation in nitrogen recycling pathways among *Blochmannia*.¹⁶ In *B. vafer*, ammonia can be used as nitrogen donor for CarAB without glutamine and may participate the initial steps of arginine biosynthesis catalyzed by enzymes (CarAB and ArgI).¹⁶ Here, the presence of the *B. chromaiodes* CarAB and ArgI enzymes in our dataset (Supplemental Table 2) supports the existence of this alternative pathway.

GroEL overexpression

The high abundance of GroEL in *B. chromaiodes* is consistent with numerous studies showing constitutive over-expression of this chaperonin in bacterial endosymbionts.^{55–60} Notably, endosymbionts experience rapid rates of protein evolution and show signatures of the fixation of deleterious changes by genetic drift.^{61,62} As a result, their proteins tend to accumulate destabilizing amino acid substitutions that reduce the efficiency of protein folding and may lead to misfolding and aggregation.⁶³ GroEL overexpression in obligate endosymbionts may be a compensatory mechanism to cope with degraded protein structures.⁶¹ High expression of GroEL in insect endosymbionts may also relate to expanded roles of this chaperonin, including potential interactions with insect viruses during passage through the insect hemolymph (as shown in whiteflies^{64,65}) or its putative role as a histidine kinase sensor molecule in symbionts.⁶⁶

As expected from the co-transcription of overlapping *groES* and *groEL* genes, bacteria often show comparable levels of the two transcripts. For example, the *B. floridanus* transcriptome showed similar levels of *groES* and *groEL* (sMTM of 19.69 and 20.15, respectively). Likewise, the *Buchnera* transcriptome revealed comparable levels of the two transcripts under heat-shock and control conditions.¹⁹ At a protein level, we might expect a molar ratio of ~2:1 (GroEL:GroES), based on the structure of the GroESL complex.⁶⁷ However, in both the *Blochmannia* and *Buchnera* proteomes, GroEL is ~8-fold more abundant than GroES (*Buchnera* NSAF = 0.087:0.011²⁴, and *Blochmannia chromaiodes* fmol/μg = 1422.2:162.4 quantified here). The particularly high abundance of GroEL compared to GroES may relate to the distinct roles of this chaperonin in endosymbiotic bacteria (see above).

Ant host functions

Our quantitative estimation of numerous (161) *Camponotus* proteins provides, to our knowledge, the first shotgun proteomics data for this ant genus. We predict that the *Camponotus* proteins detected here occurred in tissues and cells associated with the midgut, including midgut epithelial cells and/or bacteriocytes. Specifically, our preparation method removed other host tissues in the first step of dissection and rinsing of the ant midgut. In addition, the high reproducibility of results across independent samples argues against

random contamination from other host tissues. However, we cannot rule out the possibility that host proteins from other tissues entered the sample, nor can we rule out the intriguing possibilities that some host proteins interact with surface proteins of *Blochmannia* cells or are actually imported into the bacterial cells.

Transferrin, a multifunctional glycoprotein involved in iron homeostasis (reviewed by Nichol et al.⁶⁸), was among the most abundant proteins we quantified. Several studies suggest that transferrin is involved in the insect innate immune response, perhaps by sequestering iron from bacterial associates. Specifically, transferrin expression increases upon bacterial infection in mosquito cells in culture⁴⁶, in *D. melanogaster*⁶⁹, in the tsetse fly⁴⁷, and in the silkworm.⁷⁰ In addition, the promoter region of *D. melanogaster* transferrin contains sequences that are known to bind nuclear factor-kappa B-like transcription factors that mediate immune responses.⁷¹ Although transferrin is not known to regulate mutualistic symbionts in insects, recent work highlights the significance of insect immune functions in regulating long-term, vertically transmitted bacterial endosymbionts. For example, an antimicrobial peptide (coleopteracin-A) in weevils selectively targets and regulates the growth of bacteriocyte-associated mutualists.⁷² Though speculative, it is possible that *Camponotus* transferrin may play a comparable role in regulating *Blochmannia* replication and preventing the spread of endosymbionts to nearby midgut cells.

Protein quantification methods

Proteomics has increasingly been adopted to quantify global changes in protein abundance between biological samples subjected to various conditions. Many studies use “bottom-up” shotgun proteomics with multidimensional liquid chromatography coupled with MS to separate peptides. Perhaps the most popular label-free method using DDA data compares the number of MS/MS spectra assigned to each protein, normalized to protein length (NSAF). Significant correlations have been shown between spectral counts and relative protein concentration with standard proteins.⁷³ The high (68%) coverage of the *Buchnera* proteome with nanoLC-LTQ-Orbitrap mass spectrometry using DDA acquisition confirms this traditional but very powerful method in proteomics research.²⁴ However, the dynamic range and quantitative precision of spectral counting is limited, especially for proteins with low spectral counts.⁷⁴ Compared to Poliakov et al.²⁴, we obtained lower proteome coverage using DDA, which may be explained by the fact that, unlike the *Buchnera* study, we used a gel-free approach and did not perform extensive sample pre-fractionation with 1D SDS-PAGE. We specifically avoided the use of SDS-PAGE gels because of the high quantitative variability (~50% CV) in extraction of proteotypic peptides from the gels, which makes relative quantitation between samples more difficult.

