Molecular Biology and Evolution, 2025, **42**, 1–14
https://doi.org/10.1093/molbev/msaf125
Advance access publication 30 May 2025
Discoveries

MBE OXFORD

Accelerated Pseudogenization in the Ancient Endosymbionts of Giant Scale Insects

Jinyeong Choi [1,*] **Pradeep Palanichamy** [1] **Hirotaka Tanaka** [2,3] **Takumasa Kondo** [4] **Matthew E. Gruwell** [5] **Filip Husnik** [1,*]

[1]Evolution, Cell Biology, and Symbiosis Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa 904-0495, Japan
[2]Faculty of Agriculture, Ehime University, Matsuyama, Ehime 790-8566, Japan
[3]The Kyushu University Museum, Higashi-ku, Fukuoka 812-8581 Japan
[4]Corporación Colombiana de Investigación Agropecuaria-Agrosavia, Centro de Investigación Palmira, Valle del Cauca, Colombia
[5]Behrend College, School of Science, Penn State Erie, Erie, PA, USA

***Corresponding authors**: E-mails: cjy0784@gmail.com; filip.husnik@oist.jp.
Associate editor: Purificacion Lopez-Garcia

Abstract

Symbiotic microorganisms are subject to a complex interplay of environmental and population-genetic pressures that drive their gene loss. Despite the widely held perception that ancient symbionts have stable genomes, even tiny genomes experience ongoing pseudogenization. Whether these tiny genomes also experience bursts of rapid gene loss is, however, less understood. Giant scale insects (Monophlebidae) feed on plant sap and rely on the symbiotic bacterium *Walczuchella*, which provides them with essential nutrients. When compared with other ancient symbionts with similar genome sizes, such as *Karelsulcia*, *Walczuchella*'s genome was previously reported as unusually pseudogene-rich (10% of coding sequences). However, this result was based on only one genome assembly, raising questions about the assembly quality or a recent ecological shift such as co-symbiont acquisition driving the gene loss. Here, we generated six complete genomes of *Walczuchella* from three genera of giant scales, each with distinct co-symbiotic partners. We show that all the genomes are highly degraded, and particularly genes related to the cellular envelope and energy metabolism seem to be undergoing pseudogenization. Apart from general mechanisms driving genome reduction, such as the long-term intracellular lifestyle with transmission bottlenecks, we hypothesize that a more profound loss of DNA replication and repair genes, together with recent co-obligate symbiont acquisitions, likely contribute to the accelerated degradation of *Walczuchella* genomes. Our results highlight that even ancient symbionts with small genomes can experience significant bursts of gene loss when stochastic processes erase a gene that accelerates gene loss or when the selection pressure changes such as after co-symbiont acquisition.

Keywords: genome degradation, selective pressure, symbiosis, *Walczuchella*, Bacteroidota

Introduction

Genome erosion is a hallmark of symbiotic microorganisms (Moran and Bennett 2014). Its outcome depends on an interplay of multiple evolutionary forces affecting host-restricted populations that are bottlenecked every generation (Moran 1996). Under relaxed purifying selection, deleterious mutations are more likely to occur and can become easily fixed in populations with small effective sizes and accumulate through Muller's ratchet effect on asexual populations (Moran 1996; Rispe and Moran 2000). The enhanced mutation rate is highly associated with prokaryotic genome reduction (Bourguignon et al. 2020) because the persistent accumulation of nucleotide substitutions disrupts the open reading frames of genes under relaxed selective constraints and eventually leads to the loss of their function (pseudogenization) (Li et al. 1981; Petrov and Hartl 2000). For instance, the genomes of recently evolved symbionts can contain over 50% of pseudogenes (McCutcheon and Moran 2012). Moreover, the reduction of DNA replication and repair genes in symbionts can further accelerate their pseudogenization and overall genome degradation (McCutcheon and Moran 2012).

Intracellular symbionts that are maternally transmitted often contain extremely small genomes (Moran and Bennett 2014). These ancient symbionts have often been maintained in their hosts for hundreds of millions of years due to their crucial role in synthesizing essential nutrients such as amino acids and B vitamins necessary for host development and reproduction (Moran and Bennett 2014). Despite their minimal gene sets, some ancient nutritional symbionts have been reported to experience an ongoing loss of essential genes (Moran and Bennett 2014). In hosts with degraded ancient symbionts, co-occurring microbial partners often complement important metabolic pathways of the original symbionts. Such multipartite symbiotic consortia have been observed in various plant sap-feeding insects, for example, adelgids (Dial et al. 2022), aphids (Manzano-Marín et al. 2023), mealybugs (Husnik and McCutcheon 2016; Garber et al. 2021), and diverse Auchenorrhyncha (Bennett and Moran 2013; Michalik et al. 2021). However, the establishment of these new partnerships may induce additional degradation of the ancient symbiont due to the availability of metabolites (or even proteins) provided by the new partner (Manzano-Marín et al. 2023).

Received: November 25, 2024. **Revised:** April 30, 2025. **Accepted:** May 5, 2025

The Flavobacteriales are a major group of insect symbionts, some of which have evolved long-term relationships with their hosts. Examples of these ancient symbionts include *Candidatus* Karelsulcia in Auchenorrhynchan insects and *Blattabacterium* in cockroaches and *Mastotermes* termites, which play important roles in nutrient provisioning and nitrogen recycling (Moran et al. 2005; Sabree et al. 2009). Other Flavobacteriales, such as *Candidatus* Skilesia from *Geopemphigus* aphids, and *Candidatus* Shikimatogenerans and *Candidatus* Bostrichicola from bostrichid beetles, are associated with nutritional provisioning or cuticle hardness (Chong and Moran 2018; Kiefer et al. 2023). Many species of scale insects also house symbiotic Flavobacteriales (Gruwell et al. 2007, 2010; Rosenblueth et al. 2012; Dhami et al. 2013; Szklarzewicz et al. 2020; Choi and Lee 2022). However, the flavobacterial symbionts from different scale insect families form distinct clades that are incongruent with the host phylogeny, suggesting multiple acquisitions and replacements of different symbionts during scale insect evolution (Rosenblueth et al. 2012; Vea and Grimaldi 2016; Choi and Lee 2022; Szklarzewicz et al. 2022). Unfortunately, only two genomes of *Candidatus* Walczuchella and *Candidatus* Uzinuria have been analyzed so far from these diverse flavobacterial symbionts of scale insects (Sabree et al. 2013; Rosas-Pérez et al. 2014).

Monophlebidae, also known as giant scales, is a group of scale insects comprised of 267 species in 49 genera (García Morales et al. 2016). The family, for example, includes the enigmatic cottony cushion scale, *Icerya purchasi* Maskell, which is one of a few androdioecious insects and can cause significant damage to a variety of economically important plants, particularly citrus trees (Mongue et al. 2021; Grafton-Cardwell et al. 2022). Giant scales house nutritional symbionts, including *Walczuchella* and multiple co-symbionts related to *Cedecea*, *Sodalis*, and *Wolbachia* (Matsuura et al. 2009; Rosenblueth et al. 2012 ; Rosas-Perez et al. 2017). *Walczuchella* and other co-obligate symbionts are found in the bacteriomes of the host insects and are vertically transmitted to their offspring (Matsuura et al. 2009; Rosas-Pérez et al. 2014; Rosas-Perez et al. 2017). The genome of *Walczuchella* is 309 kbp in size and contains 271 protein-coding sequences (CDSs) (Rosas-Pérez et al. 2014), but it has a higher number of pseudogenes when compared with other anciently established symbionts, which typically have small genomes with relatively few pseudogenes. Interestingly, the reason for the high level of pseudogenization in *Walczuchella* remains unclear. Here, we addressed this question via metagenome sequencing of six species of giant scales and also determined the symbiont composition of these insects.

Materials and Methods

DNA Extraction and Genome Sequencing

Six species from three genera of Monophlebidae were used for metagenome sequencing (supplementary table S1, Supplementary Material online). All the samples were preserved in 99% ethanol at −20 °C. For DNA extraction, 1 to 8 individuals of each species were selected under a dissecting microscope. To eliminate any potential surface contaminants, wax secretions were removed, and the specimens were washed with 99% ethanol. The insects were then ground in liquid nitrogen with a mortar and pestle. DNA was extracted using DNeasy Blood and Tissue Kit (Qiagen) and the MasterPure Complete DNA Purification Kit (Epicenter), following the manufacturers' protocols. DNA quantity and quality were checked with NanoDrop (Thermo Fisher Scientific) and Qubit 4 fluorometer (Invitrogen). PCR-free DNA libraries were prepared with the NEBNext Ultra II kit (NEB). The libraries were multiplexed and sequenced on the Illumina NovaSeq 6000 and MiSeq sequencers at the Okinawa Institute of Science and Technology Graduate University and Pennsylvania State University (supplementary table S1, Supplementary Material online). The quality of raw Illumina reads was assessed using FastQC v0.11.7. (Andrews 2010), and low-quality reads and adapters were removed using Fastp v.0.20.0 (Chen et al. 2018).

Genome Assembly and Annotation

Raw Illumina reads were assembled using SPAdes v3.15 (Prjibelski et al. 2020) using multiple k-mers (supplementary table S1, Supplementary Material online). The taxonomic assignment of assembled scaffolds was conducted using megablast search (NCBI-BLAST v.2.11.0) against the NCBI nucleotide database. *Walczuchella* genome scaffolds were extracted based on the taxon-annotated-GC-coverage plots generated by Blobtools v1.1.1 (Laetsch and Blaxter 2017). The circular genomes of *Walczuchella* were visually assessed with Bandage v.0.9 (Wick et al. 2015). Further statistical assessments were performed for sequencing read coverage and variant calling on the draft genome assemblies, followed by polishing using Pilon v.1.24 (Walker et al. 2014). Read coverage plots were also generated by mapping filtered reads to the assemblies of *Walczuchella*. The genome assemblies were then annotated with Prokka v.1.14.6 (Seemann 2014). Some hypothetical or unannotated proteins identified by Prokka were manually re-annotated through BLASTp searches against the NCBI RefSeq database (Pruitt et al. 2007). Remnant or missing intact CDSs were identified in intergenic regions by BLASTx implemented in Pseudofinder v1.1.0 (Syberg-Olsen et al. 2022). CDSs were sorted into clusters of orthologous groups (COGs) using eggNOG-mapper v2 (Cantalapiedra et al. 2021) with default parameters. The transfer RNAs predictions were confirmed using tRNAscan-SE v2.0.11 (Chan et al. 2021). Genome maps were created using DNAPlotter (Carver et al. 2009). A synteny plot of *Walczuchella* genomes was created using Processing3 (https://processing.org/) based on Blast genome alignments and Prokka annotations. The comparative analyses were performed using genome assemblies of other ancient symbionts obtained from NCBI.

