Reductive Evolution of Bacterial Genome in Insect Gut Environment

Naruo Nikoh[1], Takahiro Hosokawa[2], Kenshiro Oshima[3], Masahira Hattori[3], and Takema Fukatsu*[,2]

[1]Department of Liberal Arts, The Open University of Japan, Chiba, Japan

[2]Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan

[3]Center for Omics and Bioinformatics, Graduate School of Frontier Sciences, University of Tokyo, Kashiwa, Japan

*Corresponding author: E-mail: t-fukatsu@aist.go.jp.

Abstract

Obligate endocellular symbiotic bacteria of insects and other organisms generally exhibit drastic genome reduction. Recently, it was shown that symbiotic gut bacteria of some stinkbugs also have remarkably reduced genomes. Here, we report the complete genome sequence of such a gut bacterium *Ishikawaella capsulata* of the plataspid stinkbug *Megacopta punctatissima*. Gene repertoire and evolutionary patterns, including AT richness and elevated evolutionary rate, of the 745,590 bp genome were strikingly similar to those of obligate γ-proteobacterial endocellular insect symbionts like *Buchnera* in aphids and *Wigglesworthia* in tsetse flies. *Ishikawaella* was suggested to supply essential amino acids for the plant-sucking stinkbug as *Buchnera* does for the host aphid. Although *Buchnera* is phylogenetically closer to *Wigglesworthia* than to *Ishikawaella*, in terms of gene repertoire *Buchnera* was similar to *Ishikawaella* rather than to *Wigglesworthia*, providing a possible case of genome-level convergence of gene content. Meanwhile, several notable differences were identified between the genomes of *Ishikawaella* and *Buchnera*, including retention of TCA cycle genes and lack of flagellum-related genes in *Ishikawaella*, which may reflect their adaptation to distinct symbiotic habitats. Unexpectedly, *Ishikawaella* retained fewer genes related to cell wall synthesis and lipid metabolism than many endocellular insect symbionts. The plasmid of *Ishikawaella* encoded genes for arginine metabolism and oxalate detoxification, suggesting the possibility of additional *Ishikawaella* roles similar to those of human gut bacteria. Our data highlight strikingly similar evolutionary patterns that are shared between the extracellular and endocellular insect symbiont genomes.

Key words: *Ishikawaella capsulata*, *Megacopta punctatissima*, extracellular gut symbiosis, genome reduction, genome sequencing.

Introduction

Insects represent the majority of eukaryotic biodiversity in the terrestrial ecosystem (Grimardi and Engel 2005), and many insects harbor symbiotic bacteria in their gut, body cavity, or cells (Buchner 1965; Bourtzis and Miller 2003). Many bacteriocyte-associated endocellular symbiotic bacteria like *Buchnera* of aphids and *Wigglesworthia* of tsetse flies are essential for survival and reproduction of their host insects, showing stable maternal inheritance and host–symbiont cospeciation. These obligate endocellular bacteria generally exhibit peculiar genetic traits, such as AT-biased nucleotide composition, accelerated molecular evolution, and drastically reduced genome size less than 1 Mb (Shigenobu et al. 2000;

Akman et al. 2002; Gil et al. 2003; van Ham et al. 2003; Degnan et al. 2005; Nakabachi et al. 2006; Perez-Brocal et al. 2006; Wu et al. 2006; McCutcheon and Moran 2007, 2010; López-Sánchez et al. 2009; McCutcheon et al. 2009; Sabree et al. 2009; Kirkness et al. 2010). These genetic traits are hypothesized to be the result of stable and nutrition-rich endocellular environment and also the consequence of attenuated purifying selection due to small population size and strong bottleneck, which are associated with the endosymbiotic lifestyle of the vertically transmitted symbionts (Wernegreen 2002; Moran et al. 2008; Moya et al. 2008). Meanwhile, facultative insect endosymbionts like *Wolbachia*, *Sodalis*, *Hamiltonella*, *Regiella*, *Serratia*, and

Arsenophonus have larger genomes of 1–4 Mb and exhibit these genetic traits to much lesser extents (Wu et al. 2004; Toh et al. 2006; Klasson et al. 2008, 2009; Degnan et al. 2009, 2010; Wilkes et al. 2010; Burke and Moran 2011), which may be relevant to less specialization and occasional horizontal transmission of the facultative microbial associates (Wernegreen 2002; Moran et al. 2008; Moya et al. 2008).

In this context, extracellular symbiotic microbes in the alimentary tract of host insects may seem unlikely to exhibit such reductive genome evolution, considering that such associations are unstable outside the host body cavity, their lifestyle seems closer to that of free-living microbes, and thus, more genes are to be needed to cope with environmental fluctuations. However, recent studies have revealed that some stinkbugs harbor a specific γ-proteobacterial symbiont in the midgut cavity, which is essential for host growth and reproduction, is vertically transmitted through host generations, exhibits host–symbiont cospeciation, and shows reduced genome sizes in the range of 0.7–0.9 Mb (Hosokawa et al. 2006, 2010; Kikuchi et al. 2009). In particular, stinkbugs of the family Plataspidae are known for their unique mechanism for vertical transmission called "symbiont capsule." When female insects lay eggs on their host plant, small brownish particles are deposited together, wherein the symbiotic bacteria are encased. The bacteria in the capsules are ingested by newborn nymphs and colonize the midgut. In the developmental course, oddly, the nymphal midgut is constricted and separated into anterior and posterior parts. In adult insects, the anterior midgut is free of the symbiont, and the posterior midgut is transformed into a voluminous organ for harboring a huge amount of symbiont cells in the cavity. Judging from the peculiar anatomy, the plant sap ingested by the insect is completely absorbed in the anterior midgut, the waste is excreted through the Malpighian tubules into the hindgut, and there is no food flow through the posterior midgut (Hosokawa et al. 2005, 2006). Structural, functional, and evolutionary details of the symbiont genome are of interest but totally unknown.

In this study, we determined the complete genome sequence of such a gut bacterium "*Candidatus* Ishikawaella capsulata" (hereafter referred to *Ishikawaella* for simplicity) associated with the plataspid stinkbug *Megacopta punctatissima*, which unveiled strikingly similar evolutionary patterns shared between the extracellular and endocellular insect symbiont genomes.

Materials and Methods

Symbiont Genomic DNA Preparation

We used an inbred strain of *M. punctatissima*, which was initially collected at Kobe, Japan, in 1999 from the kudzu vine (*Pueraria lobata*) and maintained in the laboratory on soybean plants (*Glycine max*) and pea pods (*Pisum sativum*).

An adult female was dissected in a phosphate-buffered saline, and a symbiotic section of posterior midgut was isolated and subjected to DNA extraction. Aliquots of the DNA sample were subjected to quantitative polymerase chain reaction (PCR) of the symbiont *groEL* gene and the host elongation factor 1α gene as described (Hosokawa, Kikuchi, and Fukatsu 2007).

Genome Sequencing, Gene Prediction, and Annotation

The DNA sample was subjected to whole-genome shotgun sequencing as described (Akman et al. 2002; Toh et al. 2006). We constructed small-insert (2 kb) genomic libraries and generated 12,247 sequences, giving 15-fold coverage from both ends of the genomic clones. Sequence assembly was carried out using the PHRED-PHRAP-CONSED package (Gordon et al. 2001). Remaining gaps were closed by sequencing of clones that spanned the gaps or by direct sequencing. To exclude the possibility of sequence error, we assessed and confirmed the quality of the assembled sequence. The correct assembly was confirmed by genomic restriction fragment patterns of pulsed field gel electrophoresis. Putative protein-coding sequences (CDSs) were predicted using Glimmer2.0 (Delcher et al. 1999). The annotation of CDSs was based on results of BlastP searches against *Escherichia coli* genome and the NCBI nonredundant protein database. The CDSs exhibiting database matches to a functional gene of other bacteria, but interrupted by frameshifts and/or stop codons, were annotated as pseudogenes. To find pseudogenes in spacer regions between CDSs, BlastX searches were also conducted against the database using the spacer as the query sequence. Transfer RNA (tRNA) genes were predicted by tRNAscan-SE (Lowe and Eddy 1997). Other noncoding RNAs were identified by similarity to *E. coli* homologs. Repetitive sequence regions were identified using Tandem Repeats Finder 4.0 (Benson 1999). The theoretical isoelectric points of proteins were calculated using Compute pI/Mw tool (Bjellqvist et al. 1993). The bacterial insertion sequences were predicted using ISfinder (Siguier et al. 2006).

