# Endosymbiont gene functions impaired and rescued by polymerase infidelity at poly(A) tracts

Ivica Tamas*[†][‡], Jennifer J. Wernegreen[†][§][¶], Björn Nystedt*, Seth N. Kauppinen[§], Alistair C. Darby*[∥], Laura Gomez-Valero*[,**], Daniel Lundin[††], Anthony M. Poole[¶][††], and Siv G. E. Andersson*[¶]

*Department of Molecular Evolution, Evolutionary Biology Centre, Uppsala University, SE-752 36 Uppsala, Sweden; [§]Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, MA 02543; and [††]Department of Molecular Biology and Functional Genomics, Stockholm University, SE-106 91 Stockholm, Sweden

Among host-dependent bacteria that have evolved by extreme reductive genome evolution, long-term bacterial endosymbionts of insects have the smallest (160–790 kb) and most A + T-rich (>70%) bacterial genomes known to date. These genomes are riddled with poly(A) tracts, and 5–50% of genes contain tracts of 10 As or more. Here, we demonstrate transcriptional slippage at poly(A) tracts within genes of *Buchnera aphidicola* associated with aphids and *Blochmannia pennsylvanicus* associated with ants. Several tracts contain single frameshift deletions; these apparent pseudogenes showed patterns of constraint consistent with purifying selection on the encoded proteins. Transcriptional slippage yielded a heterogeneous population of transcripts with variable numbers of As in the tract. Across several frameshifted genes, including *B. aphidicola* cell wall biosynthesis genes and a *B. pennsylvanicus* histidine biosynthesis gene, 12–50% of transcripts contained corrected reading frames that could potentially yield full-length proteins. *In situ* immunostaining confirmed the production of the cell wall biosynthetic enzyme UDP-N-acetylmuramyl pentapeptide synthase encoded by the frameshifted *murF* gene. Simulation studies indicated an overrepresentation of poly(A) tracts in endosymbiont genomes relative to other A + T-rich bacterial genomes. Polymerase infidelity at poly(A) tracts rescues the functionality of genes with frameshift mutations and, conversely, reduces the efficiency of expression for in-frame genes carrying poly(A) regions. These features of homopolymeric tracts could be exploited to manipulate gene expression in small synthetic genomes.

homopolymeric tracts | pseudogenes | *Blochmannia pennsylvanicus* | transcriptional slippage | *Buchnera aphidicola*

The transfer of genetic information is vulnerable to error at each of three enzyme-mediated steps: DNA replication, transcription, and translation. Homopolymeric tracts, particularly stretches of nine or more As or Ts, are prone to enzyme slippage that can alter the tract length compared with the template molecule (1–3). Long considered "hotspots" for mutation, recent studies have begun to clarify the evolutionary and functional implications of these error-prone genomic regions.

Variation generated by polymerase slippage can play an important functional role. During DNA replication, slipped-strand mispairing generates hypervariability at virulence genes of some pathogenic bacteria (4), and reversible frameshift mutations allow bacteria to toggle the expression of "contingency" genes (1). In the endosymbiont *Buchnera*, a poly(A) tract in the promoter of the heat-shock gene *ibpA* undergoes a recurrent nucleotide deletion that severely reduces transcriptional response to heat stress and is advantageous to aphid hosts under constant cool conditions (5). Likewise, transcriptional slippage, or slippage of RNA polymerase along a homopolymeric stretch, generates a heterogeneous pool of mRNA products with varied tract lengths and reading phases. Such slippage can restore the reading frame of genes with frameshifts in the genomic DNA, such as some IS elements (6), and may lead to the

synthesis of alternate functional proteins, such as the γ and τ subunits of *Thermus thermophilus* DNA polymerase III (7).

Despite these salient examples, polymerase slippage is generally considered detrimental. Evidence includes the underrepresentation of homopolymeric tracts in coding regions (compared with noncoding regions) of most bacterial genomes and especially low frequencies in highly expressed genes (6). These patterns suggest purifying selection against slippage-prone regions. The synthesis of mRNA with aberrant reading phases not only imposes a metabolic cost but also may lead to the production of potentially deleterious truncated polypeptides.