Pseudogene Identification and Substitution Rate Estimation

The putative pseudogenes in the seven complete and circularized *Walczuchella* genomes were predicted as truncated or fragmented genes in PseudoFinder using DIAMOND BlastP/BlastX searches v.2.0.4.142 (Buchfink et al. 2021) against the nonredundant protein database. As a reference, pseudogenes were also predicted using the same settings in the published genome assemblies of other ancient symbionts. Poly(A) or (T) tracts were screened in the predicted fragmented genes of *Walczuchella* to confirm the potential transcriptional slippage that restores the disrupted reading frame [46]. Using the CDSs of the seven genomes of *Walczuchella*, we identified 190 single-copy orthologous genes with OrthoFinder v.2.5.4 (Emms and Kelly 2019). Of these, 181 single-copy orthologous

genes shared by all seven genomes were selected for substitution rate and diversifying selection analysis (after excluding pseudogenes). Two different algorithms, CODEML implemented in PAML v.4.9 (Yang 2007) and BUSTED in HYPHY v.2.5 (Pond et al. 2005; Murrell et al. 2015) were used and the results compared. The unrooted tree of the seven *Walczuchella* genomes was generated using RAxML v.8.2.12 (Stamatakis 2014) with 1,000 bootstrap replicates and the PROTGAMMAGTR model. The orthologous gene sequences were aligned by MAFFT v.7 (Katoh and Standley 2013). For the analysis with CODEML, nucleotide sequences were translated into amino acid sequences using Geneious Prime v.2023.0.4 (Kearse et al. 2012), and then aligned with MUSCLE v.5.1 (Edgar 2004). The codon-based DNA alignment was generated using PAL2NAL v.14 (Suyama et al. 2006). The software packages CODEML (M0 model) and BUSTED were used to calculate the average ratio of nonsynonymous to synonymous substitutions ($\omega = dN/dS$) across the whole gene. The BUSTED analysis produced two values of dN/dS, one based on relative GTR branch lengths and nucleotide substitution biases, and the other based on improved branch lengths, nucleotide substitution biases, and a full codon model. Genes with a dN/dS ratio between 0.95 and 0.1 were considered to be evolving under relaxed purifying selection (Vasquez and Bennett 2022). dN and dS values were also obtained from CODEML. To screen the genes that have undergone positive selection at some sites, two comparisons based on the site models were performed (M1a vs. M2a and M7 vs. M8) using the constrained (M1a and M7) and unconstrained (M2a and M8) models. Positive selection was further tested using BUSTED based on the branch-site model.

Additional Analyses of DNA Replication and Repair Genes

A total of 33 genes involved in DNA replication and repair were searched against the genome assemblies of seven genera of ancient symbionts (supplementary table S4, Supplementary Material online) downloaded from NCBI, as well as *Walczuchella*. The genome assemblies were annotated using Prokka and BLASTp searches, and their orthologous genes were sorted using OrthoFinder. Further analyses were done for *dnaEQ* and *dnaN* genes of *Walczuchella*, compared with *Karelsulcia*, *Blattabacterium* and free-living *Flavobacterium*. Gene trees were reconstructed based on the aligned amino acid sequences using IQ-tree with 1,000 bootstrap replicates (Nguyen et al. 2015). Ancestral sequence reconstruction was performed using GRASP (Foley et al. 2022) to infer insertion and deletion events within the two genes, and the substitution variants were identified through marginal reconstruction in GRASP. Protein structures of *dna* genes were predicted for representative lineages of each symbiont using Phyre2 (Kelley et al. 2015). Multiple sequence alignments were visualized with the ETE toolkit (Huerta-Cepas et al. 2016).

Phylogenetic Analyses

The 16S rRNA and 23S rRNA gene sequences of symbionts were extracted using Barrnap v3 (https://github.com/tseemann/barrnap) from metagenome assemblies of giant scales and other bacterial genomes (or individual sequences) available from NCBI. The two genes were aligned individually using MAFFT v7. Ambiguously aligned positions were removed using trimAL v.1.4.1 (Capella-Gutiérrez et al. 2009) with the *gappyout* option. The trimmed alignments of the two genes were concatenated using Phyutility v.2.7.1 (Smith and Dunn 2008). For the genome-based phylogenetic analysis of *Walczuchella*, amino acid sequences of 134 single-copy orthologous genes were aligned using MAFFT and concatenated into a single matrix with Phyutility v.2.7.1. The sequence matrix was trimmed using trimAL v1.4.1 with the strict option. For the genome-scale phylogeny of host insects, the near-universal single-copy orthologs (USCOs) were obtained from each metagenome assembly using Patchwork v0.5.1 (Thalén et al. 2023). The initial set of 2,157 USCOs was recovered from the chromosome-level genome assembly of *Phenacoccus solenopsis* (Li et al. 2020) by BUSCO v.5.4.2 (Manni et al. 2021) utilizing the BUSCO sequences for Hemiptera. Using this initial USCOs set, 1,772 to 2,097 USCOs were mined from the assemblies of ingroups and outgroups using Patchwork. Protein sequences of the USCOs were aligned with MAFFT and concatenated. The concatenated dataset was trimmed with trimAL v1.4.1 with the *nogaps* option. Maximum likelihood trees for all the above datasets were inferred using IQ-tree under the best-fitting models selected automatically (Kalyaanamoorthy et al. 2017). Branch support was estimated using 1,000 replicates of the ultrafast bootstrap approximation (Hoang et al. 2018). The resulting trees were visualized using Figtree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

Fluorescence In Situ Hybridization

To examine the localization of *Walczuchella* in the host insects, fluorescence in situ hybridization (FISH) was performed on *Crypticerya multicicatrices* and *Icerya purchasi*. The 16S rRNA probe CFB319 (Moran et al. 2005) that matches *Walczuchella* was used. All insect samples were fixed in 4% paraformaldehyde in phosphate-buffered saline (PBS-1X) and Carnoy's solution overnight, respectively. The samples were washed with 80% ethanol and bleached with 6% hydrogen peroxide solution (Koga et al. 2009) for 3 weeks with several replacements of the solution. They were then washed with absolute ethanol and PBST. Except for the samples to be whole-mounted, the remaining tissues were embedded in paraffin and sectioned to 10 µm with a rotary microtome (HM 340E, Epredia). They were dewaxed with Clear Plus (Falma, Japan) and rehydrated with 100%, 70%, and 50% ethanol series, after drying the sections on slides. The sections were incubated with 300 µL of hybridization buffer containing DAPI (1 µg/µL) and probes (100 nM) in a humidity chamber overnight at 45 °C. The samples were washed with PBST and mounted with ProLong Diamond Antifade Mountant (Thermo Fisher Scientific), and the tissues were examined under the Nikon Eclipse Ti2-E inverted microscope.

Results

Characteristics of *Walczuchella* Genomes

All genomes of *Walczuchella* from six host species were closed into circular-mapping chromosomes (Fig. 1a; Table 1). The genome size of *Walczuchella* ranges from 281,644 bp to 309,299 bp, including the largest genome, *Walczuchella* LLAX, previously reported in Rosas-Pérez et al. (2014). The average mapped read coverage of the assemblies ranges from 12 to 2,939. The coverage plots showed overall uniform depth across the assemblies, except for regions with assembly gaps (represented as "N"s), single nucleotide polymorphisms, or repetitive sequences (supplementary fig. S1, Supplementary

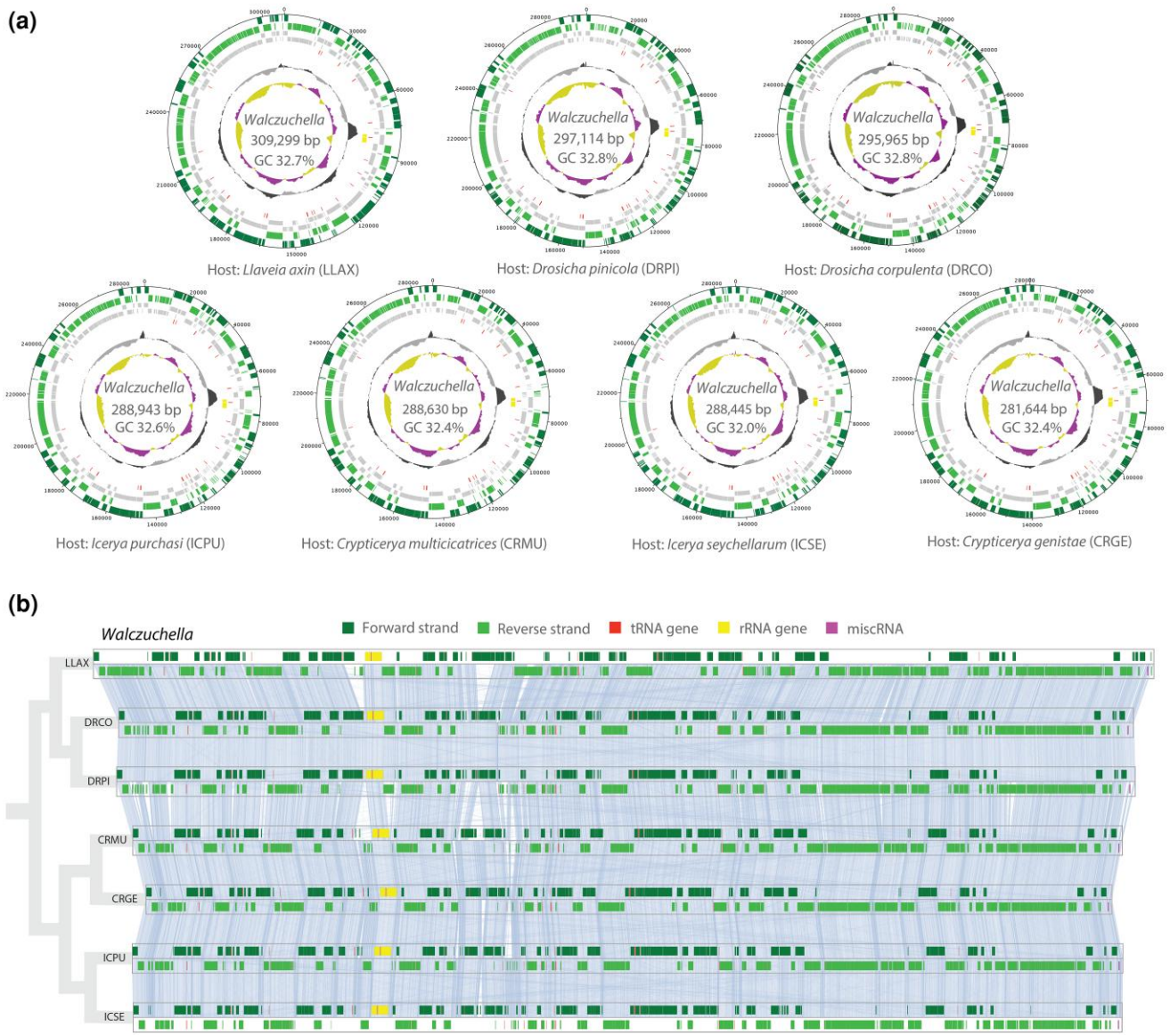

Fig. 1. Genome assemblies of *Walczuchella*. a) Circular genome maps of *Walczuchella*. The lines from outside to inside represent: (i) CDSs on the forward strand, (ii) CDSs on the reverse strand, (iii) genes on the forward strand, (iv) genes on the reverse strand, (v) tRNA genes, (vi) rRNA genes, (vii) GC plot (above-average shown as outward peaks and below-average as inward peaks) and (viii) GC skew (above-average shown as outward peaks and below-average as inward peaks). Genome sizes and GC% contents are listed in the center of circles. b) Linear genome alignments of *Walczuchella* with lines connecting colinear genes. *Walczuchella* genomes are collinear without distinct gene rearrangements. Alignments are ordered by the phylogenetic relationship of *Walczuchella* (supplementary fig. S7, Supplementary Material online). Color codes are provided on the linear genome alignments.