Molecular Phylogenetic and Evolutionary Analyses

A set of 50 ribosomal protein genes, for which orthologs were commonly identified in *Ishikawaella* and other γ-proteobacterial representatives, was selected for phylogenetic analyses (supplementary table S1, Supplementary Material online). Each of the ortholog sets was aligned using MAFFT 5.6 (Katoh et al. 2005), and all the alignments were concatenated. The Whelan and Goldman (WAG)+Γ+Inv substitution model for the amino acid sequences was selected under the Akaike criterion using ProtTest v1.4 (Abascal et al. 2005). Molecular phylogenetic analyses were conducted by three methods, maximum likelihood, Bayesian, and neighbor-joining, using RAxML Version 7.0.0 (Stamatakis 2006), MrBayes 3.1.2 (Ronquist and Huelsenbeck 2003), and PHYLIP 3.6 (Felsenstein 2005). Bootstrap values

for maximum likelihood and neighbor-joining phylogenies were obtained by 1,000 resamplings. Posterior probabilities were estimated for Bayesian phylogeny. A relative rate test on the basis of amino acid distances calculated from the concatenated alignment of the 50 ribosomal protein sequences was performed using RRTree (Robinson-Rechavi and Huchon 2000). Gene content clustering was conducted as described (Wolf et al. 2002). A set of clusters of orthologous groups of proteins (COGs) (Tatusov et al. 2003) represented in each genome was determined. The gene content similarity index S was defined as follows: given that the set of COGs contained in the bacterial genomes A and B are shown as a and b, respectively, S between the genomes A and B was calculated as $|a \cap b|/|a \cup b|$. The gene content distance D was obtained as $1 - S$, and a pairwise distance matrix for the analyzed bacterial genomes was subjected to neighbor-joining clustering. Bootstrap values were obtained by generating 1,000 replicates of the entire set of 4,873 COGs.

Functional Analysis of Predicted Coding Genes

The predicted protein-coding genes were classified on the basis of COG classification with some modification. Metabolic pathways were examined and verified according to the pathway descriptions in the EcoCyc and KEGG databases (Kanehisa et al. 2008; Keseler et al. 2009).

Reconstruction of Ancestral Genomes

To determine orthologous coding genes, we performed reciprocal BlastP searches between genes encoded in the *E. coli* K12 genome and genes encoded in each of nine insect symbiont genomes (1 *Ishikawaella*, 1 *Baumannia*, 4 *Buchnera*, 2 *Blochmannia*, and 1 *Wigglesworthia*) with a cutoff E value at 10^{-5} and retained genes that exhibited significant hits in both comparisons. For duplicated genes, their orthologous relationships were also verified by molecular phylogenetic analyses. To search pseudogenes, TBlastN searches were conducted against the *E. coli* protein data set using spacer sequence between estimated open reading frames (ORFs) in the symbiont genomes as query. ORFs shorter than 40 amino acids with no significant homology to any bacterial genome sequences were discarded. Gene content was determined as distinct orthologous gene groups (OGGs) for each of the genomes of *E. coli* and the insect symbionts. By compiling all the OGGs from nine insect symbiont genomes without duplication, we obtained 1,021 distinct OGGs in total. Then, we inferred the status of the OGGs on each of the ancestral nodes in the symbiont phylogeny by applying a parsimony principle to minimize total number of gene loss events, where we ignored the possibility of acquisition of new genes via lateral gene transfer in the streamlined insect symbiont genomes. Genome size of an ancestral node was estimated by adding the genome size of its descendant to the sum of the OGG lengths lost between the ancestral node and the descendant. The length of the *E. coli* homolog of the OGG was assumed to be
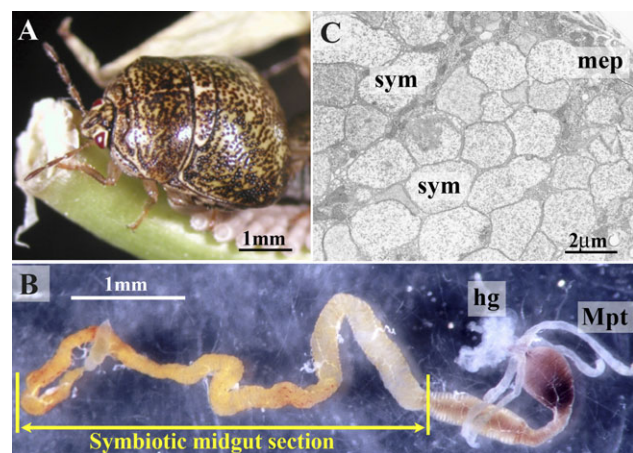

Fig. 1.—(*A*) An adult female of *Megacopta punctatissima*. (*B*) A posterior midgut dissected from an adult female. The symbiotic midgut section used for shotgun library construction is shown. (*C*) A transmission electron micrograph of the symbiotic midgut section, wherein symbiont cells are densely packed. Abbreviations: hg, hindgut; mep, midgut epithelium; Mpt, Malpighian tubule; and sym, symbiont cell.

the length of the lost gene. Because an ancestral node has two descendants, the estimated genome size of the ancestral node was obtained by averaging the values for two descendants.

Results and Discussion

Preparation and Sequencing of the *Ishikawaella* Genome

An adult female of *M. punctatissima* (fig. 1*A*) was dissected, and a symbiotic section of posterior midgut was carefully isolated (fig. 1*B*). The large midgut section contained a huge amount of *Ishikawaella* cells (fig. 1*C*) from which 0.82 μg of total DNA was prepared. Quantitative PCR assays evaluated the relative abundance of 9,100 symbiont *groEL* gene copies per host elongation factor 1α gene copy in the DNA sample. Given the symbiont genome size as 0.8 Mb (Hosokawa et al. 2006) and the host genome size as presumably 500 Mb or so, purity of the *Ishikawaella* genome in the DNA sample was estimated to be $[9{,}100 \times 0.8]/[9{,}100 \times 0.8 + 1 \times 500] \times 100 = 87\%$. The DNA sample was subjected to shotgun library construction and Sanger sequencing. Of 12,247 sequence reads determined, 10,665 and 361 were assembled into a circular bacterial chromosome and a circular plasmid, respectively. Hence, 90% (11,026/12,247) of the reads represented the *Ishikawaella* genome, which agreed with the quantitative PCR estimate.

General Features of the *Ishikawaella* Genome

The main genome of *Ishikawaella* consisted of a circular 745,590 bp chromosome encoding 611 putative protein-coding ORFs with an average size of 987 bp, which covers 81% of the whole genome. Of these, 568 were assigned to
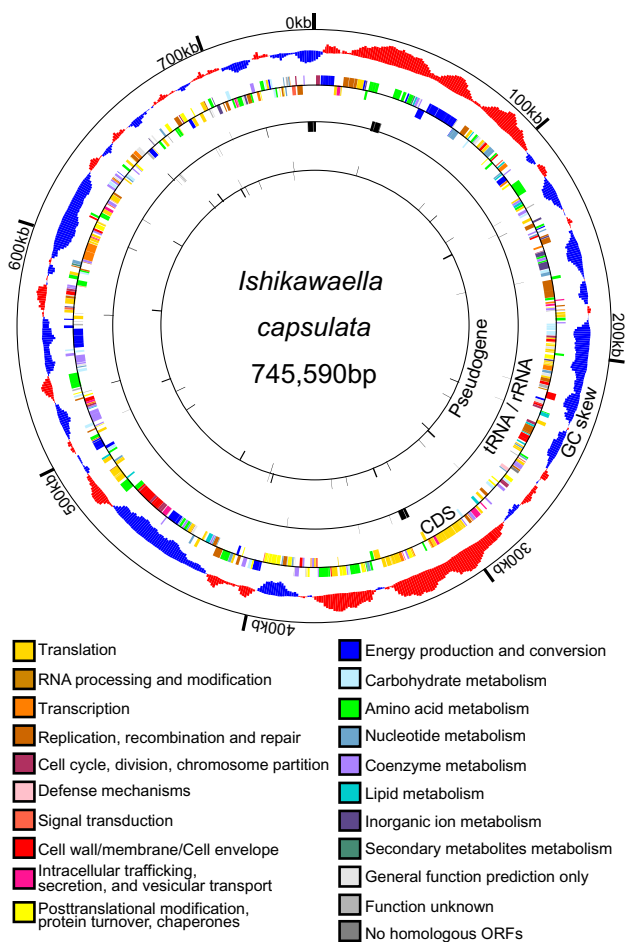
Translation
RNA processing and modification
Transcription
Replication, recombination and repair
Cell cycle, division, chromosome partition
Defense mechanisms
Signal transduction
Cell wall/membrane/Cell envelope
Intracellular trafficking, secretion, and vesicular transport
Posttranslational modification, protein turnover, chaperones
Energy production and conversion
Carbohydrate metabolism
Amino acid metabolism
Nucleotide metabolism
Coenzyme metabolism
Lipid metabolism
Inorganic ion metabolism
Secondary metabolites metabolism
General function prediction only
Function unknown
No homologous ORFs

Fig. 2.—A circular view of the *Ishikawaella* genome. On the GC skew circle, red and blue indicate GC-rich and poor, respectively. On the CDS circle, colors indicate functional categories as shown at the bottom.