DIA label-free methods (such as MS^E and HDMS^E) typically utilize integrated LC-MS peak intensities of peptide ions to demonstrate the protein abundances in complex samples.^{25,26,35} Here, our comparison of DDA, MS^E and HDMS^E approaches clearly shows that HDMS^E out-performed both the DDA and MS^E on this instrument platform in identifying peptides and proteins. That is, all proteins identified by MS^E in *B. chromaiodes* were also identified by HDMS^E, and only one unique protein was identified by DDA alone. These data support the idea that high peak-capacity techniques such as multi-dimensional LC and ion-mobility separation step coupled with tandem mass spectrometry are quite useful in order to achieve a high coverage of target proteome such as the endosymbiont *Blochmannia* with gel-free and label-free approach. A similar quantitative method, although utilizing MS^E instead of HDMS^E, was successfully used in the human pathogen *Chlamydia trachomatis*.³⁵

In sum, studies of bacterial-insect symbioses have been bolstered by the availability genome sequences for the microbial and animal partners, making it possible to explore the functional interactions of both interacting organisms.⁷⁵ While previous functional genomic studies of

insect endosymbionts have largely focused on transcript-based approaches, protein abundances more directly reflect levels of active enzymes. Here, we presented a new, reliable, gel-free and label-free proteomic method to identify and quantify proteins with molar quantities in a challenging experimental system: an unculturable bacterial endosymbiont of ants. This approach yielded robust and biologically meaningful data regarding the relative quantities of expressed *Blochmannia* proteins. The excellent reproducibility among independent biological samples validates the utility of this approach. This work sets the stage for examining functional plasticity of this symbiont to environmental variation, including distinct life stages, diets, and experimental treatments of its ant host species. Clearer functional perspectives on this and other bacterial symbionts will provide new insights into the physiological and molecular mechanisms that shape host-microbe interactions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by grants to JJW from NSF (MCB-1103113) and NIH (R01GM062626) and by funding from the IGSP at Duke. We are grateful to Benoit Guenard and Andrea Lucky for pointing us to local field sites for studying and collecting ants.

References

1. Amann RI, Ludwig W, Schleifer KH. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol Rev.* 1995; 59(1):143–69. [PubMed: 7535888]
2. Cannon SA, Giovannoni SJ. High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. *Appl Environ Microbiol.* 2002; 68(8):3878–85. [PubMed: 12147485]
3. Su C, Lei L, Duan Y, Zhang KQ, Yang J. Culture-independent methods for studying environmental microorganisms: Methods, application, and perspective. *Appl Microbiol Biotechnol.* 2012; 93(3): 993–1003. [PubMed: 22189863]
4. Schneider T, Riedel K. Environmental proteomics: analysis of structure and function of microbial communities. *Proteomics.* 2010; 10(4):785–98. [PubMed: 19953545]
5. Keller M, Hettich R. Environmental proteomics: a paradigm shift in characterizing microbial activities at the molecular level. *Microbiol Mol Biol Rev.* 2009; 73(1):62–70. [PubMed: 19258533]
6. Wilmes P, Bond PL. Microbial community proteomics: elucidating the catalysts and metabolic mechanisms that drive the Earth's biogeochemical cycles. *Curr Opin Microbiol.* 2009; 12(3):310–7. [PubMed: 19414280]
7. VerBerkmoes NC, Denev VJ, Hettich RL, Banfield JF. Systems biology: Functional analysis of natural microbial consortia using community proteomics. *Nat Rev Microbiol.* 2009; 7(3):196–205. [PubMed: 19219053]
8. Buchner P. *Rev Eng.* Interscience Publishers; New York: 1965. Endosymbiosis of animals with plant microorganisms; p. xviip. 909
9. Feldhaar H, Straka J, Krischke M, Berthold K, Stoll S, Mueller MJ, Gross R. Nutritional upgrading for omnivorous carpenter ants by the endosymbiont *Blochmannia*. *BMC Biol.* 2007; 5:48. [PubMed: 17971224]
10. de Souza DJ, Bezier A, Depoix D, Drezen JM, Lenoir A. *Blochmannia* endosymbionts improve colony growth and immune defence in the ant *Camponotus fellah*. *BMC Microbiol.* 2009; 9:29. [PubMed: 19200360]
11. McCutcheon JP, Moran NA. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol.* 2012; 10:13–26. [PubMed: 22064560]

12. Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. Nature. 2000; 407(6800):81–6. [PubMed: 10993077]
13. Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T, Hattori M, Aksoy S. Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. Nat Genet. 2002; 32(3):402–7. [PubMed: 12219091]
14. Gil R, Silva FJ, Zientz E, Delmotte F, Gonzalez-Candelas F, Latorre A, Rausell C, Kamerbeek J, Gadau J, Holldobler B, van Ham RC, Gross R, Moya A. The genome sequence of *Blochmannia floridanus*: Comparative analysis of reduced genomes. Proc Natl Acad Sci U S A. 2003; 100(16): 9388–93. [PubMed: 12886019]
15. Degnan PH, Lazarus AB, Wernegreen JJ. Genome sequence of *Blochmannia pennsylvanicus* indicates parallel evolutionary trends among bacterial mutualists of insects. Genome Res. 2005; 15(8):1023–33. [PubMed: 16077009]
16. Williams LE, Wernegreen JJ. Unprecedented loss of ammonia assimilation capability in a urease-encoding bacterial mutualist. BMC Genomics. 2010; 11:687. [PubMed: 21126349]
17. Stoll S, Feldhaar H, Gross R. Transcriptional profiling of the endosymbiont *Blochmannia floridanus* during different developmental stages of its holometabolous ant host. Environmental Microbiology. 2009; 11(4):877–88. [PubMed: 19040455]
18. Zientz E, Beyaert I, Gross R, Feldhaar H. Relevance of the endosymbiosis of *Blochmannia floridanus* and carpenter ants at different stages of the life cycle of the host. Appl Environ Microbiol. 2006; 72(9):6027–33. [PubMed: 16957225]
19. Wilcox JL, Dunbar HE, Wolfinger RD, Moran NA. Consequences of reductive evolution for gene expression in an obligate endosymbiont. Mol Microbiol. 2003; 48(6):1491–500. [PubMed: 12791133]
20. Moran NA, Dunbar HE, Wilcox JL. Regulation of transcription in a reduced bacterial genome: Nutrient-provisioning genes of the obligate symbiont *Buchnera aphidicola*. J Bacteriol. 2005; 187(12):4229–37. [PubMed: 15937185]
21. Gygi SP, Rochon Y, Franza BR, Aebersold R. Correlation between protein and mRNA abundance in yeast. Molecular and Cellular Biology. 1999; 19(3):1720–1730. [PubMed: 10022859]
22. Griffin TJ, Gygi SP, Ideker T, Rist B, Eng J, Hood L, Aebersold R. Complementary profiling of gene expression at the transcriptome and proteome levels in *Saccharomyces cerevisiae*. Molecular & Cellular Proteomics. 2002; 1(4):323–333. [PubMed: 12096114]
23. Filiatrault MJ. Progress in prokaryotic transcriptomics. Curr Opin Microbiol. 2011; 14(5):579–86. [PubMed: 21839669]
24. Poliakov A, Russell CW, Ponnala L, Hoops HJ, Sun Q, Douglas AE, van Wijk KJ. Large-scale label-free quantitative proteomics of the pea aphid-*Buchnera* symbiosis. Mol Cell Proteomics. 2011; 10(6):M110 007039. [PubMed: 21421797]
25. Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, Resing KA, Ahn NG. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. Mol Cell Proteomics. 2005; 4(10):1487–502. [PubMed: 15979981]
26. Blackburn K, Mbeunkui F, Mitra SK, Mentzel T, Goshe MB. Improving protein and proteome coverage through data-independent multiplexed peptide fragmentation. J Proteome Res. 2010; 9(7):3621–37. [PubMed: 20450226]
27. Bonasio R, Zhang G, Ye C, Mutti NS, Fang X, Qin N, Donahue G, Yang P, Li Q, Li C, Zhang P, Huang Z, Berger SL, Reinberg D, Wang J, Liebig J. Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. Science. 2010; 329(5995):1068–71. [PubMed: 20798317]
28. Venable JD, Dong MQ, Wohlschlegel J, Dillin A, Yates JR. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. Nat Methods. 2004; 1(1):39–45. [PubMed: 15782151]
29. Silva JC, Gorenstein MV, Li GZ, Vissers JP, Geromanos SJ. Absolute quantification of proteins by LCMSE: A virtue of parallel MS acquisition. Mol Cell Proteomics. 2006; 5(1):144–56. [PubMed: 16219938]