Table 1 Genome statistics of *Walczuchella* endosymbionts from seven giant scales

Symbiont	Genome size (bp)	Average read coverage	GC (%)	CDS	Pseudogenized CDS (truncated/fragmented)	CDS coding density (%)	rRNA	tRNA
Walczuchella CRMU	288,630	84	32.4	262	36 (20/16)	79	3	34
Walczuchella CRGE	281,644	12	32.4	263	27 (21/6)	81	3	34
Walczuchella DRCO	295,965	72	32.8	314	84 (59/25)	81	3	34
Walczuchella DRPI	297,114	1,226	32.8	322	83 (62/21)	81	3	34
Walczuchella ICPU	288,943	1,523	32.6	270	32 (20/12)	81	3	35
Walczuchella ICSE	288,445	2,939	32.0	282	42 (29/13)	80	3	34
Walczuchella LLAX	309,299	-	32.7	298	35 (30/5)	86	3	34

Material online). The genomes encode 262 to 322 CDSs, with 11 to 25 hypothetical proteins. Among the CDSs, 27 to 84 were predicted to be pseudogenes due to truncation (71%) or fragmentation (29%). The ends of the five pseudogenes were positioned within the potential misassembly regions, indicating uneven read depth on the coverage plots or being flagged as "low coverage," "ambiguous," and/or "deletion" in the variant calling using Pilon (supplementary table S2, Supplementary Material online). Each *Walczuchella* genome has 1 to 3 pseudogenes (*carB*, *trpC*, and *mnmE*) with long

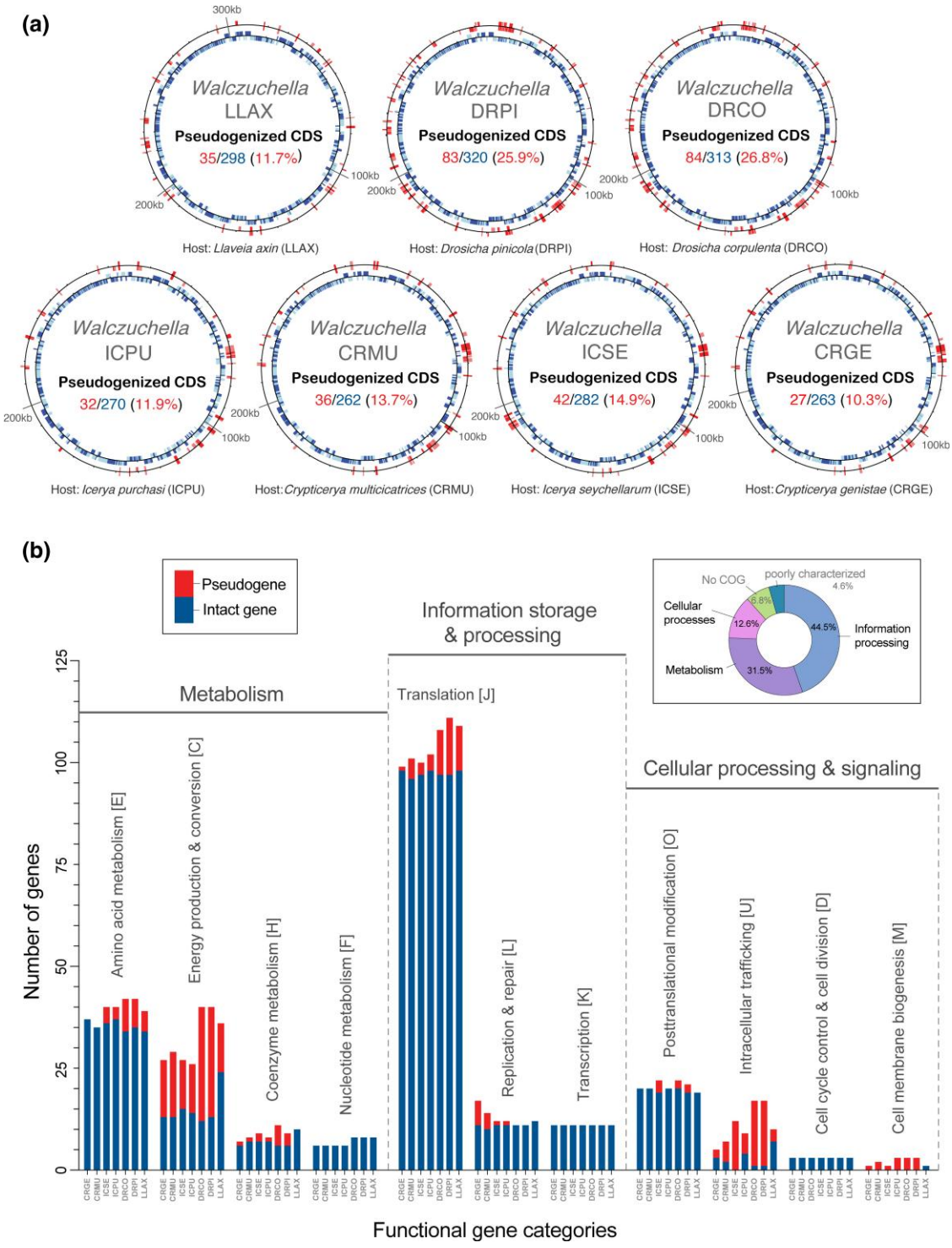

Fig. 2. Pseudogenization of *Walczuchella* genomes. a) Genome maps of *Walczuchella* genomes visualized with pseudogenes and intact genes on the outer and the inner circles, respectively. b) Protein-coding genes of *Walczuchella* genomes classified into the COG categories. Intact genes and pseudogenes are depicted in the bar plots. The average COG proportion of *Walczuchella* genomes is additionally included in the top right corner.

poly(A) or (T) tracts (>8 bp), potentially rescuable by transcriptional slippage. All seven *Walczuchella* genomes were completely syntenic (Fig. 1b). The largest proportion of genes in *Walczuchella* genomes was observed in the COG category of information storage and processing (44.5%), followed by metabolism (31.5%) and cellular processes and signaling (12.6%) (Fig. 2b).

Gene Loss and Pseudogenization of *Walczuchella*

The *Walczuchella* genomes showed a high frequency of gene loss and pseudogenization in the following functional COG categories: the biosynthesis of the cellular membrane (M), intracellular trafficking, secretion, and vesicular transport (U), and energy production and conversion (C) (Fig. 2; supplementary table S3, Supplementary Material online).

In the category (M), the genes *lgt* and *rseP* were only found, but they are pseudogenized or missing in most *Walczuchella* genomes. For the category (U), the majority of 11 genes detected are either lost (in *Walczuchella* DRCO, DRPI, and ICSE) or pseudogenized in the remaining species. In the category (C), out of 38 genes detected in *Walczuchella* genomes, most showed a high frequency of gene loss and pseudogenization, with only a few genes remaining intact. The other COGs showed lower degradation, but we note stochastic losses and pseudogenization of essential genes involved in central informational processes, such as DNA replication and repair, transcription, and translation. Degradation of genes encoding the DNA mismatch repair protein MutS, DNA topoisomerase I, DNA helicase, and endodeoxyribonuclease was observed in the DNA replication and repair category. Missing transcription regulator-encoding genes (*norR* and *glnG*) were also noted in the transcription category. Interestingly, some genes encoding ribosomal proteins of the large subunit were lost or pseudogenized, while the genes encoding the small subunit ribosomal proteins remained intact. Some losses of aminoacyl tRNA synthetases were also observed. In the translation category, more than one genome of *Walczuchella* showed gene loss or pseudogenization in genes related to tRNA biosynthesis, ribonuclease, elongation factor 4, peptide chain release factor 1, and translation initiation factor IF-2. The translation initiation factor IF-1 gene sequence was found as a gene fusion downstream of the CDS annotated as *secY* in *Walczuchella* CRMU and CRGE. Some losses/pseudogenes were also observed in genes responsible for amino acid transport and metabolism.

Substitution Rates and Selection Pressure on Protein-Coding Genes of *Walczuchella*

The substitution rates of 181 single-copy orthologous genes were estimated to average $dN = 0.0429$ and $dS = 0.3477$ (supplementary fig. S2, supplementary table S3, Supplementary Material online). The energy production-related gene *pfkA* had the highest dN (0.225) and the second highest dS (1.0945), while the 30S ribosomal protein gene *rpsU* had the highest dS (1.2213). The results from BUSTED and CODEML showed different averages and ranges for the ω values (supplementary fig. S3, supplementary table S3, Supplementary Material online). The average ω values from BUSTED (0.1670 and 0.1521) were higher than the average from CODEML (0.1340). However, they were congruent in that about 70% of genes ($n = 121$) showed $\omega > 0.1$, indicating that they were under relaxed purifying selection. The information storage and processing category had the highest number of genes under relaxed purifying selection with 50 genes, followed by the metabolism category with 34 genes, and the cellular processing and signaling category with 12 genes (Fig. 3a; supplementary table S3, Supplementary Material online). Among translation-related genes, six showed relatively high ω values (0.245 to 0.396; hereafter ω from CODEML), encoding the 50S and 30S ribosomal proteins, a ribosome-recycling factor, and ribonuclease P protein component. In the DNA replication and repair category, the gene *yqeN* had the highest ω value (0.244). In the metabolism category, the genes encoding histidine biosynthesis protein (*hisI*) and anthranilate phosphoribosyl transferase (*trpD*) showed relatively high ω values (0.504 and 0.288, respectively). In the posttranslational modification, tRNA biosynthesis protein TsaB had the highest ω value (0.363). Of the 181 single-copy orthologous genes, 63 may have experienced diversifying selection (Fig. 3b; supplementary table S3, Supplementary Material online). Both analyses of CODEML supported the positive selection of five genes at a *P*-value of 0.01 and 37 genes at a *P*-value of 0.05. In the results of BUSTED, the evidence for the positive selection was significant for six genes at a *P*-value of 0.01 and 9 genes at a *P*-value of 0.05. In all the analyses using CODEML and BUSTED, *hisI*, *korB*, *gltX*, and *rpsU* were found to be strongly supported as unconstrained genes.