Ishikawaella shares a common ancestry with the obligate endocellular symbionts and that their common ancestor might have already experienced reductive genome evolution to some extent. It should be noted that comparison of these symbiont genomes provides an ideal opportunity to investigate how the extracellular condition in the gut cavity and the endocellular conditions in the bacteriocytes have affected their genome evolution.

Similarity of the *Ishikawaella* Genome to Endocellular Insect Symbiont Genomes

General genomic features of *Ishikawaella* were strikingly similar to those of endocellular symbiotic bacteria, commonly exhibiting small genome sizes, high AT contents, high inferred pI values for encoded proteins, and few mobile genetic elements (table 1). The evolutionary rate of *Ishikawaella* was significantly higher than those of free-living bacteria and was equivalent to those of the endocellular symbiotic bacteria: nearly equal to those of *Buchnera* and *Baumannia* and slightly lower than those of *Blochmannia* and *Wigglesworthia* (supplementary table S3, Supplementary Material online). The genome of *Ishikawaella* was about six times smaller than the genome of *E. coli*, and almost all the *Ishikawaella* ORFs (97.0%; 589/607) had their orthologs in the genome of *E. coli*, indicating that the *Ishikawaella* genome is a small subset of the genome of free-living γ-proteobacteria, like the genomes of *Buchnera* and obligate endocellular symbionts of other insects (supplementary fig. S2A–D, Supplementary Material online).

Metabolic Capacity and Putative Biological Role of *Ishikawaella*

Despite the drastic genome reduction, the *Ishikawaella* genome retained many genes responsible for basic cellular processes such as translation, replication, energy production, etc., as many endocellular symbiont genomes (supplementary table S4, Supplementary Material online). Many genes involved in metabolism of amino acids were conserved in the *Ishikawaella* genome (figs. 4 and 5; supplementary table S4, Supplementary Material online). Gene content analysis inferred that *Ishikawaella* is capable of synthesizing almost all essential amino acids and some nonessential amino acids and some vitamins and cofactors (figs. 4 and 6). Only the final step enzyme, *ilvE*, involved in the synthesis of branched essential amino acids, namely isoleucine, leucine, and valine, was missing in the *Ishikawaella* genome (fig. 6A). It should be noted that the gene is also lacking in the *Buchnera* genome (Shigenobu et al. 2000). Presumably, absence of the gene is complemented by corresponding enzymes either from *Ishikawaella* itself or from the host insect (supplementary table S5, Supplementary Material online). Given that plataspid stinkbugs feed exclusively on plant sap devoid of essential amino acids and some vitamins, *Ishikawaella* probably compensates for the nutritional deficiency of

putative biological functions, 34 matched hypothetical proteins of unknown function, and 9 were unique to *Ishikawaella*. Some of the putative *Ishikawaella*-specific ORF might not be true genes, considering that eight of nine were less than 190 bp in size. Three ribosomal operons, three structural RNA genes, and 37 tRNA genes assigning all 20 amino acids were identified (supplementary table S1, Supplementary Material online; fig. 2). Besides these genes, 35 ORFs and one 16S rRNA gene were truncated and/or interrupted by stop codons, which are probably pseudogenes (supplementary table S2, Supplementary Material online). In addition, *Ishikawaella* has a 9,139 bp circular plasmid, named pAst, containing eight ORFs (supplementary fig. S1, Supplementary Material online). Phylogenetic analyses based on 50 ribosomal proteins revealed that *Ishikawaella* is placed in a well-supported clade in the γ-*Proteobacteria* together with obligate endocellular symbiotic bacteria of other insects, such as *Buchnera* of aphids, *Baumannia* of sharpshooters, *Blochmannia* of ants, and *Wigglesworthia* of tsetse flies (fig. 3). The phylogenetic relationship suggested that the extracellular symbiont
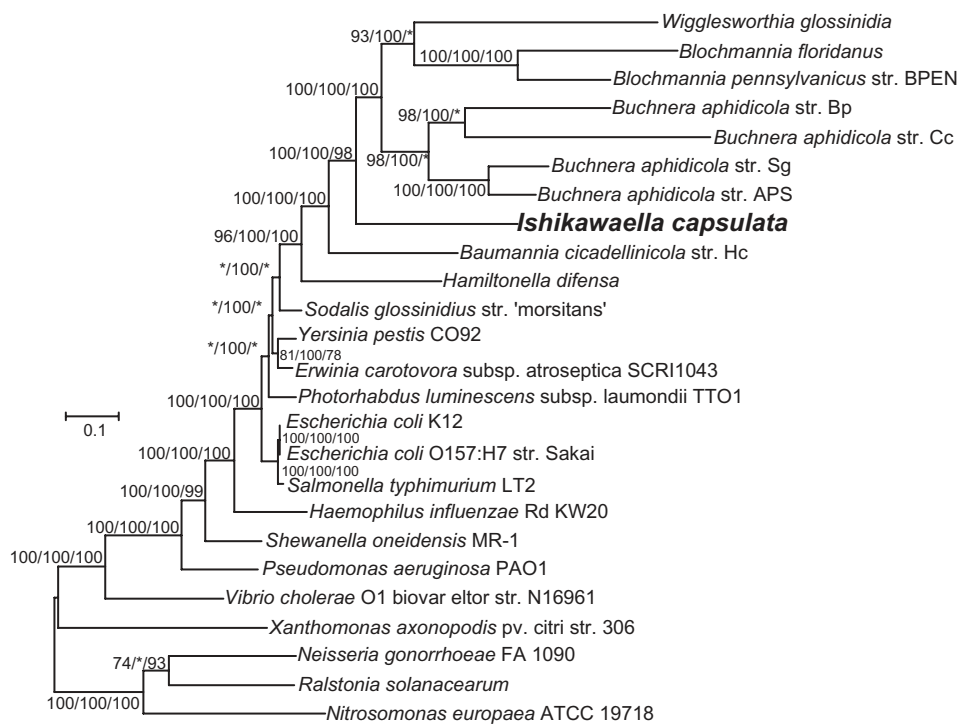
Fig. 3.—Phylogenetic placement of *Ishikawaella* in the γ-Proteobacteria. A maximum likelihood phylogeny inferred from concatenated 50 ribosomal protein sequences (6,435 aligned amino acid sites) is shown. Statistical supports (>70%) for each clade are shown at each node in the order of maximum likelihood, Bayesian, and neighbor-joining analyses. Asterisks denote statistical support values lower than 70%.

the diet, as *Buchnera* does for the host aphid (Douglas 1998; Baumann 2005).

Similarity in Gene Repertoire between *Ishikawaella* and Endocellular Insect Symbionts

The extracellular symbiotic life in the gut cavity seems to entail molecular, cellular, and physiological requirements different from the endocellular life in the bacteriocyte. For example, the environment in the gut cavity may be less homeostatic than that in the cytoplasm, which would lead to retention of genes needed for free-living life, such as transcription factors for metabolic regulation. The extracellular symbiont may be more frequently exposed to mechanical and osmotic stresses, which would result in conservation of genes related to cell wall synthesis and formation. The extracellular symbiont in the gut cavity is less likely to be affected by host innate immunity, which might affect its cell surface molecules. Hence, we initially expected that the gene repertoire of *Ishikawaella* might be divergent from those of obligate endocellular symbionts of other insects. However, simple comparison of their gene contents indicated no such explicit tendencies (supplementary fig. S2E–J, Supplementary Material online). Cluster analysis of their gene contents revealed that 1) the gene repertoire of *Ishikawaella* was the most similar to those of obligate endocellular insect symbionts *Buchnera*, *Baumannia*, *Blochmannia*, and *Wigglesworthia*, 2) among

them, the *Ishikawaella* genome was the closest to the *Buchnera* genome and the most distant from the *Wigglesworthia* genome, 3) obligate endocellular symbiotic bacteria of deep-sea mollusks and termite-associated protozoans were placed just outside the cluster, and thus 4) on account of the gene repertoire, *Ishikawaella* was, despite its extracellular niche, nested within microbial clusters exclusively consisting of endocellular symbiotic bacteria (fig. 7). These patterns may provide some insights into evolutionary aspects of the reduced symbiont genomes. Firstly, regardless of their endocellular or extracellular habitats, the reductive genome evolution was commonly observed. Secondly, regardless of their phylogenetic affinities, diverse symbiont lineages have experienced such reductive genome evolution. Hence, it appears likely that the reductive genome evolution is attributable neither to their cellular habitats nor to their phylogenetic placements but rather to some general traits commonly associated with diverse obligate bacterial symbionts, namely the small population size, the lack of recombination, and the deletional bias inherent in all bacterial genomes (Wernegreen 2002; Moran et al. 2008; Moya et al. 2008). We also suggest that the host insect ecology and the physiological involvement of the symbiont might have also affected the evolutionary patterns. Plataspid stinkbugs and aphids are phloem sap feeders, wherein the symbionts supply essential amino acids for the hosts (Douglas 1998; Baumann 2005). Tsetse