The small and AT-rich genomes of long-term bacterial endosymbionts of insects (8–16) offer a striking exception to this pattern of clear selection against slippage-prone tracts. Among the 108 bacterial genomes included in a previous analysis of poly(A) and poly(T) stretches, endosymbionts of aphids (*Buchnera*) and tsetse flies (*Wigglesworthia*) showed the highest abundance of slippery regions, even compared with other AT-rich species, and contained many long (>10-bp) tracts within coding regions (6). To explain this persistence of error-prone tracts in *Wigglesworthia* coding regions, it was suggested that transcriptional slippage might not occur in this species (6). Alternatively, if slippage occurs, long homopolymeric tracts might persist because of reduced efficacy of selection against deleterious regions in endosymbionts with small effective population sizes and/or because the tracts confer an advantage under some circumstances (14).

In this study, we clarify the functional and evolutionary significance of slippage-prone tracts where they are most abundant: the AT-rich genomes of long-term endosymbionts. We describe the experimental analysis of transcriptional slippage in long-term endosymbionts, focusing on two lineages: *Buchnera* of aphids and *Blochmannia* of ants. We simulate the abundance of long poly(A) tracts and examine the molecular evolution of
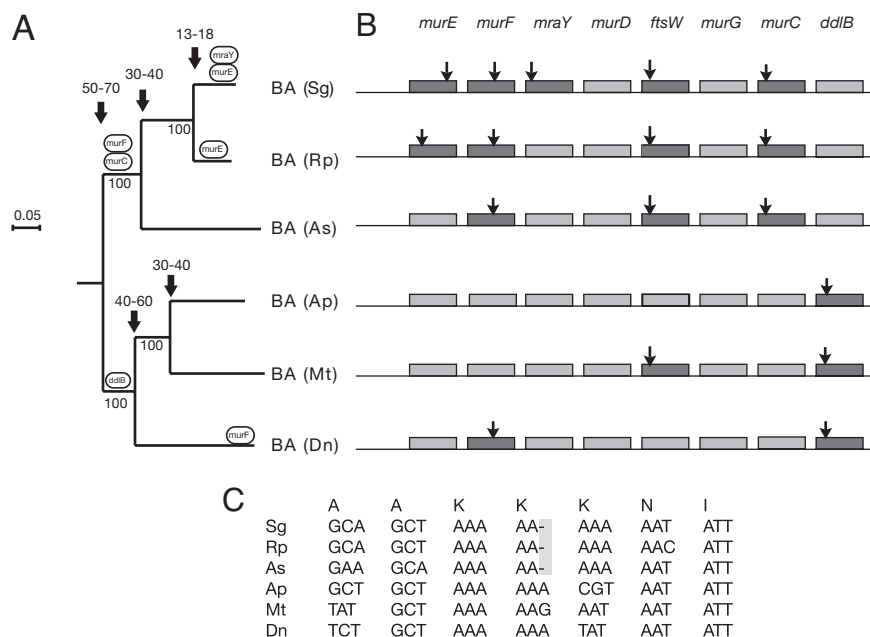
**Fig. 1.** Phylogenetic comparison reveals the acquisition and conservation of frameshift mutations in a cluster of genes for enzymes in cell wall biosynthesis. (*A*) The inference of the maximum likelihood phylogeny was performed by using a concatenated dataset of the aligned protein sequences for *murEFDGC*, *mraY*, *ftsW*, and *ddlB* from *B. aphidicola* (BA). See *Materials and Methods* for abbreviations of host aphid species. Encircled gene names indicate the node at which the individual genes acquired the deletion mutation and numbers indicated with arrows the divergence dates estimated from a date of 50–70 Myr for the split of BA (Sg) and BA (Ap) (10). Numbers below nodes refer to support values from 100 bootstrap replicates. B, *aphidicola* from *Baizongia pistaciae* (Bp) was used to root the tree. (*B*) Boxes illustrate the order of genes involved in cell wall biosynthesis, with frameshifted genes in dark gray and nonframeshifted genes in light grey. Arrows show the position of the deletion mutations inside the gene. (*C*) Nucleotide alignment of the poly(A) tract in the *murC* gene with the indicated amino acid sequence for a corrected frame in *B. aphidicola* (Sg) (Rp) and (As).