30. Silva JC, Denny R, Dorschel C, Gorenstein MV, Li GZ, Richardson K, Wall D, Geromanos SJ. Simultaneous qualitative and quantitative analysis of the *Escherichia coli* proteome: A sweet tale. *Mol Cell Proteomics*. 2006; 5(4):589–607. [PubMed: 16399765]
31. Li GZ, Vissers JP, Silva JC, Golick D, Gorenstein MV, Geromanos SJ. Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics*. 2009; 9(6):1696–719. [PubMed: 19294629]
32. Geromanos S, Vissers J, Silva J, Dorschel C, Li G, Gorenstein M, Bateman R, Langridge J. The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. *Proteomics*. 2009; 9(6):1683–95. [PubMed: 19294628]
33. Gillet LC, Navarro P, Tate S, Rost H, Selevsek N, Reiter L, Bonner R, Aebersold R. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: A new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics*. 2012; 11(6):O111 016717. [PubMed: 22261725]
34. Valentine S, Ewing M, Dilger J, Glover M, Geromanos S, Hughes C, Clemmer D. Using ion mobility data to improve peptide identification: intrinsic amino acid size parameters. *J Proteome Res*. 2011; 10(5):2318–29. [PubMed: 21417239]
35. Saka HA, Thompson JW, Chen YS, Kumar Y, Dubois LG, Moseley MA, Valdivia RH. Quantitative proteomics reveals metabolic and pathogenic properties of *Chlamydia trachomatis* developmental forms. *Mol Microbiol*. 2011; 82(5):1185–203. [PubMed: 22014092]
36. Stapels M, Piper C, Yang T, Li M, Stowell C, Xiong ZG, Saugstad J, Simon RP, Geromanos S, Langridge J, Lan JQ, Zhou A. Polycomb group proteins as epigenetic mediators of neuroprotection in ischemic tolerance. *Sci Signal*. 2010; 3(111):ra15. [PubMed: 20197544]
37. Gilar M, Olivova P, Daly AE, Gebler JC. Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. *J Sep Sci*. 2005; 28(14):1694–703. [PubMed: 16224963]
38. Gilar M, Olivova P, Daly AE, Gebler JC. Orthogonality of separation in two-dimensional liquid chromatography. *Anal Chem*. 2005; 77(19):6426–34. [PubMed: 16194109]
39. Dowell JA, Frost DC, Zhang J, Li L. Comparison of two-dimensional fractionation techniques for shotgun proteomics. *Anal Chem*. 2008; 80(17):6715–23. [PubMed: 18680313]
40. Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem*. 2003; 75(17):4646–58. [PubMed: 14632076]
41. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem*. 2002; 74(20):5383–92. [PubMed: 12403597]
42. Wall DP, Fraser HB, Hirsh AE. Detecting putative orthologs. *Bioinformatics*. 2003; 19(13):1710–1. [PubMed: 15593400]
43. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res*. 2003; 31(13):3497–500. [PubMed: 12824352]
44. Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci*. 1997; 13(5):555–6. [PubMed: 9367129]
45. Serres MH, Riley M. MultiFun, a multifunctional classification scheme for *Escherichia coli* K-12 gene products. *Microb Comp Genomics*. 2000; 5(4):205–22. [PubMed: 11471834]
46. Yoshiga T, Hernandez V, Fallon A, Law J. Mosquito transferrin, an acute-phase protein that is up-regulated upon infection. *Proc Natl Acad Sci U S A*. 1997; 94(23):12337–42. [PubMed: 9356450]
47. Guz N, Attardo G, Wu Y, SA. Molecular aspects of transferrin expression in the tsetse fly (*Glossina morsitans morsitans*). *J Insect Physiol*. 2007; 53(7):715–23. [PubMed: 17498733]
48. Ghazalpour A, Bennett B, Petyuk VA, Orozco L, Hagopian R, Mungrue IN, Farber CR, Sinsheimer J, Kang HM, Furlotte N, Park CC, Wen PZ, Brewer H, Weitz K, Camp DG 2nd, Pan C, Yordanova R, Neuhaus I, Tilford C, Siemers N, Gargalovic P, Eskin E, Kirchgessner T, Smith DJ, Smith RD, Lusk AJ. Comparative analysis of proteome and transcriptome variation in mouse. *Plos Genetics*. 2011; 7(6):e1001393. [PubMed: 21695224]