Table 1

General Features of the Genomes of *Ishikawaella*, Endocellular Insect Symbionts, and Free-Living γ-proteobacteria

Bacterium	*Ishikawaella capsulata*	*Buchnera aphidicola* APS	*Baumannia cicadellinicola*	*Blochmannia floridanus*	*Wigglesworthia glossinidia*	*Escherichia coli* K12	*Vibrio cholerae* O1
Host insect	Stinkbug (Plataspidae)	Aphid (Aphididae)	Sharpshooter (Cicadellidae)	Carpenter ant (Formicidae)	Tsetse flies (Glossinidae)	Not applicable	Not applicable
Symbiotic niche	Extracellular (midgut cavity)	Endocellular (bacteriocyte)	Endocellular (bacteriocyte)	Endocellular (bacteriocyte)	Endocellular (bacteriocyte)	Free-living (vertebrate gut)	Free-living (vertebrate gut)
Biological function	Essential amino acids (inferred)[a]	Essential amino acids; riboflavin (confirmed)[b,c]	Vitamins and cofactors (inferred)[d]	Essential amino acids (confirmed)[e]	B vitamins and cofactors (confirmed)[f]	Not applicable	Not applicable
Chromosome (bp)	745,590	640,681	686,192	705,557	697,724	4,639,675	4,033,464
Plasmid	1	2	0	0	1	0	0
A+T content (%)	70	74	67	73	77	49	53
Coding content (%)	83	88	89	84	89	88	88
Predicted proteins	611 (8)	564 (10)	596	583	611 (6)	4,243	3,835
Ribosomal RNAs	9	3	6	3	6	22	25
tRNAs	37	32	39	37	34	89	98
Small RNA genes	3	4	2	3	2	74	34
Average pI of proteins	8.4	9.4	8.6	8.9	9.8	6.9	6.8
Pseudogenes	36	13	9	6	14	99	159
Insertion sequence (IS) elements	0	0	0	0	0	42	16

[a] This study.
[b] Douglas (1998).
[c] Nakabachi and Ishikawa (1999).
[d] Wu et al. (2006).
[e] Feldhaar et al. (2007).
[f] Nogge (1981).

flies, on the other hand, live on vertebrate blood, where the symbiont mainly provides B vitamins to the host (Nogge 1981; Akman et al. 2002). The common nutritional physiology of the host insects may be relevant to the higher similarity of gene content between *Ishikawaella* and *Buchnera* than that between *Ishikawaella* and *Wigglesworthia* (fig. 7), although *Buchnera* is phylogenetically closer to *Wigglesworthia* than to *Ishikawaella* (fig. 3). We suggest that this may be an example of convergent genomic evolution in response to the common ecological necessity (e.g., Xu et al. 2007; López-Sánchez et al. 2009; McCutcheon and Moran 2010).

Differences in Gene Repertoire between *Ishikawaella* and Endocellular Insect Symbionts

Although the gene reduction patterns in *Ishikawaella* were strikingly similar to those in the endocellular symbionts, detailed pairwise comparisons revealed several noteworthy differences that may be relevant to functional and ecological aspects of these insect symbionts. For example, *Ishikawaella* possessed more genes for the synthesis of amino acids and cofactors than *Buchnera* and *Blochmannia* (supplementary table S6A and C, Supplementary Material online), which may indicate the broader nutritional capability of the extracellular symbiont than the endocellular symbionts and/or the younger coevolutionary history of the former than the latter. Meanwhile, *Ishikawaella* possessed more amino acid genes and less cofactor genes than *Baumannia*

and *Wigglesworthia* (supplementary table S6C and D, Supplementary Material online), which clearly reflects their distinct biological roles; *Ishikawaella* as supplier of essential amino acids, whereas *Baumannia* and *Wigglesworthia* as suppliers of cofactors (Nogge 1981; Akman et al. 2002; Wu et al. 2006). Unexpectedly, the extracellular symbiont *Ishikawaella* retained fewer genes related to cell wall synthesis and lipid metabolism than the endocellular symbionts *Blochmannia*, *Baumannia*, and *Wigglesworthia* (supplementary table S6B–D, Supplementary Material online; fig. 4). How *Ishikawaella* stands the extracellular condition with a small number of cell wall and membrane genes is an enigma. A previous histological work revealed that a symbiont capsule of *M. punctatissima* consists of three structural components: symbiont cells, secretion matrix, and chitinous envelope. The symbiont cells are embedded in the secretion matrix, and the envelope layer covers the surface of the capsule (Hosokawa et al. 2005). It seems that, although speculative, the secretion matrix may play some roles in symbiont preservation outside the host body. *Ishikawaella* exhibited more genes related to energy production and conversion than the endocellular symbionts (supplementary table S6A–D, Supplementary Material online), which was attributable to, at least partly, the complete TCA cycle genes in *Ishikawaella* in contrast to lack of those genes in the endocellular symbionts (supplementary fig. S3, Supplementary Material online; figs. 4 and 5). Plausibly, availability of

Metabolic pathway	Ishikawaella	Buchnera	Baumannia	Blochmannia	Wigglesworthia	E. coli	V. cholerae
Amino acid biosynthesis							
Essential amino acid							
Arginine [10]	10	10	0	1	0	10	10
Threonine [4]	4	4	0	4	2	4	4
Methionine [4]	4	1	4	4	0	4	4
Isoleucine [6]	5	4	0	6	0	6	6
Valine [5]	4	4	0	5	0	5	5
Leucine [9]	8	8	0	9	0	9	9
Lysine [9]	9	9	1	8	7	9	9
Histidine [8]	8	8	8	8	0	8	8
Phenylalanine [2]	2	1	1	2	1	2	2
Tryptophan [5]	5	5	0	5	0	5	5
Non-essential amino acid							
Tyrosine [2]	2	0	1	2	1	2	2
Cysteine [2]	2	2	2	2	0	2	2
Glycine [1]	1	1	1	1	1	1	1
Aspartate [1]	1	0	1	1	1	1	1
Glutamate [1]	1	2*(2)	1	0	1	1	1
Alanine [1]	1	1	1	1	1	1	1
Asparagine [1]	0	0	0	0	1	1	1
Glutamine [1]	0	0	0	1	1	1	1
Proline [3]	0	0	0	0	0	3	3
Serine [3]	1	1	1	1	1	3	3
Coenzyme biosynthesis							
Biotin [4]	4	3	4	0	4	4	4
Tetrahydrofolate [9]	8	2	9	9	6	9	9
Coenzyme A [8]	4	5	7	1	7	8	8
Flavin [6]	5	6	6	6	6	6	6
Pyridoxal 5'-phosphate [7]	6	3	7	7	7	7	7
Ubiquinone [9]	1	0	0	8	8	9	9
NAD [5]	0	2	5	0	4	5	5
Molybdopterin [8]	6	0	0	0	0	8	7
Fe-S cluster assembly [6]	6	4*(5)	6	6	6	6	5*(5)
Glutathione [2]	2	2	2	0	2	2	2
Heme [10]	4	2	4	3	10	10	10
Thiamine [6]	0	1	6	0	6	6	6
Lipoate [2]	2	2	2	0	2	2	2
Nucleotide biosynthesis							
Purine nucleotide [10]	10	8*(9)	10	10	10	10	10
Pyrimidine nucleotide [18]	18	16	18	11*(12)	18	18	17*(17)
Carbohydrate metabolism							
Tricarboxylic acid cycle [14]	14	3	5	11	11	14	14
Pentose-phosphate pathway [7]	7	7	7	7	4	7	7
Glycolysis [9]	9	9	9	9	8	9	9
Energy metabolism							
NADH dehydrogenase [12]	12	12	12	12	1*(1)	1*(1)	1*(1)
Cytochrome oxidase [4]	4	4	4	4	4	4	4
F_0F_1 ATP synthase [8]	8	8	8	8	8	8	8
Other molecules metabolism							
Chorismate [7]	7	7	7	7	2	7	7
Lipid metabolism							
Fatty acid [11]	11	5	11	11	11	11	10
Polyisoprenoid [8]	1	8	8	4	8	8	8
Phospholipid [7]	2	1	7	6	7	7	7
Cell wall structure							
UDP-N-acetyl-D-glucosamine [3]	0	3	3	2	3	3	3
KDO2-Lipid A [13]	0	0	0	11	10	13	13
ADP-heptose [4]	0	2	0	3	0	4	4
Peptidoglycan [10]	9	9	10	9	10	10	10
Flagellar							
Flagellar apparatus [32]	0	23	0	0	30	32	28
Transport							
Lipoprotein transport [5]	0	2	5	4	5	5	5
LPS transport [5]	0	0	0	5	4	5	5
Sugar transport							
Glucose uptake [4]	0	4	2	2	0	4	4
Mannitol uptake [1]	0	1	0	0	0	1	1
Mannose uptake [5]	0	0	5	5	0	5	2
Galactose (+Glucose) uptake [1]	1	0	0	0	0	1	0