## Results and Discussion

**Purifying Selection Acts on Frameshifted Genes with Poly(A) Tracts.**
Several poly(A) tracts in endosymbiont genomes contain frame-shift mutations; these sequences have been annotated as pseu-dogenes, although their functional status is unknown. For ex-ample, in *Buchnera aphidicola* (Sg) (10), the division/cell wall (*dcw*) gene cluster contains single-nucleotide deletions within poly($A_{10}$) tracts in coding portions of the *murCEF* genes. We tested whether such loci evolve under purifying selection, con-sistent with retention of protein function. Our analysis of the *dcw* cluster from six *B. aphidicola* isolates, four sequenced as part of this study (Rp, As, Mt, Dn), revealed a pattern of cumulative fixations of frameshift deletion mutations over the past 50 million years of diversification (Fig. 1). All genes in the cluster showed dN/dS ≪ 1, indicating purifying selection [supporting information (SI) Table S1]. Likewise, the *Blochmannia pennsyl-vanicus* genome revealed frameshifts within poly(A) tracts at *hisH*, *ubiF*, *ybiS*, and *ytfM* (14). These genes were conserved across the two sequenced *Blochmannia* genomes (13, 14), with protein divergences comparable to genome-wide averages. Puz-zlingly, no additional nonsense or frameshift mutations were detected in the affected genes of either species. These compar-isons suggest that such "pseudogenes" might be functional, despite the presence of frameshift mutations.

**Heterogeneous mRNA Pool Is Produced from Genes with Poly(A) Tracts.** To examine whether the frameshifted "pseudogenes" are transcribed in *Buchnera* and *Blochmannia*, we isolated total RNA from *Rhopalosiphum padi* aphids and *Camponotus penn-*

*sylvanicus* ants, respectively. We synthesized cDNA and PCR-amplified frameshifted genes of *B. aphidicola* (*murCEF*) and *B. pennsylvanicus* (*hisH*, *ytfM*, *ubiF*, *ybiS*) and sequenced cDNA clones for each gene. The sites of the *B. aphidicola* frameshifts were tracts of 10 A, whereas in *B. pennsylvanicus* these ranged from 9 to 12 A. Three poly(A) tracts, two of which were in-frame, were analyzed in the *murC* gene of *B. aphidicola* (Rp) and one frameshifted tract each in *murE* and *murF*. Two poly(A) tracts in the *ytfM* and *ybiS* genes of *B. pennsylvanicus* were analyzed; one tract in each gene was frameshifted and one in-frame.

Cloned cDNA sequences from all frameshifted genes dis-played poly(A) tract length heterogeneity indicating transcrip-tional slippage. Frequencies of slippage ranged from 19% (for the 9 A tract) to 40–70% (for tracts of 10–12 As in length). Corrected reading frames were observed in 29–48% of the mRNAs from *B. aphidicola* (Rp) (Table 1) and in 12–33% of *B. pennsylvanicus* transcripts (Table 2). For comparison, we esti-mated the slippage frequency to 7% when the site covering the frameshifted 10 A-mer in the *murC* gene was cloned and expressed in *Escherichia coli* (Table 1). We also demonstrated slippage for in-frame poly(A) tracts, resulting in a disruption of the reading frame. Slippage from 10 to 11 A at these sites ranged from 3% to 17%. The overall proportions of full-length, in-frame gene transcripts for *ybiS*, *ytfM*, and *murC* were estimated to be 7%, 17%, and 28%, respectively, assuming that slippage at distinct poly(A) sites is independent.