DNA Replication and Repair Genes of *Walczuchella* and Other Ancient Symbionts

The genomes of ancient symbionts were found to contain 33 genes related to DNA replication and repair (Fig. 4; supplementary table S4, Supplementary Material online). DNA polymerase-related genes are relatively conserved across small genomes of all ancient symbionts. *Walczuchella* retains most polymerase genes, but the genes *dnaB* and *dnaG* were absent in all seven genomes. *dnaB* is still present in *Candidatus* Nasuia and *Candidatus* Tremblaya, and other ancient symbionts, except *Candidatus* Carsonella and some *Karelsulcia*. Among the repair genes, only *mutS* encoding DNA mismatch repair protein is present in three genomes of *Walczuchella*. Genetic properties of *dnaEQ* and *dnaN* genes indicated that mutation accumulation and deletions had occurred in their amino acid sequences, although overall protein structures were found to be similar to those of free-living bacteria (supplementary figs. S4 and S5, Supplementary Material online).

Phylogenetic Analyses of *Walczuchella* and Their Host Insects

In the phylogenetic tree based on the 16S and 23S rRNA gene sequences, *Walczuchella* formed a distinct clade comprised of two subclades with *Walczuchella* from the host genera *Drosicha* + *Llaveia* and *Crypticerya* + *Icerya* (Fig. 5a, supplementary fig. S6, Supplementary Material online). This topology was confirmed by the phylogenomic analysis based on 134 single-copy genes (supplementary fig. S7a, Supplementary Material online). Additionally, the phylogenenomic tree of host insects, reconstructed using 2,097 USCOs with high bootstrap values, was found to be congruent with the phylogenomic tree of *Walczuchella* (Fig. 5b). The clade of *Walczuchella* was found to be sister to *Candidatus* Hoataupuhia, an obligate symbiont of the Coelostomidiidae (supplementary fig. S6, Supplementary Material online). Flavobacterial symbionts of *Cryptococcus* scale insects and *Geopemphigus* aphids were sister to the clade of *Hoataupuhia* + *Walczuchella*.

Phylogenetic and Genome Analysis of co-symbionts

The metagenomes of giant scales revealed the presence of co-symbionts from multiple bacterial lineages, including Alphaproteobacteria and Gammaproteobacteria (Fig. 5a, supplementary fig. S6, Supplementary Material online). Acetobacteraceae-related symbionts were recognized in *Icerya seychellarum* and were phylogenetically close to *Bombella*. *Wolbachia* was detected in *Drosicha pinicola* and *Icerya purchasi*. *Arsenophonus* was found in *Crypticerya multicicatrices*. A long-branched *Sodalis*-like symbiont was detected in *Icerya seychellarum*. An *Enterobacter*-related symbiont was identified in *Icerya purchasi*. Two species of *Drosicha* showed *Cedecea*-related symbionts. *Serratia* was found in *Crypticerya genistae*, although its contigs had low

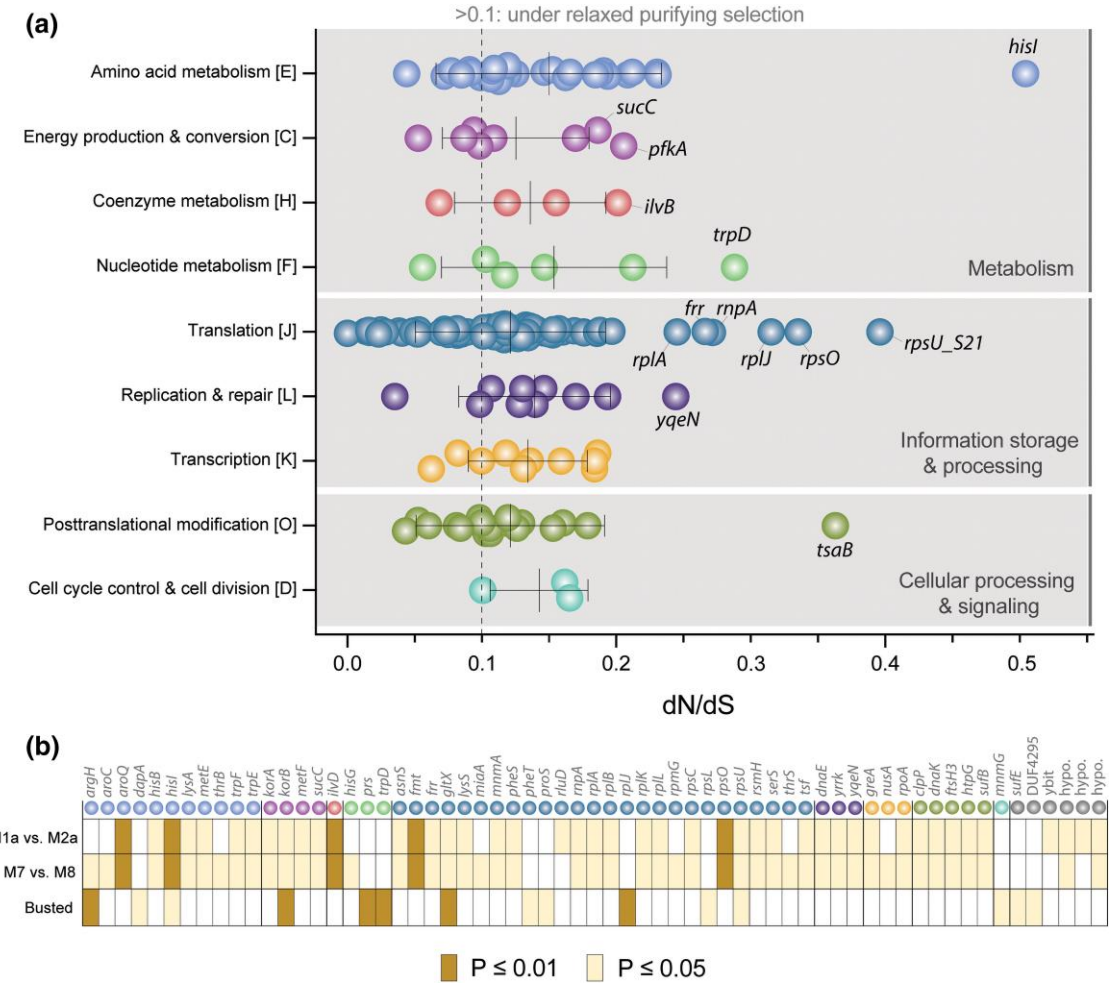

Fig. 3. Selective pressures on protein-coding genes of *Walczuchella*. a) Ratios of nonsynonymous to synonymous substitutions (dN/dS; ω) of single-copy orthologous genes of *Walczuchella*, sorted into the COG categories. Each circle shows ω of the gene estimated using Codeml based on the sequences of seven *Walczuchella*. The names of genes are given around circles showing higher ω in each category. The genes with $\omega > 0.1$ are considered under relaxed purifying selection. b) 63 genes predicted to have experienced diversifying selection. The first and second rows represent the two results from different approaches (M1a vs. M2a and M7 vs. M8) with Codeml. The third row shows the result of Busted. Each box of genes is filled with different colors following their statistical significance.

sequencing coverage. Several species of Coelostomidiidae showed *Sodalis*, *Wolbachia*, and *Erwinia*-related symbionts, some of which clustered close to the Enterobacterales symbionts from giant scales. The available genome assemblies of co-symbionts varied from 1.2 to 4.8 Mbp and had 34% to 57% GC content (supplementary table S5, Supplementary Material online).

Localization of *Walczuchella* in Giant Scales

Walczuchella localization and vertical transmission were confirmed in *Crypticerya multicicatrices* and *Icerya purchasi* using FISH and microscopy (Fig. 6). *Walczuchella* signal was localized in the posterior terminal area of the early stage of eggs (Fig. 6a). Its population was divided into two groups in the more developed egg (Fig. 6b) and located on both sides of the posterior area, eventually shaping bacteriomes during embryogenesis (Fig. 6c). The first instar nymph showed clear localization of *Walczuchella* in about five paired and multilobed structures in the abdomen (Fig. 6d). The localization of *Walczuchella* in the sectioned tissues was found to be within the bacteriocytes, which had larger polyploid nuclei compared with the normal host cells (Fig. 6e to h).

Discussion

Walczuchella is an Ancient Symbiont With a Large Number of Pseudogenes

Our results support initial reports from the first sequenced *Walczuchella* genome indicating the ongoing genome erosion in *Walczuchella* (Rosas-Pérez et al. 2014). We show that all *Walczuchella* genomes from different host species contain a high proportion of pseudogenes, with up to 27% of total CDSs. This is in stark contrast to other ancient symbionts, which usually contain fewer than 5% pseudogenized CDS (supplementary table S6, Supplementary Material online). Furthermore, we show that many intact genes are facing potential deactivation.

Approximately 70% of single-copy genes in *Walczuchella* are under relaxed purifying selection ($\omega > 0.1$) (Fig. 3; supplementary table S3, Supplementary Material online). Despite its already highly reduced genome of <300 kbp and ~150 My of co-evolution with scale insects, *Walczuchella* seems to be experiencing ongoing and rapid genome erosion. Similar levels of genome erosion are, to our knowledge, often associated with a rapid shift of the host diet or co-symbiont acquisition. One example of such a change is *Tremblaya princeps* with coding density of 66% in a 138 kbp genome (Husnik and McCutcheon 2016).

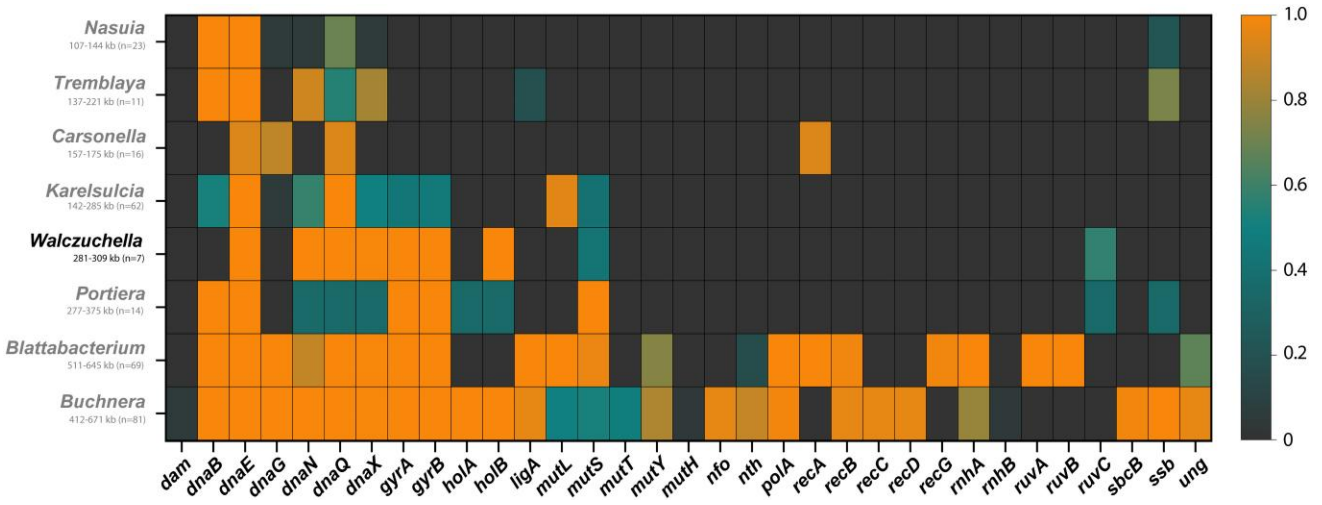

Fig. 4. Heatmap showing retention proportions of DNA replication and repair genes in genomes of ancient symbionts. The ranges of genome size and the number of analyzed genomes are noted under the names of symbionts.