Fig. 4.—Comparison of the metabolic gene repertoire between *Ishikawaella*, insect endocellular symbionts, and free-living γ-proteobacteria. The minimal number of genes for a metabolic pathway is shown in each of the brackets. Color indicates the ratio of retained genes to the minimal gene set for a metabolic pathway: green for 100%, orange for 99–75%, yellow for 74–50%, pink for 49–25%, and gray for 24–0%. Asterisk denotes that the bacterium possesses an alternative pathway for biosynthesis of the final product. Number in the parentheses shows the minimal number of genes for the alternative pathway.
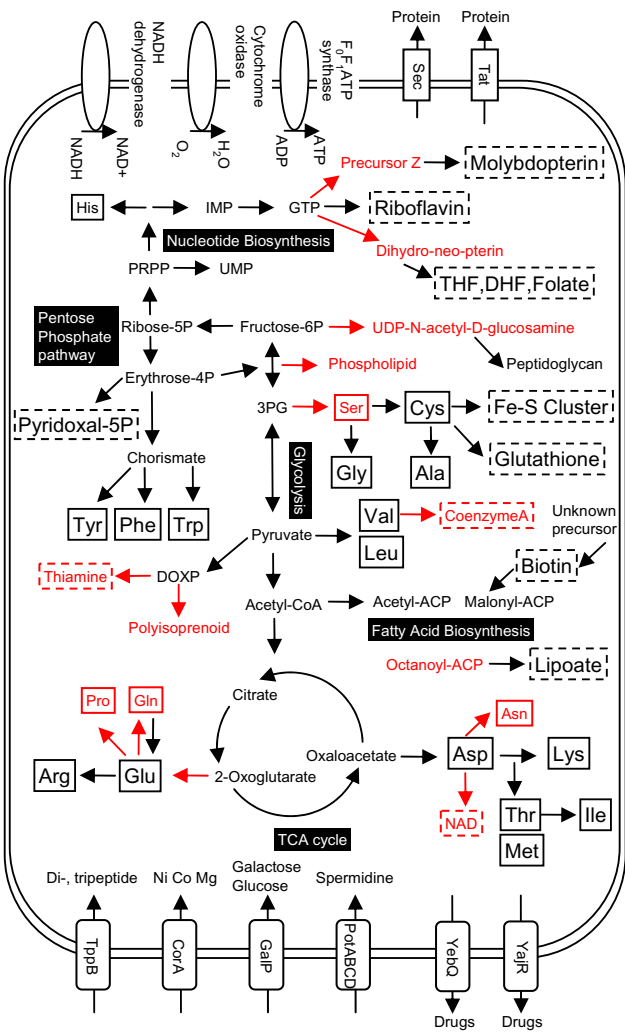
metabolic intermediates in the host cytoplasm might have resulted in the evolutionary consequence in the endocellular symbionts, whereas the extracellular symbiont has to retain its own metabolic genes. *Ishikawaella* possessed no flagellar-related cell motility genes, whereas *Buchnera* and *Wigglesworthia* retained many of them (supplementary table S6A and D, Supplementary Material online; fig. 4). In *Buchnera*, numerous flagellar basal bodies were found on the cell membrane, which are suggested to mediate material transport from and to the host cytoplasm (Maezawa et al. 2006). Because reproduction of tsetse flies entails adenotrophic viviparity, *Wigglesworthia* might require flagellar motility for vertical transmission via milk gland secretion (Akman et al. 2002; Attardo et al. 2007). Harbored extracellularly in the isolated gut cavity without food flow (Hosokawa et al. 2005, 2006), *Ishikawaella* may require neither such transporters nor flagellar motors.

Origin of *Ishikawaella*: Specialized Gut Bacterium or Ex-endocellular Symbiont?

Whether the ancestor of *Ishikawaella* was an extracellular gut bacterium or an endocellular bacterium is an unanswered question. Its extracellular habitat in the midgut cavity favors the hypothesis that *Ishikawaella* is a highly specialized gut bacterium. On the other hand, molecular phylogenetic analyses based on 50 ribosomal protein sequences revealed that *Ishikawaella* is nested in a clade of endocellular insect symbionts of mutualistic nature with drastically reduced genomes, such as *Buchnera*, *Blochmannia*, *Baumannia*, and *Wigglesworthia* (fig. 3), which raises an alternative hypothesis that their common ancestor was endocellular and *Ishikawaella* established the extracellular habitat secondarily. However, we note that the phylogenetic pattern requires careful interpretation: We cannot rule out the possibility that the apparent grouping of *Ishikawaella* with the insect endocellular symbionts was caused by their fast-evolving AT-rich gene sequences via the artifactual effect so-called long-branch attraction (Herbeck et al. 2005). To address this question, more genome data and sophisticated analyses of allied γ-proteobacterial endocellular and extracellular insect symbionts are needed.

Estimation of Genome Contents and Gene Losses in the Evolutionary Course of *Ishikawaella* and Allied Endocellular Insect Symbionts

Regardless of its endocellular or extracellular symbiotic status, the common ancestor of the *Baumannia–Ishikawaella–Buchnera–Blochmannia–Wigglesworthia* clade must have

Fig. 5.—An overview of the *Ishikawaella* metabolism and transport. The main elements of metabolic pathways and transporters that are retained and lost in the *Ishikawaella* genome are shown in black and red, respectively. Amino acids are in solid boxes. Vitamins and coenzymes are in dashed boxes.

estimated number of OGG losses was allocated to each ancestral branch (fig. 8; supplementary table S8, Supplementary Material online). The common ancestor of the insect symbionts was estimated to possess 1,021 OGGs, which is equivalent to genome size of about 1.2 Mb. As for *Ishikawaella*, at least 343 OGGs were lost and 33 OGGs were pseudogenized after divergence from the other insect symbionts, resulting in a 0.75 Mb genome with 608 OGGs. In the lineage leading to *Ishikawaella*, genes of the following categories were preferentially lost: Cell wall/membrane/envelope-related genes involved in synthesis of lipopolysaccharide, peptidoglycan, and outer membrane; cell mortility–related genes responsible for flagellar formation; and lipid and ion transport–related genes (fig. 8; supplementary table S8, Supplementary Material online). In each of the lineages leading to the other insect symbionts, preferential loss of specific gene sets was identified as follows: amino acid–related genes, cell wall/membrane/envelope-related genes, and cell motility–related genes in the *Baumannia* lineage; amino acid–related genes in the lineage leading to the *Wigglesworthia* lineage; cell motility–related genes in the *Blochmannia* spp. lineage; coenzyme-related genes, lipid and ion transport–related genes, and cell wall/membrane/envelope-related genes in the *Buchnera* spp. lineage; and coenzyme-related genes, nucleotide-related genes, and cell wall/membrane/envelope-related genes in the *Buchnera* str. Cc lineage where further genome reduction occurred (Perez-Brocal et al. 2006) (fig. 8; supplementary table S8, Supplementary Material online). Lastly, we note that the above arguments are based on the assumption that the symbiont phylogeny is correct, whereas the fast-evolving and AT-rich symbiont genomes are potentially prone to long-branch attraction and other artifacts in phylogenetic inferences (Herbeck et al. 2005).