Control experiments indicated the observed length heteroge-neity does not reflect PCR, cloning, or DNA sequencing artifacts (Table S2). Moreover, we conducted the same experiment using a synthetic RNA oligomer containing a poly($A_{11}$) tract. Our results illustrated that the commercial RT enzyme M-MLV, the same enzyme we used to generate endosymbiont cDNA, reverse transcribed the RNA oligomer with high fidelity. Namely, the frequency of the enzyme error was significantly lower than

# Table 1. RNA polymerase slippage at homopolymeric tracts of 10 As in mRNAs from *B. aphidicola* (BA) and *E. coli* (EC)

| Species | BA | BA | BA | EC | BA | BA |
|---|---|---|---|---|---|---|
| Locus | *murE* | *murF* | *murC*1 | *murC*1 | *murC*2 | *murC*3 |
| Frame* | FS | FS | FS | FS | + | + |
| 8A | 0 | 0 | 0 | 0 | 0 | 1 |
| 9A | 2 | 3 | 5 | 3 | 7 | 8 |
| 10A | **42** | **9** | **7** | **10** | **43** | **52** |
| 11A | 21 | 7 | 11 | 1 | 5 | 2 |
| 12A | 5 | 0 | 0 | 1 | 4 | 0 |
| 13A | 0 | 0 | 0 | 0 | 2 | 0 |
| 14A | 1 | 0 | 0 | 0 | 0 | 0 |
| 18A | 1 | 0 | 0 | 0 | 0 | 0 |
| Total | 72 | 19 | 23 | 15 | 61 | 63 |
| Slippage (%) | 42 | 53 | 70 | 33 | 30 | 17 |
| In-frame (%) | 29 | 37 | 48 | 6.7 | 73 | 83 |

Numbers in bold, the number of As in the original genomic DNA. Numbers underlined, the number of As in-frame.

*FS, frameshifted gene; +, open reading frame.

slippage inferred at endosymbiont genes with 10 or 11 A tracts ($P < 0.0001$, Fisher's exact test), and the few length variants observed showed a different spectrum of change. This experiment showed that reverse transcription of endosymbiont RNA using M-MLV did not generate the observed variation in cDNA sequences. Combined, these controls rule out various potential experimental artifacts. We conclude that the observed cDNA heterogeneity reflects the underlying variation in transcript sequences and is attributable to slippage by endogenous RNA polymerase in the endosymbionts.

**MurF Protein Is Present in *Buchnera* Cells and Aphid Embryos.** Having shown that a fraction of the heterogeneous mRNA population produced from frameshifted genes could serve as substrates for translation of full-length proteins, we sought to determine whether the corresponding proteins were expressed. Western blots on whole *R. padi* aphid protein extracts using polyclonal antibodies raised against a MurF-derived peptide fragment gave a single specific band corresponding to the predicted molecular weight of the full-length MurF (Fig. S1). To establish whether the detected protein was of aphid or *B. aphidicola* origin, we performed *in situ* immunofluorescence on material from *R. padi* embryos. The results clearly demonstrated that

# Table 2. RNA polymerase slippage at homopolymeric tracts of 9–12 As in mRNAs from *B. pennsylvanicus* (BP)

| Species | BP | BP | BP | BP | BP | BP |
|---|---|---|---|---|---|---|
| Locus | *hisH* | *ytfM-1* | *ubiF* | *ybiS-2* | *ytfM-2* | *ybiS-1* |
| Frame* | FS | FS | FS | FS | + | + |
| 8A | 0 | 2 | 0 | 3 | 0 | 0 |
| 9A | 1 | 4 | 4 | 21 | 0 | 8 |
| 10A | 11 | **10** | **10** | 2 | 2 | 11 |
| 11A | **25** | 8 | 5 | 0 | 7 | 4 |
| 12A | 15 | 0 | 2 | 0 | **13** | 0 |
| 13A | 4 | 0 | 0 | 0 | 4 | 0 |
| 14A | 3 | 0 | 0 | 0 | 0 | 0 |
| 16A | 1 | 0 | 0 | 0 | 0 | 0 |
| Total | 60 | 24 | 21 | 26 | 26 | 23 |
| Slippage, % | 58 | 58 | 52 | 19 | 50 | 52 |
| In-frame, % | 27 | 33 | 29 | 12 | 50 | 48 |