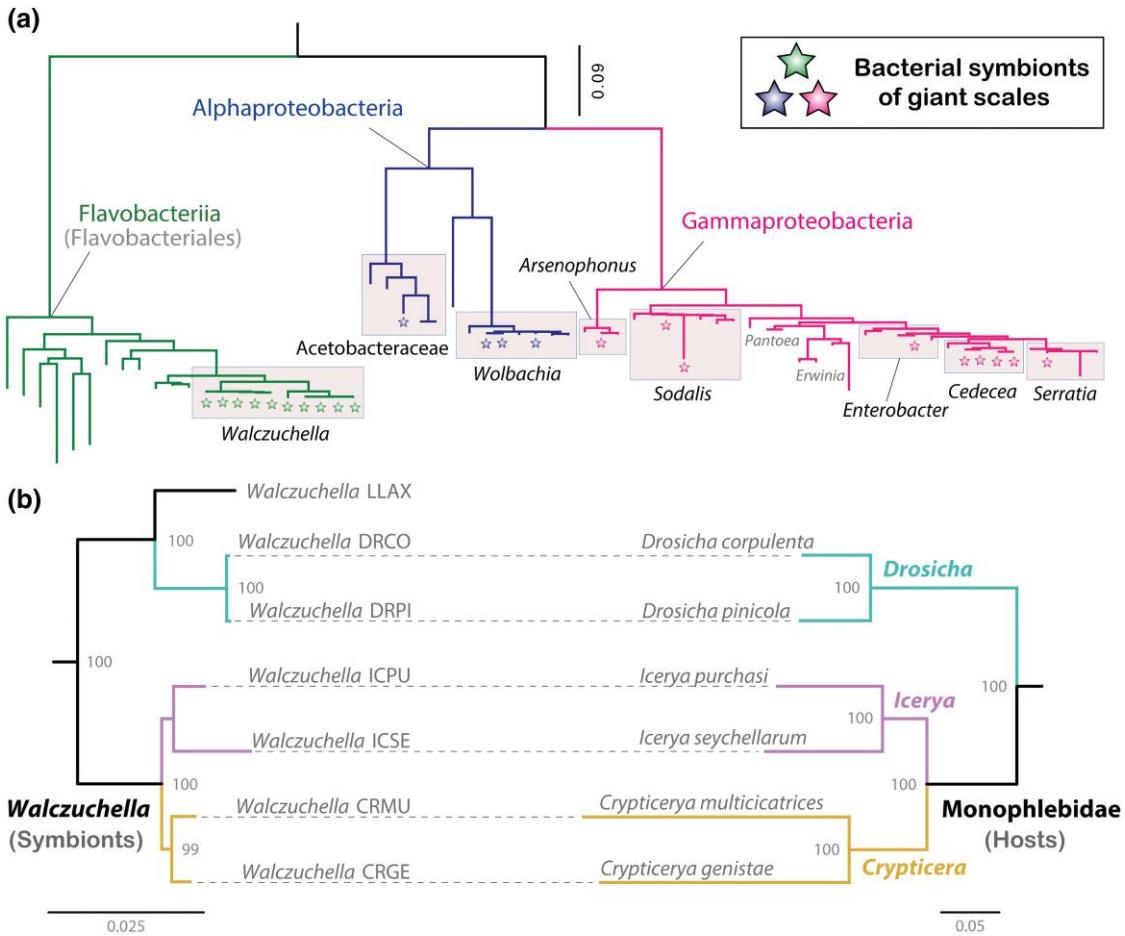

Fig. 5. Maximum likelihood analyses of giant scales and their symbionts. a) Phylogenetic tree of symbionts from giant scales based on 16S-23S rRNA sequences obtained from metagenome assemblies and NCBI. Symbionts of giant scales are indicated with stars. b) Co-phylogenetic analysis of *Walczuchella* and their hosts (detailed information provided in supplementary fig. S6, Supplementary Material online). *Walczuchella* phylogeny (left) was reconstructed based on 134 single-copy orthologous genes obtained from genome assemblies. The host phylogeny (right) was reconstructed based on 2,097 USCOs. The numbers at nodes represent bootstrap values over 95%. Symbionts are linked with dashed lines to their respective host insects. *Walczuchella* LLAX is not linked because of the lack of sequencing data of its host. Both trees were inferred with IQ-tree (see Materials and Methods for details).

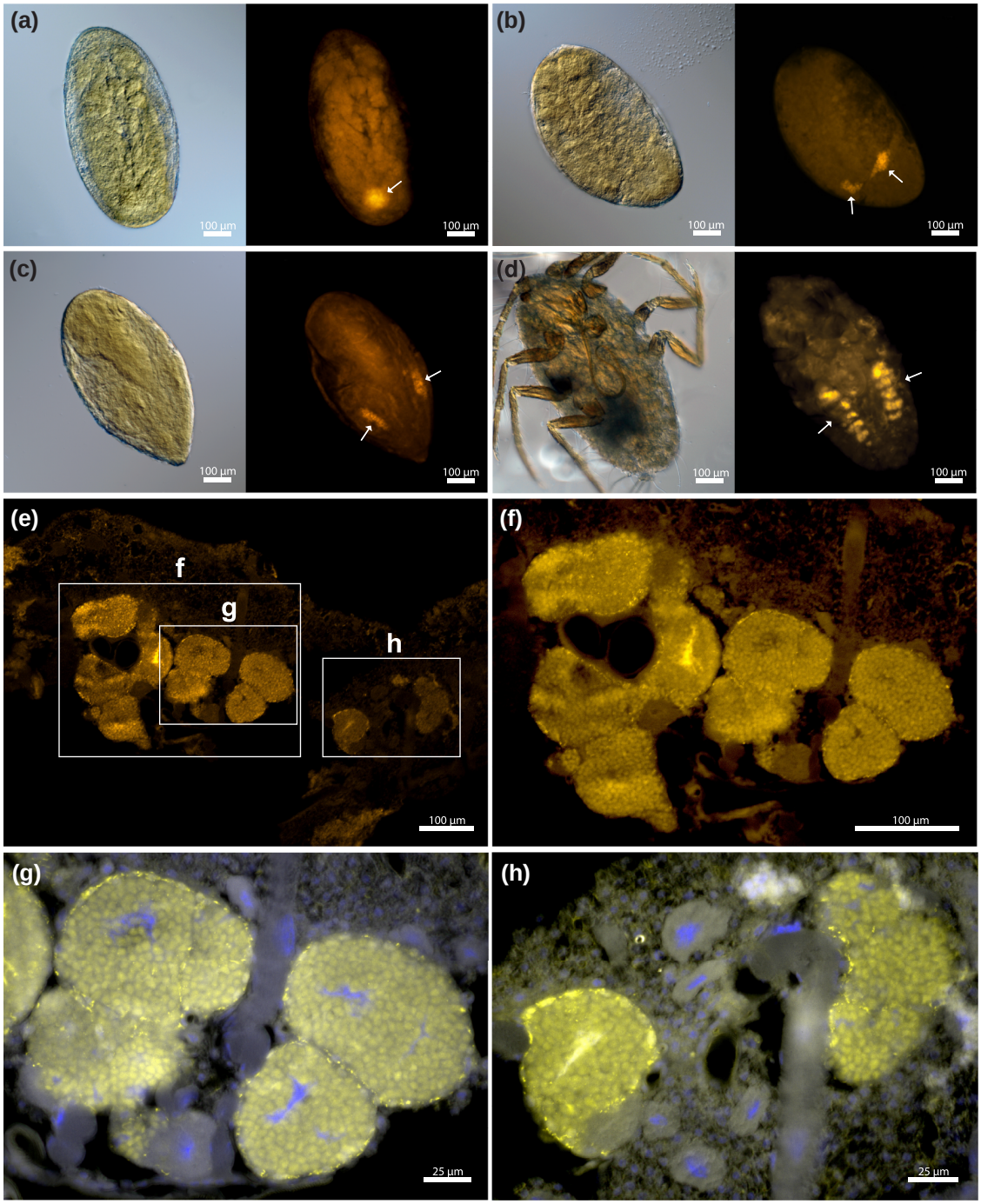

Fig. 6. Localization of *Walczuchella* in *Crypticerya multicicatrices* and *Icerya purchasi*. a) *Walczuchella* is localized in the posterior pore of the early stage of the egg of *I. purchasi*. b) *Walczuchella* is separated into two groups in the developing egg of *I. purchasi*. c) *Walczuchella* is localized on both sides of the posterior part of the egg of *I. purchasi*. d) Five paired and multilobed bacteriomes of the first instar nymph of *C. multicicatrices* including *Walczuchella*. a to d) Whole mounts. The photos on the left are light microscopy images, and the photos on the right are FISH images. The group of *Walczuchella* is indicated by an arrow. e) Entire view of sectioned tissue representing bacteriomes and bacteriocytes of *I. purchasi* in the nymphal stage. f to h) Higher magnification images of each part in e. e to h) FISH images of body sections. The bacteriomes (roundish structures) contain numerous *Walczuchella* (yellow dots). Nuclei are stained with DAPI (blue) in G and H.

This genome erosion was likely set in motion by the acquisition of intrabacterial symbionts and subsequent relaxed selection pressure on the genes of *Tremblaya*.

Which Functional Gene Categories are Eroding Away?

The pseudogenization of *Walczuchella* affects most gene categories, but a few are experiencing extensive erosion (Fig. 2). Significant deterioration was observed in genes associated with the cellular envelope, intracellular trafficking, and energy metabolism (Fig. 2b), which are functions typically lost in tiny endosymbiont genomes due to complementation by the host insect (Moran and Bennett 2014).

The genome size of *Walczuchella* between medium-sized and tiny symbionts coincides with losing the capability to synthesize peptidoglycan, phospholipids, and lipopolysaccharides,

and produce energy. Transmission electron microscopy of *Palaeococcus* species showed that *Walczuchella* is likely surrounded by a host-derived membrane (Szklarzewicz et al. 2006). Further high-resolution microscopy is required to clarify the detailed structure of its membranes. ATP is provided by the host mitochondria, similarly to other insect symbionts (Mao et al. 2018) and as suggested by the presence of numerous mitochondria in the bacteriocyte cytoplasm (Szklarzewicz et al. 2006) and pseudogenization of *Walczuchella* genes involved in energy production and conversion.