Plasmid-Encoded Genes of *Ishikawaella*

The plasmid of *Ishikawaella*, pAst, carries four genes, *astC*, *astA*, *astB*, and *astD* (supplementary fig. S1, Supplementary Material online), which encode enzymes of the ammonia-producing succinyltransferase (AST) pathway that yield glutamate and ammonia from arginine (Schneider et al. 1998). In many bacteria, the *ast* operon generally consists of five genes, *astC*, *astA*, *astD*, *astB*, and *astE*. In the case of *Ishikawaella*, the fifth gene, *astE*, is separately encoded in the chromosome (supplementary table S1, Supplementary Material online). Upon nitrogen starvation, the AST pathway plays a principal role in *E. coli* for utilizing arginine as nitrogen source (Schneider et al. 1998). The AST pathway may play a similar role in *Ishikawaella*, although it should be verified whether the food of the plataspid stinkbug, phloem sap of leguminous plants, contains a sufficient quantity of arginine for that purpose.

already experienced remarkable genome reduction. In an attempt to gain insights into the evolutionary process of the genome reduction, on the basis of whole genome sequences of *Ishikawaella* and allied eight endocellular symbionts, we inferred a phylogenetic reconstruction of gene repertoires for common ancestors of the insect symbionts. All protein-coding genes in the genomes were organized into 1,021 OGGs based on reciprocal Blast searches, of which 966 possessed homologs in *E. coli* genome, 48 represented homologs in genomes of enteric bacteria other than *E. coli*, and 7 were unique to genomes of the insect symbionts and their close relatives (supplementary table S7, Supplementary Material online). On the basis of the phylogenetic relationship (fig. 3), estimated number of OGGs was mapped on each ancestral node, and

A Essential amino acid biosynthetic pathways

B Non-essential amino acid biosynthetic pathways

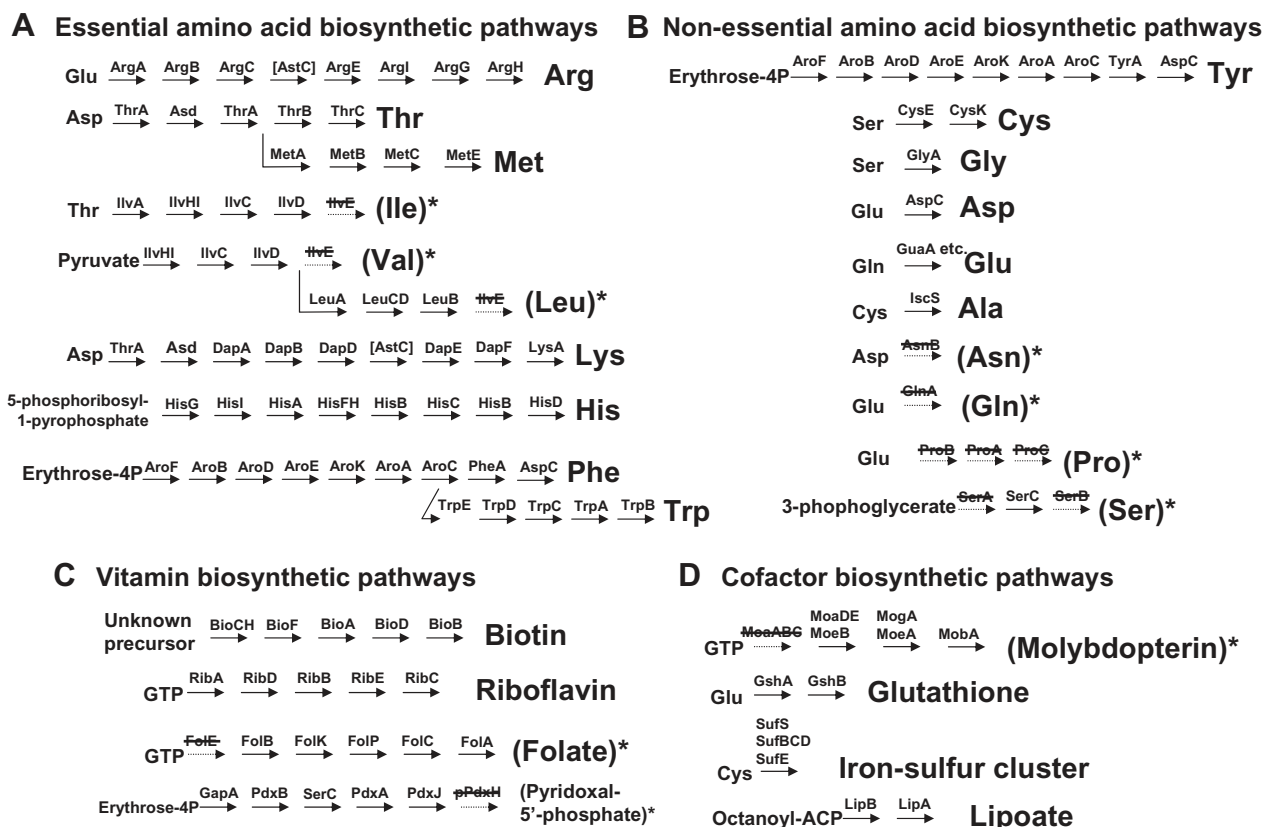**C Vitamin biosynthetic pathways**

D Cofactor biosynthetic pathways

Fıɢ. 6.—Biosynthetic pathways of essential amino acids (A), nonessential amino acids (B), vitamins (C), and cofactors (D) retained in the *Ishikawaella* genome. Plasmid-encoded enzyme genes are in brackets. Parentheses indicate that their synthetic pathways encoded in the *Ishikawaella* genome are incomplete, whereas asterisks imply that the missing final step enzymes (strikethrough) are probably complemented by corresponding enzymes of either *Ishikawaella* or host insect origin (see supplementary table S5, Supplementary Material online).

It is also notable that *Ishikawaella* may utilize *astC* for transamination in the synthetic pathways of arginine and lysine (fig. 6). In *E. coli*, not only *astC* but also *argD* catalyze these reactions (Riley and Glansdorff 1983), but in *Ishikawaella*, *argD* is encoded neither in the plasmid nor in the chromosome.

The plasmid pAst also contains *ode*, a gene encoding oxalate decarboxylase (supplementary fig. S1, Supplementary Material online), which is widely distributed among fungi and bacteria (Mäkelä et al. 2009). Oxalate is commonly present in higher plants and often accumulated at substantial concentrations in the plant biomass, the calcium salt of which can function as antiherbivore defensive agent (Franceschi and Nakata 2005). In the case of human, food-derived oxalate is degraded by a gut bacterium *Oxalobacter formigenes* (Allison et al. 1985), and missing of this gut bacterium increases the risk of hyperoxaluria and relevant disorders, such as the development of calcium oxalate kidney stones (Sidhu et al. 1999). The possibility that the plasmid of *Ishikawaella* may play a similar detoxifying role should be taken into account in future studies.