49. Schwanhaussner B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M. Global quantification of mammalian gene expression control. *Nature*. 2011; 473(7347):337–42. [PubMed: 21593866]
50. Maier T, Guell M, Serrano L. Correlation of mRNA and protein in complex biological samples. *FEBS Lett*. 2009; 583(24):3966–73. [PubMed: 19850042]
51. de Sousa Abreu R, Penalva LO, Marcotte EM, Vogel C. Global signatures of protein and mRNA expression levels. *Mol Biosyst*. 2009; 5(12):1512–26. [PubMed: 20023718]
52. Saito K. Regulation of sulfate transport and synthesis of sulfur-containing amino acids. *Curr Opin Plant Biol*. 2000; 3(3):188–95. [PubMed: 10837270]
53. Saito K. Sulfur assimilatory metabolism. The long and smelling road. *Plant Physiol*. 2004; 136(1):2443–50. [PubMed: 15375200]
54. Tsakraklides G, Martin M, Chalam R, Tarczynski MC, Schmidt A, Leustek T. Sulfate reduction is increased in transgenic *Arabidopsis thaliana* expressing 5'-adenylylsulfate reductase from *Pseudomonas aeruginosa*. *Plant J*. 2002; 32(6):879–89. [PubMed: 12492831]
55. Ishikawa H. Characterization of the protein species synthesized *in vivo* and *in vitro* by an aphid endosymbiont. *Insect Biochem*. 1984; 14:417–425.
56. Ishikawa H. Host-symbiont interactions in the protein synthesis in the pea aphid, *Acyrtosiphon pisum*. *Insect Biochem*. 1982; 12:613–622.
57. Aksoy S. Molecular analysis of the endosymbionts of tsetse flies: 16S rDNA locus and over-expression of a chaperonin. *Insect Mol Biol*. 1995; 4(1):23–9. [PubMed: 7538012]
58. Ahn TI, Lim ST, Leeu HK, Lee JE, WJK. A novel strong promoter of the groEx operon of symbiotic bacteria in *Amoeba proteus*. *Gene*. 1994; 148:43–49. [PubMed: 7926836]
59. Baumann P, Baumann L, Clark M. Levels of *Buchnera aphidicola* chaperonin GroEL during growth of the aphid *Schizaphys graminum*. *Curr Microbiol*. 1996; 32:279–285.
60. Charles H, Heddi A, Guillaud J, Nardon C, Nardon P. A molecular aspect of symbiotic interactions between the weevil *Sitophilus oryzae* and its endosymbiotic bacteria: Over-expression of a chaperonin. *Biochem Biophys Res Commun*. 1997; 239(3):769–74. [PubMed: 9367844]
61. Moran NA. Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proc Natl Acad Sci U S A*. 1996; 93(7):2873–8. [PubMed: 8610134]
62. Kuo CH, Moran NA, Ochman H. The consequences of genetic drift for bacterial genome complexity. *Genome Res*. 2009; 19(8):1450–4. [PubMed: 19502381]
63. Van Ham RC, Kamerbeek J, Palacios C, Rausell C, Abascal F, Bastolla U, Fernandez JM, Jimenez L, Postigo M, Silva FJ, Tamames J, Viguera E, Latorre A, Valencia A, Moran F, Moya A. Reductive genome evolution in *Buchnera aphidicola*. *Proc Natl Acad Sci U S A*. 2003; 100(2):581–586. [PubMed: 12522265]
64. Morin S, Ghanim M, Zeidan M, Czosnek H, Verbeek M, van den Heuvel JF. A GroEL homologue from endosymbiotic bacteria of the whitefly *Bemisia tabaci* is implicated in the circulative transmission of tomato yellow leaf curl virus. *Virology*. 1999; 256(1):75–84. [PubMed: 10087228]
65. Gottlieb Y, Zchori-Fein E, Mozes-Daube N, Kontsedalov S, Skaljac M, Brumin M, Sobol I, Czosnek H, Vavre F, Fleury F, Ghanim M. The transmission efficiency of tomato yellow leaf curl virus by the whitefly *Bemisia tabaci* is correlated with the presence of a specific symbiotic bacterium species. *J Virol*. 2010; 84(18):9310–7. [PubMed: 20631135]
66. Morioka M, Muraoka H, Yamamoto K, Ishikawa H. An endosymbiont chaperonin is a novel type of histidine protein kinase. *J Biochem*. 1994; 116(5):1075–81. [PubMed: 7896735]
67. Braig K, Otwinowski Z, Hegde R, Boisvert DC, Joachimiak A, Horwich AL, Sigler PB. The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. *Nature*. 1994; 371(6498):578–86. [PubMed: 7935790]
68. Nichol H, Law JH, Winzerling JJ. Iron metabolism in insects. *Annu Rev Entomol*. 2002; 47:535–59. [PubMed: 11729084]
69. Yoshiga T, Georgieva T, Dunkov BC, Harizanova N, Ralchev K, Law JH. Drosophila melanogaster transferrin. Cloning, deduced protein sequence, expression during the life cycle, gene localization and up-regulation on bacterial infection. *Eur J Biochem*. 1999; 260(2):414–20. [PubMed: 10095776]