Numbers in bold, the number of As in the original genomic DNA. Numbers underlined, the number of As in frame.
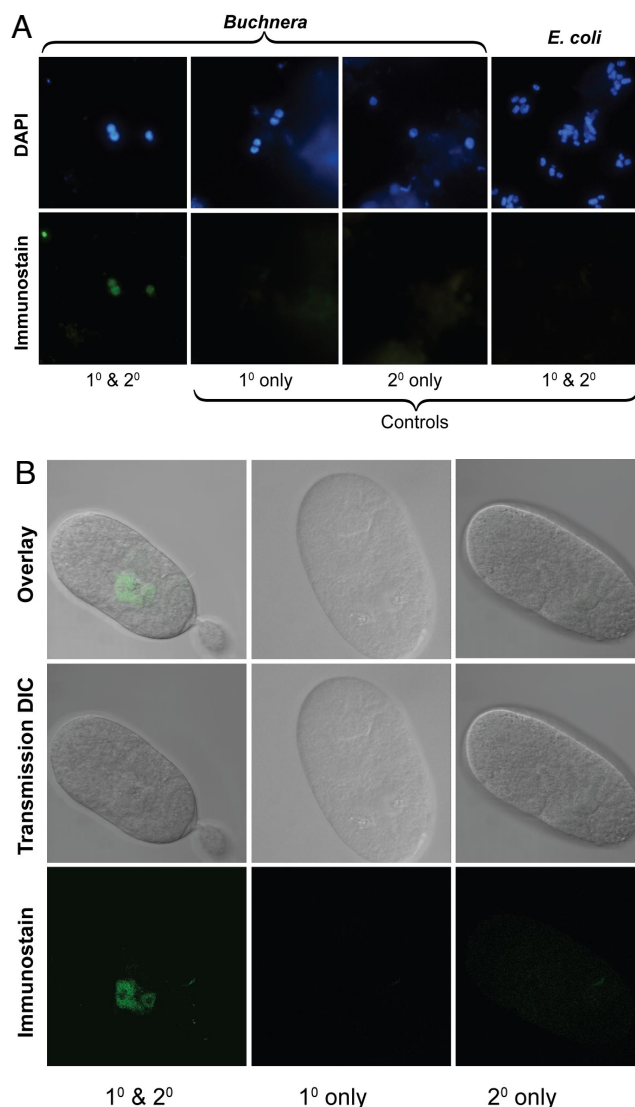
*FS, frameshifted gene; +, open reading frame.

**Fig. 2.** Immunolocalizaton of UDP-N-acetylmuramyl pentapeptide synthase encoded by the frameshifted *murF* gene. (*A*) Isolated *B. aphidicola* cells with secondary antibody detection using Alexa Fluor 488. (*B*) The mycetome of stage 12 embryos from the aphid *R. padi*, with secondary antibody detection using Alexa Fluor 633. The primary antibody was designed to the C terminus end of the UDP-*N*-acetylmuramyl pentapeptide synthase and therefore specific to the full functional protein.

the signal was localized to and specific for *B. aphidicola* (Fig. 2*A*). Further immunostaining of intact *R. padi* embryos localized the signal to mycetomes (Fig. 2*B*), the specialized structures in which *B. aphidicola* are housed within their aphid hosts. This implies that *B. aphidicola* synthesize the full-length MurF protein despite a frameshift in the *murF* gene.