Conclusion and Perspective

Taking all these results together, we conclude that drastic genome reduction can occur even in gut symbiotic associations. In this study, we demonstrated that general genomic features of the extracellular stinkbug symbiont *Ishikawaella* are very similar to those of allied endocellular insect symbionts, such as *Buchnera*, *Baumannia*, *Blochmannia*, and *Wigglesworthia*. Such genomic features include general gene repertoire, genome size, AT richness, elevated evolutionary rate, paucity of mobile genetic elements, etc. On the other hand, there are some differences between the gene repertoire of *Ishikawaella* and those of the endocellular symbionts, which may be relevant to functional and physiological adaptations of the respective symbionts and can be attributed to preferential gene losses in each of the symbiont lineages. The strikingly similar evolutionary patterns shared between the extracellular and endocellular insect symbionts despite their apparently distinct ecological niches suggest some common evolutionary processes underpinning the symbiotic associations, which may be relevant to stable and nutrition-rich environment in the host
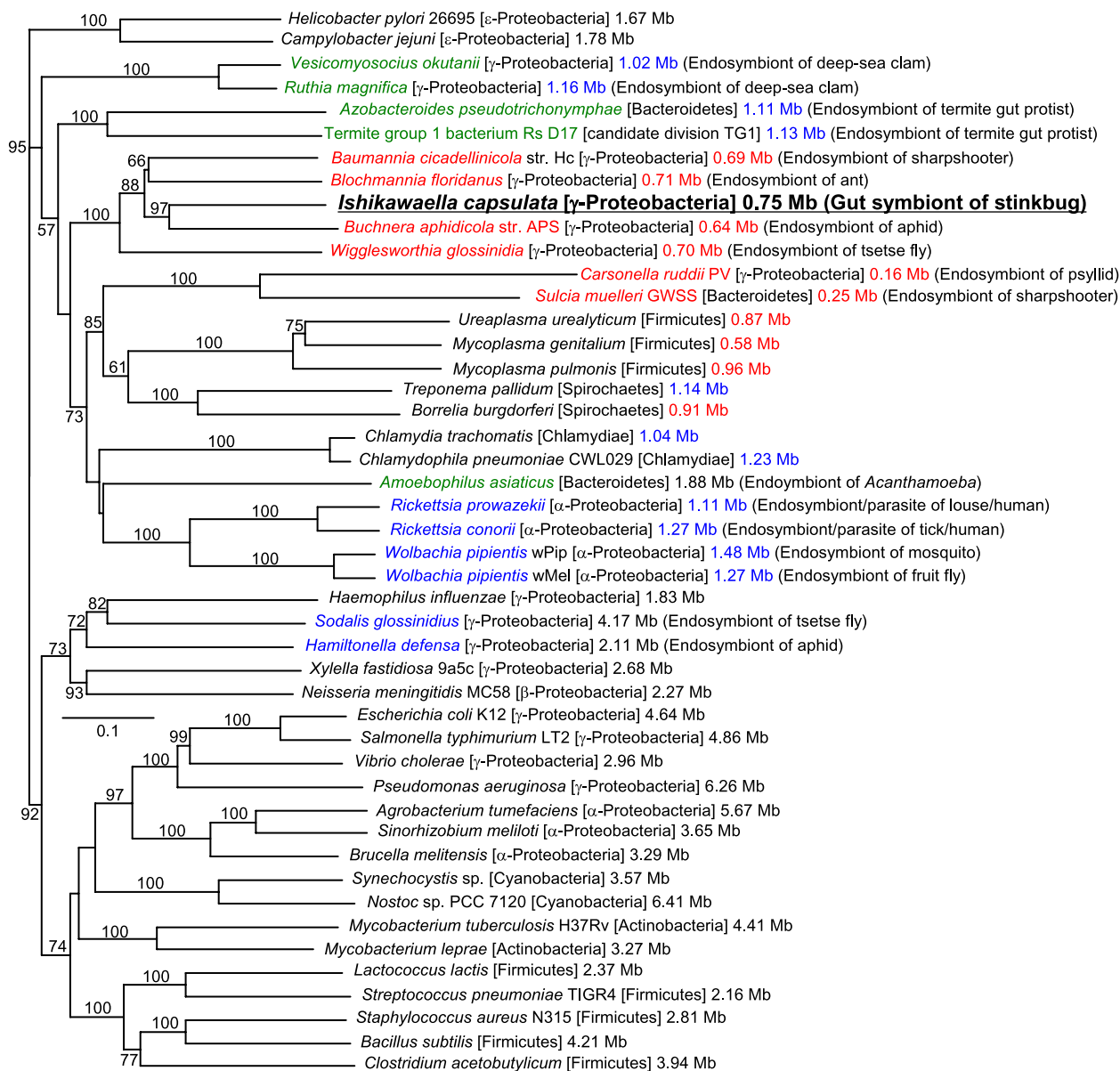
Fig. 7.—Cluster analysis of 47 bacterial genomes, including the *Ishikawaella* genome, on the basis of their gene repertoire. An index reflecting gene content similarity was calculated for each of all pairs of the bacterial genomes, a distance matrix was constructed from the similarity indices, and the bacterial genomes were clustered into a tree topology under the neighbor-joining algorism. Bacterial names are shown in italic; in brackets are bacterial phyla; in parentheses are host organisms for endosymbionts. Colored bacterial names indicate red, obligate endocellular insect symbionts; blue, facultative endocellular insect symbionts/parasites; and green, obligate endocellular symbionts of non-arthropod organisms. Colored genome sizes indicate red, smaller than 1.0 Mb; blue, smaller than 1.5 Mb.

body and also attenuated purifying selection due to small population size and strong bottleneck associated with the lifestyle of the vertically transmitted symbionts (Wernegreen 2002; Moran et al. 2008; Moya et al. 2008).

We point out that the secretion matrix embedding *Ishikawaella* cells within symbiont capsules of the plataspid stinkbug (Hosokawa et al. 2005) may be an important factor that has contributed to the reductive evolution of the symbiont genome in the gut cavity. It seems likely, although

speculative, that the secretion matrix is somehow mimicking the intracellular environment in the cytoplasm, protecting *Ishikawaella* against dehydration, irradiation, and shortage of metabolites that the symbiont cannot make on its own. In this context, biochemical composition and biological function of the secretion matrix are to be investigated in depth.

Are there any substantial differences between the evolutionary consequences of the extracellular and endocellular
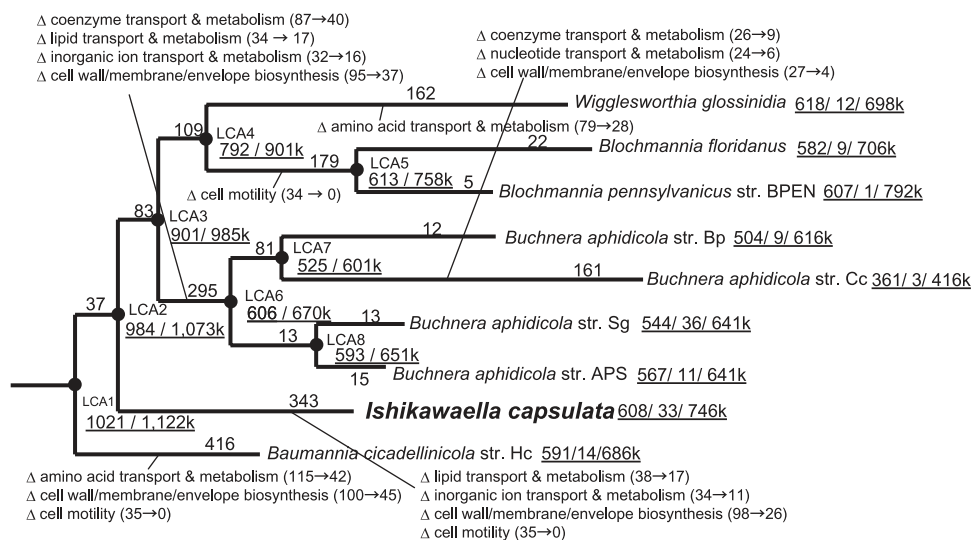
Fig. 8.—Estimation of the ancestral gene contents and gene losses in the evolutionary course of *Ishikawaella* and allied endocellular insect symbionts. The phylogenetic tree is a part of figure 3. Number above each of the branches is the number of lost OGGs since the last node. Underlined number at each of the nodes is the number of OGGs present/the expected genome size at the node. Numbers on the right side of each bacterial name indicate number of intact OGGs/number of pseudogenes/the genome size. COG categories remarkably lost and/or pseudogenized in specific branches (highlighted in supplementary table S8, Supplementary Material online) are shown.

symbiotic associations? Years ago, most sequenced genomes of obligate insect symbionts ranged from 0.6 to 1 Mb in size and contained more than 500 genes (see table 1). These values are similar to the smallest known pathogen genomes (see fig. 3), and the *Ishikawaella* genome falls into that range. However, recent studies revealed that some endocellular insect symbionts exhibit genome sizes smaller than 0.5 Mb down to 0.14 Mb, which are almost approaching to the extremely streamlined genomes of organelles (Nakabachi et al. 2006; Perez-Brocal et al. 2006; McCutcheon and Moran 2007, 2010; McCutcheon et al. 2009; McCutcheon 2010). It is difficult to imagine that such suborganellar microbes survive extracellularly in the gut cavity. How small such extracellular gut symbiont genomes can be is of evolutionary interest, deserving future surveys of obligate gut symbiotic bacteria of diverse stinkbugs and other organisms.

Finally, we point out an applied perspective of the *Ishikawaella* genome. A previous study demonstrated that the pest status of plataspid stinkbugs is determined by the *Ishikawaella* genotype rather than by the insect genotype (Hosokawa, Kikuchi, Shimada, et al. 2007). Comparative analyses of the *Ishikawaella* genomes from the pest and nonpest stinkbug species would shed light on the mechanisms underlying the symbiont-mediated pest evolution.

Supplementary Material

Supplementary tables S1–S8 and figures S1–S3 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Literature Cited

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

Akman L, et al. 2002. Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. Nat Genet. 32:402–407.

Allison MJ, Dawson KA, Mayberry WR, Foss JG. 1985. *Oxalobacter formigenes* gen. nov., sp. nov: oxalate-degrading anaerobes that inhabit the gastrointestinal tract. Arch Microbiol. 141:1–7.

Attardo GM, Guz N, Strickler-Dinglasan P, Aksoy S. 2007. Molecular aspects of viviparous reproductive biology of the tsetse fly (*Glossina morsitans morsitans*): regulation of yolk and milk gland protein synthesis. J Insect Physiol. 52:1128–1136.

Baumann P. 2005. Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. Annu Rev Microbiol. 59:155–189.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27:573–580.

Bjellqvist B, et al. 1993. The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino-acid sequences. Electrophoresis. 14:1023–1031.

Bourtzis K, Miller TA. 2003. Insect symbiosis. Boca Raton (FL): CRC Press.

Buchner P. 1965. Endosymbiosis of animals with plant microorganisms. New York: Interscience.