Where *Walczuchella* genomes seem to be more reduced when compared with symbionts with similar genome sizes is DNA replication, transcription, and translation (Fig. 2b). Especially notable are losses of select DNA repair and replication genes (Fig. 4). *Walczuchella* has lost DNA repair genes, although some genomes retain *mutS* encoding the DNA mismatch repair protein. While other ancient symbionts also experienced substantial losses in DNA repair genes, they usually retain the genes encoding DNA replication proteins. However, *Walczuchella* has lost two key genes, *dnaB* and *dnaG*, which encode the DNA helicase and primase. Among translation-related genes, significant decay was found in many aminoacyl-tRNA synthetases and 50S/30S ribosomal proteins. The glutamate-tRNA ligase and 30S ribosomal protein S21 genes were predicted to have undergone positive selection, and several other translation-related genes are under relaxed purifying selection with relatively high ω values (Fig. 3). Other key translation steps such as initiation, elongation, and peptide release (*infB*, *lepA*, and *prfA*) were also affected by the genome erosion. Strikingly, the translational initiation factor IF-1 and some ribosomal proteins genes are missing or present as pseudogenes in some *Walczuchella* genomes even though they are typically retained even in ancient symbionts with tiny genomes (Moran and Bennett 2014; Husnik and McCutcheon 2016).

We have confirmed the conservation of genes in the category of amino acid metabolism in the *Walczuchella* genomes (Fig. 2b). This supports its role as a provider of essential amino acids to the insect hosts, as previously reported (Rosas-Pérez et al. 2014). However, our findings indicate the loss or pseudogenization of genes encoding enzymes for synthesizing certain amino acids, such as arginine, lysine, methionine, and tryptophan. In addition, the histidine biosynthesis gene (*hisI*) was strongly implicated as having undergone positive selection (Fig. 3). Our analysis suggests that this gene may currently be under relaxed purifying selection, as it showed the highest ω among the genes tested in this study. Rosas-Pérez et al. (2014) noted that some substrates for the biosynthesis of several amino acids need to be supplied to complete the pathways due to the loss of several amino acid biosynthesis genes in the genome of *Walczuchella*. Our six new *Walczuchella* genomes also confirmed the absence of *argAE*, *aroEH*, *aspC*, *dapEF*, *hisCD*, *ilvE*, and *trpG*, while their respective co-symbionts have retained some or most of them (supplementary table S7, Supplementary Material online). Additionally, horizontally transferred genes of bacterial origin, such as *dapEF*, were found in the host genomes of giant scales (supplementary table S7, Supplementary Material online). These results suggest a complex metabolic patchwork among *Walczuchella,* its co-symbionts, and host genes of both native and bacterial origin, similar to other scale insects such as mealybugs (Husnik and McCutcheon 2016). Additionally, we confirm that all *Walczuchella* genomes retain *rpoN*, a regulator of genes involved in nitrogen metabolism and assimilation (Sabree et al. 2009; Rosas-Pérez et al. 2014).

Main Drivers of Genome Erosion in *Walczuchella*

Long-Term Intracellular Lifestyle and Vertical Transmission

Walczuchella is a vertically transmitted endosymbiotic bacterium. We confirmed the presence of *Walczuchella* in the eggs of *I. purchasi* and its localization during embryogenesis (Fig. 6a to c). The migration of *Walczuchella* to oocytes was previously observed in the ovarioles of the giant scales *Palaeococcus fuscipennis* (Szklarzewicz et al. 2006). In the first instar nymph of *C. multicicatrices*, *Walczuchella* is present in the paired multilobed symbiotic organ (Fig. 6d), which is similar to the symbiotic organs housing the flavobacterial symbionts of Coelostomidiidae (Dhami et al. 2012, 2013). Phylogenetic analyses reveal that Coelostomidiidae and Monophlebidae, including their symbiotic bacteria, are sister clades (supplementary fig. S6, Supplementary Material online; Dhami et al. 2012; Rosas-Pérez et al. 2014; Vea and Grimaldi 2016). Furthermore, the symbiotic bacteria and their hosts in Coelostomidiidae and Monophlebidae show co-phylogenetic patterns (Fig. 5b; Dhami et al. 2013). These results suggest that they share the most recent common ancestor that originated approximately 125 to 150 million years ago (Vea and Grimaldi 2016) and their symbionts have coevolved with these two host families since then (Fig. 7b). Similar to many other endosymbionts, the strict intracellular lifestyle and limited effective population size of *Walczuchella* result in stronger effects of random genetic drift and less effective purifying selection over time (McCutcheon and Moran 2012). We did not identify any clear differences in these processes between *Walczuchella* and other insect endosymbionts, but we cannot rule out more profound bottlenecking, for example, due to effective population sizes of giant scales or the number of symbiont cells transmitted from the mother to offspring, which are currently unknown.

Establishment of Obligate Co-Symbionts

The establishment of co-obligate endosymbionts can be an important factor contributing to the ongoing genome reduction of *Walczuchella* in giant scales. We show that various bacterial lineages of Alphaproteobacteria and Gammaproteobacteria are putative obligate co-symbionts of giant scales (Figs. 5a and 7b) based on the genome analyses, although their localizations in the hosts need to be confirmed. The composition of these co-symbionts varies depending on the species of giant scales, but *Walczuchella* is uniformly present in all samples. This finding suggests that giant scales have independently acquired diverse co-symbionts during their diversification. *Arsenophonus* and *Sodalis*- allied bacteria are known to complement the missing nutritional genes of ancient symbiotic partners in aphids, mealybugs, and planthoppers (Husnik and McCutcheon 2016; Michalik et al. 2021, 2023; Manzano-Marín et al. 2023). The genome of *Sodalis* TME1 in *Llaveia axin* contains genes involved in all pathways for biosynthesis of essential amino acids, and this *Sodalis* is thus potentially supplying nutrients to the host in cooperation with *Walczuchella* (Rosas-Perez et al. 2017). Potential metabolic complementation by co-symbionts thus likely occurs in giant scales for amino acid biosynthesis (supplementary table S7, Supplementary Material online). Furthermore, *Cedecea*-related co-symbionts were previously confirmed in two species of *Drosicha* where they are intracellular in the host cells surrounded by *Walczuchella* bacteriocytes (Fig. 7a; Matsuura et al. 2009). The acquisition of additional

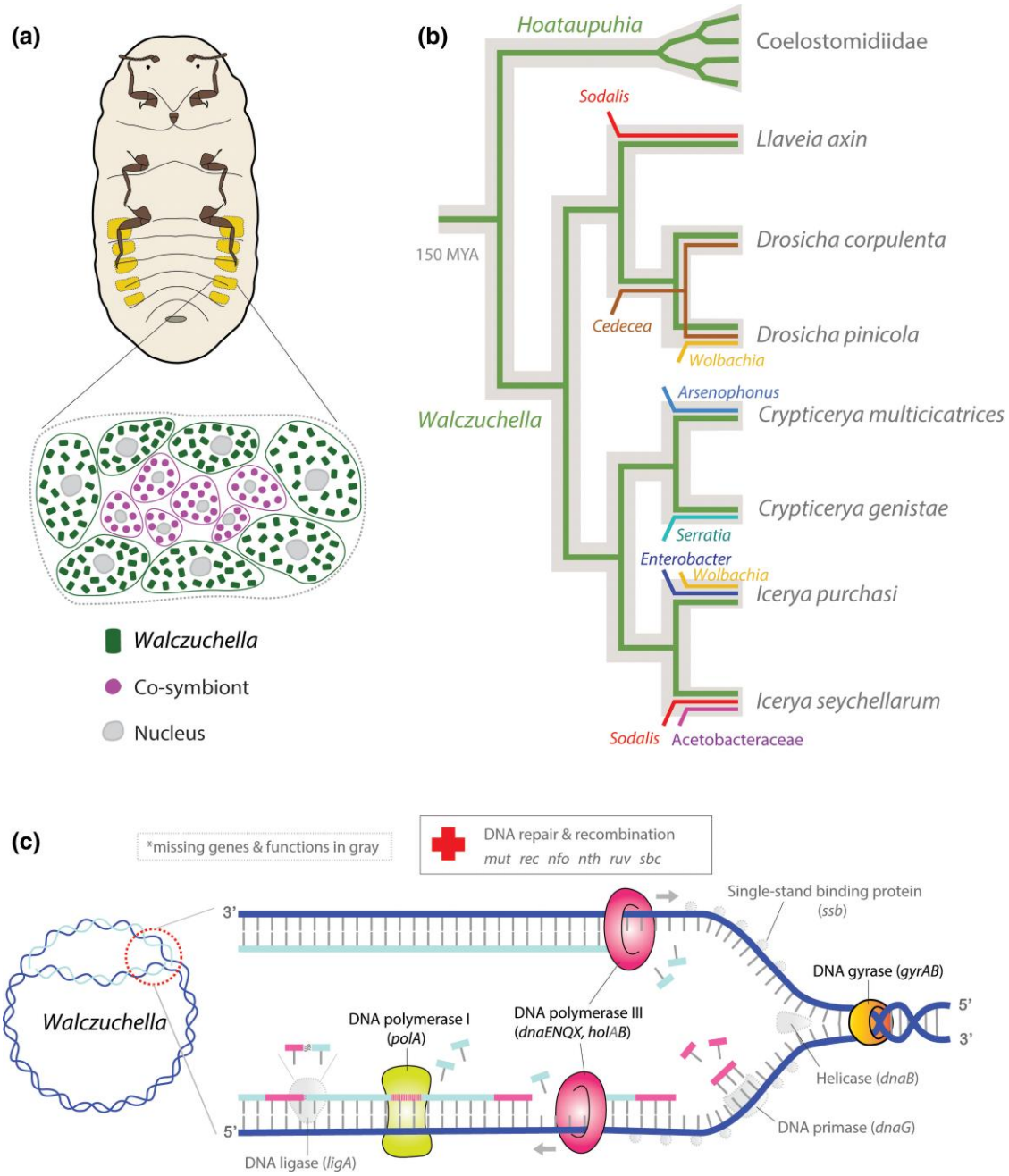

Fig. 7. Potential microbial symbiotic features of giant scales leading to the degradation of *Walczuchella* genomes. a) Intracellular lifestyle of *Walczuchella* in bacteriocytes of giant scales with co-symbionts. This illustration is based on the FISH images for *Drosicha* host species and their symbionts reported previously (Matsuura et al. 2009). b) Long-term evolutionary relationship of *Walczuchella* with hosts and additional introductions of putative co-symbionts. Colored lines with labels indicate symbionts on host phylogeny (gray background). c) Incomplete DNA replication and repair system of *Walczuchella*.

symbionts can be detrimental to *Walczuchella* as more functional "junior" symbionts with a larger gene pool may lead to further relaxation of purifying selection on redundant genes in *Walczuchella*. Co-symbiont acquisition is thus emerging as a major factor driving significant changes in the co-occurring genomes of ancient symbionts that are otherwise relatively stable.