70. Yun EY, Kang SW, Hwang JS, Goo TW, Kim SH, Jin BR, Kwon OY, Kim KY. Molecular cloning and characterization of a cDNA encoding a transferrin homolog from *Bombyx mori*. *Biol Chem*. 1999; 380(12):1455–9. [PubMed: 10661875]
71. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wortman JR, Yandell MD, Zhang Q, Chen LX, Brandon RC, Rogers YH, Blazej RG, Champe M, Pfeiffer BD, Wan KH, Doyle C, Baxter EG, Helt G, Nelson CR, Gabor GL, Abril JF, Agbayani A, An HJ, Andrews-Pfannkoch C, Baldwin D, Ballew RM, Basu A, Baxendale J, Bayraktaroglu L, Beasley EM, Beeson KY, Benos PV, Berman BP, Bhandari D, Bolshakov S, Borkova D, Botchan MR, Bouck J, Brokstein P, Brottier P, Burtis KC, Busam DA, Butler H, Cadieu E, Center A, Chandra I, Cherry JM, Cawley S, Dahlke C, Davenport LB, Davies P, de Pablos B, Delcher A, Deng Z, Mays AD, Dew I, Dietz SM, Dodson K, Doup LE, Downes M, Dugan-Rocha S, Dunkov BC, Dunn P, Durbin KJ, Evangelista CC, Ferraz C, Ferreira S, Fleischmann W, Fosler C, Gabrielian AE, Garg NS, Gelbart WM, Glasser K, Glodek A, Gong F, Gorrell JH, Gu Z, Guan P, Harris M, Harris NL, Harvey D, Heiman TJ, Hernandez JR, Houck J, Hostin D, Houston KA, Howland TJ, Wei MH, Ibegwam C, Jalali M, Kalush F, Karpen GH, Ke Z, Kennison JA, Ketchum KA, Kimmel BE, Kodira CD, Kraft C, Kravitz S, Kulp D, Lai Z, Lasko P, Lei Y, Levitsky AA, Li J, Li Z, Liang Y, Lin X, Liu X, Mattei B, McIntosh TC, McLeod MP, McPherson D, Merkulov G, Milshina NV, Mobarry C, Morris J, Moshrefi A, Mount SM, Moy M, Murphy B, Murphy L, Muzny DM, Nelson DL, Nelson DR, Nelson KA, Nixon K, Nusskern DR, Pacle JM, Palazzolo M, Pittman GS, Pan S, Pollard J, Puri V, Reese MG, Reinert K, Remington K, Saunders RD, Scheeler F, Shen H, Shue BC, Siden-Kiamos I, Simpson M, Skupski MP, Smith T, Spier E, Spradling AC, Stapleton M, Strong R, Sun E, Svirskas R, Tector C, Turner R, Venter E, Wang AH, Wang X, Wang ZY, Wassarman DA, Weinstock GM, Weissenbach J, Williams SM, Woodage T, Worley KC, Wu D, Yang S, Yao QA, Ye J, Yeh RF, Zaveri JS, Zhan M, Zhang G, Zhao Q, Zheng L, Zheng XH, Zhong FN, Zhong W, Zhou X, Zhu S, Zhu X, Smith HO, Gibbs RA, Myers EW, Rubin GM, Venter JC. The genome sequence of *Drosophila melanogaster*. *Science*. 2000; 287(5461):2185–95. [PubMed: 10731132]
72. Login FH, Balmand S, Vallier A, Vincent-Monegat C, Vigneron A, Weiss-Gayet M, Rochat D, Heddi A. Antimicrobial peptides keep insect endosymbionts under control. *Science*. 2011; 334(6054):362–5. [PubMed: 22021855]
73. Liu H, Sadygov RG, Yates JR. 3rd, A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem*. 2004; 76(14):4193–201. [PubMed: 15253663]
74. Collier TS, Randall SM, Sarkar P, Rao BM, Dean RA, Muddiman DC. Comparison of stable-isotope labeling with amino acids in cell culture and spectral counting for relative quantification of protein expression. *Rapid Commun Mass Spectrom*. 2011; 25(17):2524–32. [PubMed: 21818813]
75. Shigenobu S, Wilson AC. Genomic revelations of a mutualism: the pea aphid and its obligate bacterial symbiont. *Cellular and Molecular Life Sciences*. 2011; 68(8):1297–309. [PubMed: 21390549]

Figure 1. Proteins and peptides identified by DDA, MS^E and HDMS^E analyses of *Blochmannia chromaiodes* samples

The DDA, MS^E, and HDMS^E data from three *B. chromaiodes* samples were imported and aggregated in Scaffold v3.1 for curation. Peptide identifications were accepted at greater than 50% probability. Protein identifications were accepted at greater than 90% probability and contained at least two identified peptides. Peptide FDR was 1.2% for both ant host and bacterial peptides. A, Proteins identified. B, Peptides identified.

Figure 2. Dynamic range of the *Blochmannia chromaiodes* proteome spans three orders of magnitude, and many abundant proteins also showed high transcript levels in *B. floridanus*. The frequency distribution of quantities for 309 *B. chromaiodes* proteins is presented on a log scale. This dataset includes 45 (of the 53) genes previously identified as highly expressed in the *B. floridanus* transcriptome¹⁷ (shaded in dark grey). While notable exceptions exist, those 45 proteins are relatively abundant in the *B. chromaiodes* proteome. The median protein quantity of those 45 *B. chromaiodes* proteins (107 ± 4.1 fmol/ μ g) is significantly higher than the median value (34.3 ± 5.3 fmol/ μ g) of the remaining 264 proteins ($p < 0.000001$, Wilcoxon test). The most abundant *B. chromaiodes* proteins are labeled.