**Simulation Analyses.** The number of poly(A) tracts increases with the genomic A + T content, with up to 1,491 tracts of 10 or more As per Mb of coding sequences in *B. aphidicola* (Cc), which has a genomic G + C content of only 20% (Table S3). We performed a simulation study to test whether the overall abundance of poly(A) tracts is higher than expected by chance from the A + T bias in the endosymbiont genomes. For each genome, we generated 10,000 simulated genomes, using species-specific codon frequencies and gene lengths, comparing the distribution profiles of poly(A) tracts in the genes produced *in silico* with the occurrences in the real
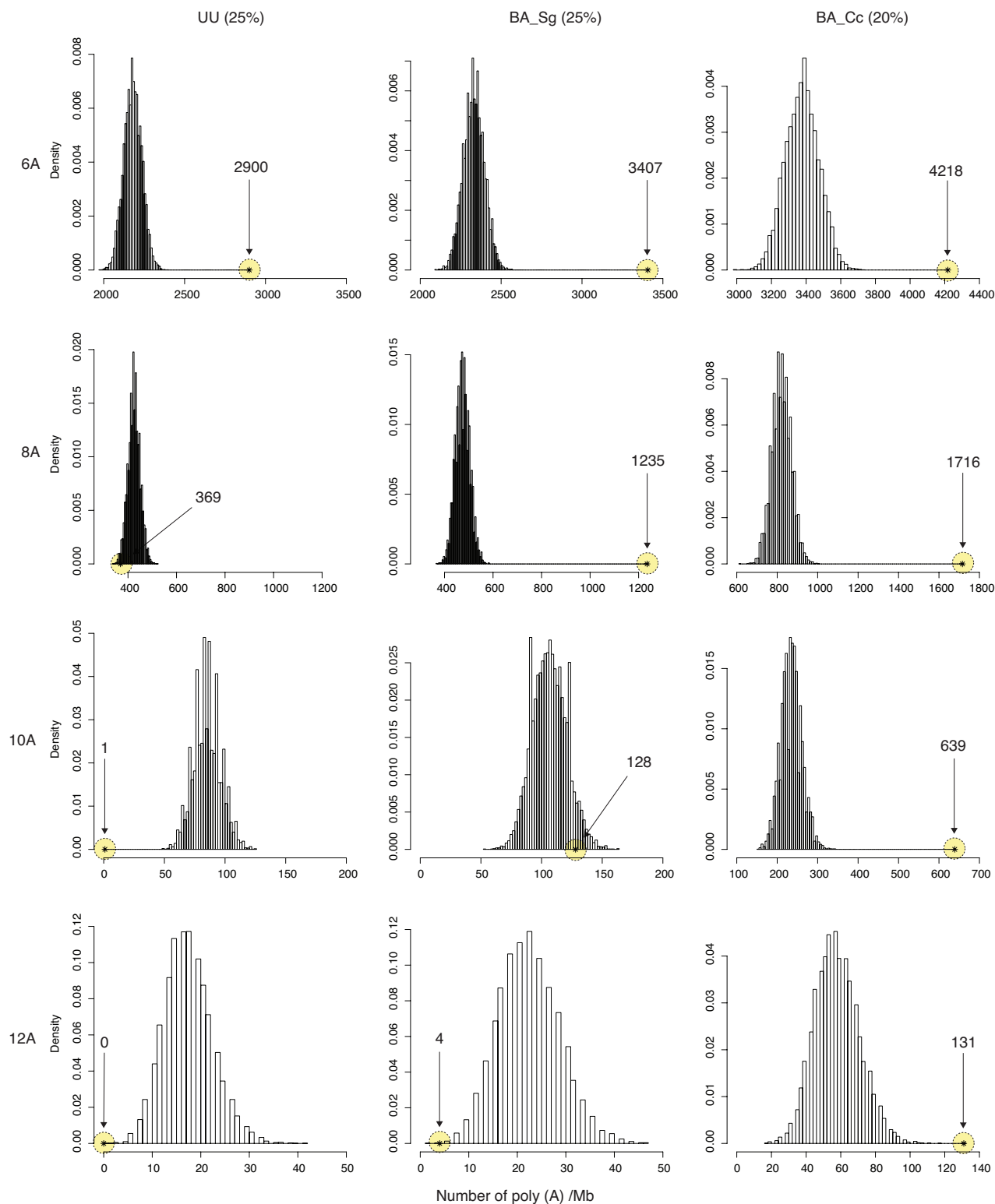
**Fig. 3.** Frequency distribution of poly(A) tracts. The observed number of poly(A) tracts per Mb of sequence data (stars inside yellow circles) compared with poly(A) density distribution profiles generated for 10,000 simulated genomes (histograms) for *U. urealyticum* (UU), *B. aphidicola* (Sg), and *B. aphidicola* (Cc). Numbers in parenthesis refer to genomic G + C content values. The lengths of the poly(A) tracts (6, 8, 10, and 12 A) are indicated at the left for each horizontal set of plots. The number of poly(A) tracts ($X$) was normalized ($X_{norm}$) to represent poly(A) tracts per 1 Mb by $X_{norm} = X \times (1,000\,000/SeqLength_{tot})$, where SeqLength$_{tot}$ is the total number of nucleotides in the dataset. Full details including statistics and additional species are presented in Fig. S2.