Loss of DNA Replication and Repair Genes

The breakdown of the DNA replication and repair system in *Walczuchella* could be a significant factor contributing to their excessive pseudogenization. Our analyses have revealed a limited set of DNA replication and repair genes (Figs. 4, 7c). Remarkably, *Walczuchella* has lost two essential genes, *dnaB* and *dnaG*, that are important for the initiation of replication. In comparison, other ancient symbionts, including *Nasuia* and *Tremblaya* with extremely reduced genomes, tend to retain at least *dnaB*. Although there are some cases of *dnaB* loss among *Carsonella* and *Karelsulcia*, their genome size is 157 to 245 kbp, which is much smaller than that of *Walczuchella* (281 to 309 kbp) (supplementary table S4, Supplementary Material online). *Karelsulcia* and *Portiera*, which have similar genome sizes to *Walczuchella*, retain *dnaB*. Mutations of *dnaB* were shown to increase genome instability and result in higher mutational frequency in *E. coli*

(Behrmann et al. 2021). We hypothesize that this early loss of *dnaB* in *Walczuchella* genomes was the tipping point of genome degradation. The remaining DNA polymerase genes have accumulated many mutations likely partly affecting their function (supplementary figs. S4 and S5, Supplementary Material online). Furthermore, *Walczuchella* does not contain DNA recombination and repair proteins found in symbionts with similarly reduced genomes. The incomplete system of DNA replication and repair leads to an increased rate of mutations that can easily accumulate over time. The enhanced mutation rate is strongly correlated with the evolution of genome size in prokaryotes (Bourguignon et al. 2020). The loss of key genes for DNA replication and repair introduces accidental early stop codons, consequently resulting in the more rapid degradation of *Walczuchella* genomes with increased pseudogenization.

Continuous Genome Instability of Ancient Symbionts

Microbial symbionts are subject to environmental conditions that inevitably result in genome reduction, integration with the host, or extinction (Husnik and Keeling 2019). The complete genomes of ancient symbionts reveal a wide range of genome sizes, indicating ongoing genome reduction (supplementary fig. S8, Supplementary Material online). Among ancient symbionts of insects, *Buchnera* is a striking example of this phenomenon, with genome sizes ranging from 412 to 671 kbp. *Walczuchella* can be considered an ancient symbiont based on its long-term symbiotic relationship with Monophlebidae and its genome size similar to *Karelsulcia*, which originated at least 260 million years ago [14]. Despite its status as an ancient symbiont, *Walczuchella* continues to show unusually high levels of genome degradation. The acceleration is likely due to the early loss of DNA replication/repair genes and the frequent establishment of functional co-symbionts, although we cannot rule out alternative reasons such as unusually strong bottlenecking during vertical transmission, accelerated cellular integration with the host cell (organellogenesis-like), or differences in the host biology such as androdioecy affecting its population-genetic processes (Fig. 7). The different rates of genome degradation among symbionts may be due to various factors, such as the mode of transmission, effective population sizes of both hosts and symbionts, differences in host biology, and the genetic bases of mutualistic features (Rispe and Moran 2000). In particular, the number of symbionts transmitted to offspring can be modulated by very specific circumstances, as exemplified by *Hodgkinia* in cicadas (Campbell et al. 2018). Therefore, further work is needed to determine if *Walczuchella* has a lower rate of transmission to offspring compared with other ancient symbionts, as reducing the number of symbionts can promote the fixation of mutations (Rispe and Moran 2000). *Walczuchella* joins a few other endosymbionts such as *Tremblaya princeps* in being the exception to the typical pattern of ancient symbionts retaining few pseudogenes (McCutcheon and Moran 2012). It is a cautionary tale that even if most symbiont genomes may reach a relatively stable state, some such as *Walczuchella* are currently spiraling down the symbiosis rabbit hole of genome instability and pseudogenization.

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

Acknowledgments

J.C. was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT; 2021R1A6A3A03038909) and the JSPS KAKENHI grant (20939772). F.H. was supported by the JSPS KAKENHI grant (23K14256) and the HFSP Early Career Grant (RGEC29/2024). T.K. was partially supported by the Corporación Colombiana de Investigación Agropecuaria—Agrosavia. Specimens from Colombia (CRMU and CRGE) were collected under a permit framework for collecting specimens of wild species of the biological diversity for non-commercial scientific research purposes (resolution No. 1466, Autoridad Nacional de Licencias Ambientales—ANLA) [Colombian National Authority Environmental Permits]. We thank Nic Schröder for assistance with a phylogenetic analysis of host insects, Akito Shima for creating a script for estimating sequence read coverage, and Dr. Yamagishi Kenzo (Meijo University, Nagoya, Japan) for providing us with samples of *Drosicha corpulenta* Kuwana. We also acknowledge the Scientific Computing (SCDA), Imaging (IMG), and Sequencing (SQC) sections of the Okinawa Institute of Science and Technology for their great support.

Author Contributions

J.C. led the study, conducted the data analyses, prepared all figures, and drafted the manuscript. F.H. designed the study, supervised the analyses, and revised the manuscript. P.P. performed DNA extraction and data analysis. H.T., T.K., and M.E.G. provided samples and metagenome data. All authors edited the manuscript.

Data Availability

The genome assemblies of *Walczuchella* and Illumina raw reads are available under NCBI BioProject PRJNA1129605.

References

Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010 [accessed 2022 Nov 30]. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Behrmann MS, Perera HM, Hoang JM, Venkat TA, Visser BJ, Bates D, Trakselis MA. Targeted chromosomal *Escherichia coli*:*dnaB* exterior surface residues regulate DNA helicase behavior to maintain genomic stability and organismal fitness. *PLoS Genet*. 2021:17(11):e1009886. https://doi.org/10.1371/journal.pgen.1009886.

Bennett GM, Moran NA. Small, smaller, smallest: the origins and evolution of ancient dual symbioses in a phloem-feeding insect. *Genome Biol Evol*. 2013:5(9):1675–1688. https://doi.org/10.1093/gbe/evt118.

Bourguignon T, Kinjo Y, Villa-Martín P, Coleman NV, Tang Q, Arab DA, Wang Z, Tokuda G, Hongoh Y, Ohkuma M, *et al*. Increased mutation rate is linked to genome reduction in prokaryotes. *Curr Biol*. 2020:30(19):3848–3855.e4. https://doi.org/10.1016/j.cub.2020.07.034.

Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods*. 2021:18(4):366–368. https://doi.org/10.1038/s41592-021-01101-x.

Campbell MA, Łukasik P, Meyer MC, Buckner M, Simon C, Veloso C, Michalik A, McCutcheon JP. Changes in endosymbiont complexity drive host-level compensatory adaptations in cicadas. *mBio*. 2018:9(6):e02104-18. 10.1128/mbio.02104-18.

Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. Eggnog-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol*

Evol. 2021:38(12):5825–5829. https://doi.org/10.1093/molbev/msab293.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. Trimal: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009:25(15):1972–1973. https://doi.org/10.1093/bioinformatics/btp348.

Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics*. 2009:25(1):119–120. https://doi.org/10.1093/bioinformatics/btn578.

Chan PP, Lin BY, Mak AJ, Lowe TM. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res*. 2021:49(16):9077–9096. https://doi.org/10.1093/nar/gkab688.

Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018:34(17):i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

Choi J, Lee S. Higher classification of mealybugs (Hemiptera: Coccomorpha) inferred from molecular phylogeny and their endosymbionts. *Syst Entomol*. 2022:47(2):354–370. https://doi.org/10.1111/syen.12534.

Chong RA, Moran NA. Evolutionary loss and replacement of *Buchnera*, the obligate endosymbiont of aphids. *ISME J*. 2018:12(3):898–908. https://doi.org/10.1038/s41396-017-0024-6.

Dhami MK, Buckley TR, Beggs JR, Taylor MW. Primary symbiont of the ancient scale insect family Coelostomidiidae exhibits strict cophylogenetic patterns. *Symbiosis*. 2013:61(2):77–91. https://doi.org/10.1007/s13199-013-0257-8.

Dhami MK, Turner AP, Deines P, Beggs JR, Taylor MW. Ultrastructural and molecular characterization of a bacterial symbiosis in the ecologically important scale insect family Coelostomidiidae. *FEMS Microbiol Ecol*. 2012:81(3):537–546. https://doi.org/10.1111/j.1574-6941.2012.01378.x.

Dial DT, Weglarz KM, Aremu AO, Havill NP, Pearson TA, Burke GR, Von Dohlen CD. Transitional genomes and nutritional role reversals identified for dual symbionts of adelgids (Aphidoidea: Adelgidae). *ISME J*. 2022:16(3):642–654. https://doi.org/10.1038/s41396-021-01102-w.

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004:32(5):1792–1797. https://doi.org/10.1093/nar/gkh340.

Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019:20(1):238. https://doi.org/10.1186/s13059-019-1832-y.

Foley G, Mora A, Ross CM, Bottoms S, Sützl L, Lamprecht ML, Zaugg J, Essebier A, Balderson B, Newell R, *et al.* Engineering indel and substitution variants of diverse and ancient enzymes using Graphical Representation of Ancestral Sequence Predictions (GRASP). *PLOS Comput Biol*. 2022:18(10):e1010633. https://doi.org/10.1371/journal.pcbi.1010633.

Garber AI, Kupper M, Laetsch DR, Weldon SR, Ladinsky MS, Bjorkman PJ, McCutcheon JP. The evolution of interdependence in a four-way mealybug symbiosis. *Genome Biol Evolu*. 2021:13(8):evab123. https://doi.org/10.1093/gbe/evab123.

García Morales M, Denno BD, Miller DR, Miller GL, Ben-Dov Y, Hardy NB. ScaleNet: a literature-based model of scale insect biology and systematics. *Database (Oxford)*. 2016:2016:bav118. https://doi.org/10.1093/database/bav118.

Grafton-Cardwell EE, Gu P, Daugherty MP. Impact of *Icerya purchasi* (Hempitera: Monophlebidae) on Navel Orange, *Citrus sinensis*, production and fruit quality. *J Econ Entomol*. 2022:115(5):1627–1636. https://doi.org/10.1093/jee/toac129.

Gruwell ME, Hardy NB, Gullan PJ, Dittmar K. Evolutionary relationships among primary endosymbionts of the mealybug subfamily phenacoccinae (Hemiptera: Coccoidea: Pseudococcidae). *Appl Environ Microbiol*. 2010:76(22):7521–7525. https://doi.org/10.1128/AEM.01354-10.

Gruwell ME, Morse GE, Normark BB. Phylogenetic congruence of armored scale insects (Hemiptera: Diaspididae) and their primary endosymbionts from the phylum Bacteroidetes. *Mol Phylogenet Evol*. 2007:44(1):267–280. https://doi.org/10.1016/j.ympev.2007.01.014.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Bio Evol*. 2018:35(2):518–522. https://doi.org/10.1093/molbev/msx281.