Figure 3. Quantitative results indicate high reproducibility among three biological replicates A, Percentage of *Blochmannia* and *Camponotus* protein within three biological samples, showing high enrichment for *Blochmannia* proteins and reproducible sample preparation. All proteins quantified as the unit of fmol were added together according to their ant or bacterial origin, and percent contribution to the total protein was calculated. Bars indicate the mean of three biological samples (mean ± SEM). Across the three samples, *Blochmannia* proteins comprised $76.6\% \pm 0.9\%$ (1.2% RSD) of the total (ant and *Blochmannia*) protein, while the host protein constituted $23.4\% \pm 0.9\%$. B, Frequency and cumulative %RSD for the 309 quantified *Blochmannia* proteins. X-axis shows the number of proteins within 5% RSD intervals. Y-axes indicate (left) the number of proteins within each 5% RSD bin, and (right) the cumulative distribution, or percent of proteins that had a given %RSD or lower. For example, approximately 80% of the proteins quantified had less than 30% RSD across three biological samples.

Figure 4. Comparison of genome, proteome content, and protein abundances according to functional categories

Broad functional groups were assigned according to a modification of the MultiFun schema developed for *E. coli*⁴⁵ (see text). *Buchnera* protein abundances are based on NSAF results in the Bu-2 sample of Poliakov et al.²⁴ Orthologs identified and functional assignments are listed in Supplemental Table 4. Values <1% are not labeled on the pie chart. A, Percent contribution of each functional category to the 609 annotated proteins in the *B. chromaiodes* genome. B, Percent contribution of each functional category to the 330 proteins detected in the *B. chromaiodes* proteome, with detected proteins counting equally regardless of abundance. C, Percent contribution of each functional category to the total protein abundance (quantified as fmol/μg) in the *B. chromaiodes* proteome. D, Percent contribution of each functional category to the total protein abundance (quantified as NSAF) in the *Buchnera*-APS proteome.²⁴

Figure 5. Quantitative analysis of sulfate assimilation enzymes in the *Blochmannia chromaiodes* proteome

Upper panel: Sulfate reduction I (assimilatory) pathway. In this pathway, reduction of sulfate to sulfite is catalyzed by ATP sulfurylase (CysD and CysN), adenosine-5'-phosphosulfate (APS) kinase (CysC), and 3'-phosphoadenosine-5'-phosphosulfate (PAPS) reductase (CysH). Sulfite reductase (CysJ, CysI) generates the final product of hydrogen sulfide, which is used to biosynthesize the sulfur-containing amino acid cysteine. Lower panel: Quantification of *B. chromaiodes* enzymes in the sulfate assimilation pathway.

Abundance of *Blochmannia chromaiodes* proteins within select functional categories

Table 1

For comparison, values are listed based on the representation of protein-coding sequences (CDS) in the *Blochmannia* genome, in the pool of proteins detected in the *B. chromaiodes* proteome, and in the total protein abundance based on quantitative data. Values for the Chaperonins/Adaptation to Stress category are included in Figure 4. Other groups listed are those sub-categories of Metabolism that are particularly relevant for symbiont nutritional functions. Proteins within select categories are listed in Supplemental Tables 5–7.

Function	# CDS in genome (609 total)	% of CDS in genome	# proteins detected (330 total)	% of all proteins detected	protein abundance (fmol/μg)	% total protein abundance
Chaperonins/Adaptation to Stress	29	4.8%	20	6.1%	3,732	15.0%
Essential amino acid biosynthesis	50	8.2%	41	12.4%	2,320	9.3%
Nonessential amino acid biosynthesis	18	3.0%	14	4.2%	797	3.2%
Sulfate assimilation	6	1.0%	5	1.5%	292	1.2%
Fatty acid biosynthesis	13	2.1%	8	2.4%	336	1.4%
Nucleotide biosynthesis	14	2.3%	10	3.0%	441	1.8%
Nitrogen recycling	10	1.6%	4	1.2%	270	1.1%

Abundance of *Blochmannia chromaiodes* proteins contributing to amino acid biosynthesis in *B. chromaiodes*

Functional roles of proteins were assigned according to the MultiFun schema.⁴⁵ All annotated proteins contributing to the biosynthesis of a given amino acid are listed, even if the protein was not detected in the proteome. Proteins are listed more than once, if they contribute to the synthesis of more than one amino acid. Columns list (L to R): protein name in *B. chromaiodes*; numerical *B. chromaiodes* gene ID in annotated genome; status in this proteomic study as Quantified (Q, white background), Not Detected (ND, dark grey shading), or Detected but Not Quantified (DNQ, light grey shading); mean quantity (in fmol/ μ g) for proteins that were quantified; standard deviation of this mean; specific amino acid synthesized; status of amino acid as essential or nonessential for animals.