Burke GR, Moran NA. 2011. Massive genomic decay in *Serratia symbiotica*, a recently evolved symbiont of aphids. Genome Biol Evol. 3:195–208.

Degnan PH, Lazarus AB, Wernegreen JJ. 2005. Genome sequence of *Blochmannia pennsylvanicus* indicates parallel evolutionary trends among bacterial mutualists of insects. Genome Res. 15:1023–1033.

Degnan PH, et al. 2010. Dynamics of genome evolution in facultative symbionts of aphids. Environ Microbiol. 12:2060–2069.

Degnan PH, Yu Y, Sisneros N, Wing RA, Moran NA. 2009. *Hamiltonella defensa*, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. Proc Natl Acad Sci U S A. 106:9063–9068.

Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. 27:4636–4641.

Douglas AE. 1998. Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria *Buchnera*. Annu Rev Entomol. 43:17–37.

Feldhaar H, et al. 2007. Nutritional upgrading for omnivorous carpenter ants by the endosymbiont *Blochmannia*. BMC Biol. 5:48.

Felsenstein J. 2005. PHYLIP (Phylogeny Inference Package) version 3.6. Seattle (WA): Department of Genetics, University of Washington. [Internet]. [cited 2011 July 11]. Available from: http://evolution.genetics.washington.edu/phylip.html

Franceschi VR, Nakata PA. 2005. Calcium oxalate in plants: formation and function. Annu Rev Plant Biol. 56:41–71.

Gil R, et al. 2003. The genome sequence of *Blochmannia floridanus*: comparative analysis of reduced genomes. Proc Natl Acad Sci U S A. 100:9388–9393.

Gordon D, Desmarais C, Green P. 2001. Automated finishing with autofinish. Genome Res. 11:614–625.

Grimardi DA, Engel MS. 2005. Evolution of the insects. Cambridge: Cambridge University Press.

Herbeck JT, Degnan PH, Wernegreen JJ. 2005. Nonhomogeneous model of sequence evolution indicates independent origins of primary endosymbionts within the enterobacteriales (γ-Proteobacteria). Mol Biol Evol. 22:520–532.

Hosokawa T, Kikuchi Y, Fukatsu T. 2007. How many symbionts are provided by mothers, acquired by offspring, and needed for successful vertical transmission in an obligate insect-bacterium mutualism? Mol Ecol. 16:5316–5325.

Hosokawa T, Kikuchi Y, Meng XY, Fukatsu T. 2005. The making of symbiont capsule in the plataspid stinkbug *Megacopta punctatissima*. FEMS Microbiol Ecol. 54:471–477.

Hosokawa T, et al. 2010. Phylogenetic position and peculiar genetic traits of a midgut bacterial symbiont of the stinkbug *Parastrachia japonensis*. Appl Environ Microbiol. 76:4130–4135.

Hosokawa T, Kikuchi Y, Nikoh N, Shimada M, Fukatsu T. 2006. Strict host-symbiont cospeciation and reductive genome evolution in insect gut bacteria. PLoS Biol. 4:e337.

Hosokawa T, Kikuchi Y, Shimada M, Fukatsu T. 2007. Obligate symbiont involved in pest status of host insect. Proc R Soc B Biol Sci. 274:1979–1984.

Kanehisa M, et al. 2008. KEGG for linking genomes to life and the environment. Nucleic Acids Res. 36:D480–D484.

Katoh K, Kuma K, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33:511–518.

Keseler IM, et al. 2009. EcoCyc: a comprehensive view of *Escherichia coli* biology. Nucleic Acids Res. 37:D464–D470.

Kikuchi Y, et al. 2009. Host-symbiont co-speciation and reductive genome evolution in gut symbiotic bacteria of acanthosomatid stinkbugs. BMC Biol. 7:2.

Kirkness EF, et al. 2010. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. Proc Natl Acad Sci U S A. 107:12168–12173.

Klasson L, et al. 2008. Genome evolution of *Wolbachia* strain *w*Pip from the *Culex pipiens* group. Mol Biol Evol. 25:1877–1887.

Klasson L, et al. 2009. The mosaic genome structure of the *Wolbachia* *w*Ri strain infecting *Drosophila simulans*. Proc Natl Acad Sci U S A. 106:5725–5730.

López-Sánchez MJ, et al. 2009. Evolutionary convergence and nitrogen metabolism in *Blattabacterium* strain Bge, primary endosymbiont of the cockroach *Blattella germanica*. PLoS Genet. 5:e1000721.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25:955–964.

Maezawa K, et al. 2006. Hundreds of flagellar basal bodies cover the cell surface of the endosymbiotic bacterium *Buchnera aphidicola* sp strain APS. J Bacteriol. 188:6539–6543.

Mäkelä MR, Hildén K, Hatakka A, Lundell TK. 2009. Oxalate decarboxylase of the white-rot fungus *Dichomitus squalens* demonstrates a novel enzyme primary structure and non-induced expression on wood and in liquid cultures. Microbiology 155:2726–2738.

McCutcheon JP. 2010. The bacterial essence of tiny symbiont genomes. Curr Opin Microbiol. 13:73–78.

McCutcheon JP, McDonald BR, Moran NA. 2009. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. Proc Natl Acad Sci U S A. 106:15394–15399.

McCutcheon JP, Moran NA. 2007. Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. Proc Natl Acad Sci U S A. 104:19392–19397.

McCutcheon JP, Moran NA. 2010. Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. Genome Biol Evol. 2:708–718.

Moran NA, McCutcheon JP, Nakabachi A. 2008. Genomics and evolution of heritable bacterial symbionts. Annu Rev Genet. 42:165–190.

Moya A, Peretó J, Gil R, Latorre A. 2008. Learning how to live together: genomic insights into prokaryote-animal symbioses. Nat Rev Genet. 9:218–229.

Nakabachi A, Ishikawa H. 1999. Provision of riboflavin to the host aphid, *Acyrthosiphon pisum*, by endosymbiotic bacteria, *Buchnera*. J Insect Physiol. 45:1–6.

Nakabachi A, et al. 2006. The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. Science 314:267.

Nogge G. 1981. Significance of symbionts for the maintenance of an optimal nutritional state for successful reproduction in hematophagous arthropods. Parasitology 82:101–104.

Perez-Brocal V, et al. 2006. A small microbial genome: the end of a long symbiotic relationship? Science 314:312–313.

Riley M, Glansdorff N. 1983. Cloning the *Escherichia coli* K-12 *argD* gene specifying acetylornithine δ-transaminase. Gene. 24:335–339.

Robinson-Rechavi M, Huchon D. 2000. RRTree: relative-rate tests between groups of sequences on a phylogenetic tree. Bioinformatics 16:296–297.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.

Sabree ZL, Kambhampati S, Moran NA. 2009. Nitrogen recycling and nutritional provisioning by *Blattabacterium*, the cockroach endosymbiont. Proc Natl Acad Sci U S A. 106:19521–19526.

Schneider BL, Kiupakis AK, Reitzer LJ. 1998. Arginine catabolism and the arginine succinyltransferase pathway in *Escherichia coli*. J Bacteriol. 180:4278–4286.

Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H. 2000. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. Nature 407:81–86.

Sidhu H, et al. 1999. Direct correlation between hyperoxaluria/oxalate stone disease and the absence of the gastrointestinal tract-dwelling bacterium *Oxalobacter formigenes*: possible prevention by gut recolonization or enzyme replacement therapy. J Am Soc Nephrol. 10:S334–S340.

Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res. 34:D32–D36.

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

Tatusov RL, et al. 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4:41.

Toh H, et al. 2006. Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. Genome Res. 16:149–156.

van Ham RCHJ, et al. 2003. Reductive genome evolution in *Buchnera aphidicola*. Proc Natl Acad Sci U S A. 100:581–586.

Wernegreen JJ. 2002. Genome evolution in bacterial endosymbionts of insects. Nat Rev Genet. 3:850–861.

Wilkes TE, et al. 2010. The draft genome sequence of *Arsenophonus nasoniae*, son-killer bacterium of *Nasonia vitripennis*, reveals genes associated with virulence and symbiosis. Insect Mol Biol. 19(S1):59–73.

Wolf YI, Rogozin IB, Grishin NV, Koonin EV. 2002. Genome trees and the tree of life. Trends Genet. 18:472–479.

Wu D, et al. 2006. Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. PLoS Biol. 4:e188.

Wu M, et al. 2004. Phylogenomics of the reproductive parasite *Wolbachia pipientis w*Mel: a streamlined genome overrun by mobile genetic elements. PLoS Biol. 2:e69.

Xu J, et al. 2007. Evolutionary symbiotic bacteria in the distal human intestine. PLoS Biol. 5:e156.

Associate editor: Richard Cordaux