Huerta-Cepas J, Serra F, Bork P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol Biol Evol*. 2016:33(6):1635–1638. https://doi.org/10.1093/molbev/msw046.

Husnik F, Keeling PJ. The fate of obligate endosymbionts: reduction, integration, or extinction. *Curr Opin Genet Devel*. 2019:58-59:1–8. https://doi.org/10.1016/j.gde.2019.07.014.

Husnik F, McCutcheon JP. Repeated replacement of an intrabacterial symbiont in the tripartite nested mealybug symbiosis. *Proc Natl Acad Sci U S A*. 2016:113(37):E5416–E5424. https://doi.org/10.1073/pnas.1603910113.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017:14(6):587–589. https://doi.org/10.1038/nmeth.4285.

Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: improvements in performance and usability. *Mol Biol Evol*. 2013:30(4):772–780. https://doi.org/10.1093/molbev/mst010.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, *et al.* Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012:28(12):1647–1649. https://doi.org/10.1093/bioinformatics/bts199.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*. 2015:10(6):845–858. https://doi.org/10.1038/nprot.2015.053.

Kiefer JST, Bauer E, Okude G, Fukatsu T, Kaltenpoth M, Engl T. Cuticle supplementation and nitrogen recycling by a dual bacterial symbiosis in a family of xylophagous beetles. *ISME J*. 2023:17(7):1029–1039. https://doi.org/10.1038/s41396-023-01415-y.

Koga R, Tsuchida T, Fukatsu T. Quenching autofluorescence of insect tissues for *in situ* detection of endosymbionts. *Appl Entomol Zool*. 2009:44(2):281–291. https://doi.org/10.1303/aez.2009.281.

Laetsch DR, Blaxter ML. BlobTools: interrogation of genome assemblies. *F1000Research*. 2017:6(1287). https://doi.org/10.12688/f1000research.12232.1.

Li M, Tong H, Wang S, Ye W, Zicheng L, Omar MAA, Ao Y, Ding S, Zihao L, Wang Y, *et al.* A chromosome-level genome assembly provides new insights into paternal genome elimination in the cotton mealybug *Phenacoccus solenopsis*. *Mol Ecol Resour*. 2020:20(6):1733–1747. https://doi.org/10.1111/1755-0998.13232.

Li W-H, Gojobori T, Nei M. Pseudogenes as a paradigm of neutral evolution. *Nature*. 1981:292(5820):237–239. https://doi.org/10.1038/292237a0.

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol*. 2021:38(10):4647–4654. https://doi.org/10.1093/molbev/msab199.

Manzano-Marín A, Coeur d'acier A, Clamens A-L, Cruaud C, Barbe V, Jousselin E. Co-obligate symbioses have repeatedly evolved across aphids, but partner identity and nutritional contributions vary across lineages. *Peer Community J*. 2023:3:e46. https://doi.org/10.24072/pcjournal.278.

Mao M, Yang X, Bennett GM. Evolution of host support for two ancient bacterial symbionts with differentially degraded genomes in a leafhopper host. *Proc Natl Acad Sci U S A*. 2018:115(50):E11691–E11700. https://doi.org/10.1073/pnas.1811932115.

Matsuura Y, Koga R, Nikoh N, Meng X-Y, Hanada S, Fukatsu T. Huge symbiotic organs in giant scale insects of the genus *Drosicha* (Coccoidea: Monophlebidae) harbor flavobacterial and enterobacterial endosymbionts. *Zool Sci*. 2009:26(7):448–456. https://doi.org/10.2108/zsj.26.448.

McCutcheon JP, Moran NA. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol*. 2012:10(1):13–26. https://doi.org/10.1038/nrmicro2670.

Michalik A, Castillo Franco D, Kobiałka M, Szklarzewicz T, Stroiński A, Łukasik P. Alternative transmission patterns in independently acquired nutritional cosymbionts of dictyopharidae planthoppers. *mBio*. 2021:12(4):e0122821. https://doi.org/10.1128/mbio.01228-21.

Michalik A, Franco DC, Deng J, Szklarzewicz T, Stroiński A, Kobiałka M, Łukasik P. Variable organization of symbiont-containing tissue across planthoppers hosting different heritable endosymbionts. *Front Physiol*. 2023:14:1135346. https://doi.org/10.3389/fphys.2023.1135346.

Mongue AJ, Michaelides S, Coombe O, Tena A, Kim D, Normark BB, Gardner A, Hoddle MS, Ross L. Sex, males, and hermaphrodites in the scale insect *Icerya purchasi*. *Evolution*. 2021:75(11):2972–2983. https://doi.org/10.1111/evo.14233.

Moran NA. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc Natl Acad Sci U S A*. 1996:93(7):2873–2878. https://doi.org/10.1073/pnas.93.7.2873.

Moran NA, Bennett GM. The tiniest tiny genomes. *Annu Rev Microbiol*. 2014:68(1):195–215. https://doi.org/10.1146/annurev-micro-091213-112901.

Moran NA, Tran P, Gerardo NM. Symbiosis and insect diversification: an ancient symbiont of sap-feeding insects from the bacterial phylum *Bacteroidetes*. *Appl Environ Microbiol*. 2005:71(12):8802–8810. https://doi.org/10.1128/AEM.71.12.8802-8810.2005.

Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, Eren K, Pollner T, Martin DP, Smith DM, *et al*. Gene-wide identification of episodic selection. *Mol Biol Evol*. 2015:32(5):1365–1371. https://doi.org/10.1093/molbev/msv035.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015:32(1):268–274. https://doi.org/10.1093/molbev/msu300.

Petrov D, Hartl D. Pseudogene evolution and natural selection for a compact genome. *Heredity (Edinb)*. 2000:91(3):221–227. https://doi.org/10.1093/jhered/91.3.221.

Pond SLK, Frost SDW, Muse SV. Hyphy: hypothesis testing using phylogenies. *Bioinformatics*. 2005:21(5):676–679. https://doi.org/10.1093/bioinformatics/bti079.

Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes De Novo Assembler. *Curr Protocols Bioinf*. 2020:70(1):e102. https://doi.org/10.1002/cpbi.102.

Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res*. 2007:35(Database):D61–D65. https://doi.org/10.1093/nar/gkl842.

Rispe C, Moran NA. Accumulation of deleterious mutations in endosymbionts: Muller's ratchet with two levels of selection. *Am Nat*. 2000:156(4):425–441. https://doi.org/10.1086/303396.

Rosas-Pérez T, Rosenblueth M, Rincón-Rosales R, Mora J, Martínez-Romero E. Genome sequence of "*Candidatus* Walczuchella monophlebidarum" the Flavobacterial Endosymbiont of *Llaveia axin axin* (Hemiptera: Coccoidea: Monophlebidae). *Genome Biol Evol*. 2014:6(3):714–726. https://doi.org/10.1093/gbe/evu049.

Rosas-Perez T, Vera-Ponce De Len A, Rosenblueth M, Ramirez-Puebla ST, Rincon-Rosales R, Martinez-Romero J, Dunn MF, Kondorosi E, Martinez-Romero E. The symbiome of *Llaveia* Cochineals (Hemiptera: Coccoidea: Monophlebidae) includes a Gammaproteobacterial Cosymbiont *Sodalis* TME1 and the known *Candidatus* Walczuchella monophlebidarum. In: Shields VDC, editor. *Insect physiology and ecology*. Rijeka, Croatia: IntechOpen; 2017. p. 115–134.

Rosenblueth M, Sayavedra L, Sámano-Sánchez H, Roth A, Martínez-Romero E. Evolutionary relationships of flavobacterial and enterobacterial endosymbionts with their scale insect hosts (Hemiptera: Coccoidea). *J Evol Biol*. 2012:25(11):2357–2368. https://doi.org/10.1111/j.1420-9101.2012.02611.x.

Sabree ZL, Huang CY, Okusu A, Moran NA, Normark BB. The nutrient supplying capabilities of *Uzinura*, an endosymbiont of armoured scale insects: beneficial endosymbiont of armoured scale insects. *Environ Microbiol*. 2013:15(7):1988–1999. https://doi.org/10.1111/1462-2920.12058.

Sabree ZL, Kambhampati S, Moran NA. Nitrogen recycling and nutritional provisioning by *Blattabacterium*, the cockroach endosymbiont. *Proc Natl Acad Sci U S A*. 2009:106(46):19521–19526. https://doi.org/10.1073/pnas.0907504106.

Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014:30(14):2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

Smith SA, Dunn CW. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics*. 2008:24(5):715–716. https://doi.org/10.1093/bioinformatics/btm619.

Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014:30(9):1312–1313. https://doi.org/10.1093/bioinformatics/btu033.

Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res*. 2006:34(Web Server):W609–W612. https://doi.org/10.1093/nar/gkl315.

Syberg-Olsen MJ, Garber AI, Keeling PJ, McCutcheon JP, Husnik F. Pseudofinder: detection of pseudogenes in prokaryotic genomes. *Mol Biol Evol*. 2022:39(7):msac153. https://doi.org/10.1093/molbev/msac153.

Szklarzewicz T, Kalandyk-Kołodziejczyk M, Michalik A. Ovary structure and symbiotic associates of a ground mealybug, *Rhizoecus albidus* (Hemiptera, Coccomorpha: Rhizoecidae) and their phylogenetic implications. *J Anatomy*. 2022:241(3):860–872. https://doi.org/10.1111/joa.13712.

Szklarzewicz T, Kędra K, Niżnik S. Ultrastructure and transovarial transmission of endosymbiotic microorganisms in *Palaeococcus fuscipennis* (Burmeister) (Insecta, Hemiptera, Coccinea: Monophlebidae). *Folia Biol (Krakow)*. 2006:54(1):69–74. https://doi.org/10.3409/173491606777919102.

Szklarzewicz T, Michalik A, Michalik K. The diversity of symbiotic systems in scale insects. In: Kloc M, editor. *Symbiosis: cellular, molecular, medical and evolutionary aspects*. Cham: Springer International Publishing; 2020. p. 469–495.

Thalén F, Köhne CG, Bleidorn C. Patchwork: alignment-based retrieval and concatenation of phylogenetic markers from genomic data. *Genome Biol Evol*. 2023:15(12):evad227. https://doi.org/10.1093/gbe/evad227.

Vasquez YM, Bennett GM. A complex interplay of evolutionary forces continues to shape ancient co-occurring symbiont genomes. *iScience*. 2022:25(8):104786. https://doi.org/10.1016/j.isci.2022.104786.

Vea IM, Grimaldi DA. Putting scales into evolutionary time: the divergence of major scale insect lineages (Hemiptera) predates the radiation of modern angiosperm hosts. *Sci Rep*. 2016:6(1):23487. https://doi.org/10.1038/srep23487.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, *et al*. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*. 2014:9(11):e112963. https://doi.org/10.1371/journal.pone.0112963.

Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics*. 2015:31(20):3350–3352. https://doi.org/10.1093/bioinformatics/btv383.

Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007:24(8):1586–1591. https://doi.org/10.1093/molbev/msm088.