protein name	Bchrom ID	Q, ND, DNQ	Average (fmol/ μ g)	Stdev (fmol/ μ g)	AA	AA type
AroF	Bchrom_184	Q	34.4	16.8	Chorismate	essen. & nonessen [*]
AroE	Bchrom_230	Q	45.2	20.4	Chorismate	essen. & nonessen
AroA	Bchrom_404	Q	26.4	11.7	Chorismate	essen. & nonessen
AroC	Bchrom_531	Q	74.7	23.4	Chorismate	essen. & nonessen
AroB	Bchrom_607	Q	14.7	1.3	Chorismate	essen. & nonessen
AroK	Bchrom_608	ND			Chorismate	essen. & nonessen
CarA	Bchrom_127	Q	40.5	7.7	Arginine	essential
CarB	Bchrom_128	Q	46.3	4.2	Arginine	essential
ArgI	Bchrom_32	Q	52.9	2.6	Arginine	essential
CysK	Bchrom_540	Q	144.8	42.3	Cysteine	essential
CysE	Bchrom_647	ND			Cysteine	essential
HisG	Bchrom_491	Q	12.4	0.7	Histidine	essential
HisD	Bchrom_492	Q	15.1	2.5	Histidine	essential
HisC	Bchrom_493	Q	67.4	2	Histidine	essential
HisB	Bchrom_494	Q	42	9.8	Histidine	essential
HisA	Bchrom_497	Q	32.3	3.7	Histidine	essential
HisI	Bchrom_499	Q	17.1	6.3	Histidine	essential
HisH	Bchrom_496	DNQ			Histidine	essential
HisF	Bchrom_498	ND			Histidine	essential
LeuD	Bchrom_135	Q	13.4	3.4	Isoleucine	essential
LeuC	Bchrom_136	Q	31.3	14.4	Isoleucine	essential
LeuB	Bchrom_137	Q	62.7	14.2	Isoleucine	essential

protein name	Behrom ID	Q ₁ , ND, DNQ	Average (fmol/μg)	Stdev (fmol/μg)	AA	AA type
LeuA	Behrom_138	Q	117.2	5.3	Isoleucine	essential
IlvE	Behrom_634	ND			Isoleucine	essential
IlvC	Behrom_631	Q	80.2	11.6	Leucine/valine	essential
IlvA	Behrom_632	Q	50.1	2.9	Leucine/valine	essential
IlvD	Behrom_633	Q	49.7	5.8	Leucine/valine	essential
IlvG	Behrom_636	Q	55.4	6.2	Leucine/valine	essential
IlvE	Behrom_634	ND			Leucine/valine	essential
IlvM	Behrom_635	ND			Leucine/valine	essential
DapB	Behrom_126	Q	10.8	1.8	Lysine	essential
LysA	Behrom_278	Q	10.6	2.2	Lysine	essential
DapD	Behrom_284	Q	68.3	27.3	Lysine	essential
DapE	Behrom_549	Q	14.4	1.8	Lysine	essential
Asd	Behrom_611	Q	22.9	10.7	Lysine	essential
DapF	Behrom_620	DNQ			Lysine	essential
DapA	Behrom_551	ND			Lysine	essential
Asd	Behrom_611	Q	22.9	10.7	Methionine	essential
MetB	Behrom_642	Q	23.7	7.7	Methionine	essential
MetE	Behrom_675	Q	295.8	20.6	Methionine	essential
MetC	Behrom_68	Q	30.5	3	Methionine	essential
MetA	Behrom_681	ND			Methionine	essential
TyrA	Behrom_185	Q	139.4	10	Phenylalanine	essential
PheA	Behrom_186	ND			Phenylalanine	essential
ThrA	Behrom_116	Q	252.9	52.6	Threonine	essential
ThrB	Behrom_117	Q	30.6	4.3	Threonine	essential
ThrC	Behrom_118	Q	90.5	11	Threonine	essential
Asd	Behrom_611	Q	22.9	10.7	Threonine	essential
TrpE	Behrom_450	Q	55.4	1	Tryptophan	essential
TrpC	Behrom_453	Q	17.1	3.5	Tryptophan	essential
TrpB	Behrom_454	Q	99.6	15.2	Tryptophan	essential

protein name	Behrom ID	Q ₁ , ND, DNQ	Average (fmol/μg)	Stdev (fmol/μg)	AA	AA type
TrpA	Behrom_455	Q	31	6.9	Tryptophan	essential
TrpD	Behrom_452	ND			Tryptophan	essential
IleE	Behrom_634	ND			Alanine	nonessential
IscS	Behrom_568	Q	18.9	7.2	Alanine	nonessential
AspC	Behrom_446	Q	18.2	5	Aspartate	nonessential
GltP	Behrom_30	ND			Aspartate	nonessential
GlnA	Behrom_666	Q	65.8	11.9	Glutamine	nonessential
GlyA	Behrom_571	Q	24.8	4	Glycine	nonessential
ThrA	Behrom_116	Q	252.9	52.6	Homoserine	nonessential
Asd	Behrom_611	Q	22.9	10.7	Homoserine	nonessential
SerS	Behrom_406	Q	58.9	1.1	Selenocysteine	nonessential
SerC	Behrom_405	DNQ			Serine	nonessential
AroF	Behrom_184	Q	34.4	16.8	Tyrosine	nonessential
TyrA	Behrom_185	Q	139.4	10	Tyrosine	nonessential
PheA	Behrom_186	ND			Tyrosine	nonessential

* Chorismate is considered both essential and nonessential because it is precursor to tryptophan, tyrosine and phenylalanine.