genes. The results showed that poly($A_{10}$) tracts within coding regions in endosymbiont genomes were at least as abundant as expected from their genomic G + C contents (20–29% GC) (Fig. 3; Fig. S2). In contrast, no poly($A_{10}$) tracts occur within genes of the 5.7-Mb genome of *E. coli*-K12 (17), and only a single such tract exists in coding regions of *Ureaplasma urealyticum* (18) despite its

EVOLUTION

biased base composition (25% GC) (Fig. 3). This implies reduced strength or efficacy of selection against slippage-prone poly(A) tracts in endosymbiont genomes.

**Rescue of Gene Function.** We have shown that transcriptional slippage can restore the reading frame of endosymbiont genes that contain frameshifts in poly(A) tracts. In these cases, slippage has a clear advantage of rescuing gene function. Polyploidy of endosymbiont genomes (19) might also rescue gene function by raising the possibility that a minor fraction of chromosomes encode intact gene copies. However, the frameshift mutations are well supported from multiple clones in genomic DNA libraries and independent PCR verification with no indications of sequence variation. This suggests that all, or nearly all, copies of the gene within a host contain the frameshift mutation, at least at the time that the genomic DNA was sampled.

Once a frameshift mutation becomes fixed within a poly(A) tract, then retention of adenines within that tract becomes essential for restoring the reading frame via transcriptional slippage. However, because DNA polymerases are also subject to slippage errors, poly(A) tracts are thought to be highly unstable during DNA replication, and mutations correcting the frame should be frequent. This raises the question of what evolutionary forces could maintain the frameshift itself over a long evolutionary time scale.

Based on our data (Tables 1 and 2), we expect that frameshift mutations generally reduce the abundance of transcripts for full-length proteins. Might down-regulation of endosymbiont genes be advantageous? The transcript heterogeneity we observe suggests that long poly(A) tracts in coding regions are not hypermutable "ON-OFF" molecular switches (20). Occasional DNA mutations correcting a frameshift may instead be favored under specific environmental conditions (1). However, the frameshift-containing genes have no obvious link to environmental variability, and we found no evidence for variation in poly(A) tract lengths in the genomes of endosymbionts infecting closely related host species.

Another possibility is that transcriptional slippage at poly(A) tracts serves to compensate for RNA polymerase slippage at nearby sites. However, the sequencing of 300–600 bp flanking each of the poly(A) tracts in the cDNAs provided no evidence for RNA polymerase slippage outside the tract. We also addressed the potential for compensatory changes in the genomic DNA sequences, but a manual inspection of the sequence alignments revealed no additional indel mutations upstream of the poly(A) tract. This rules out selection on out-of-frame homopolymers to compensate for indel mutations and/or RNA polymerase slippage at other locations in the gene or vice versa. In rare cases, alternative decoding is used to produce two overlapping gene products in different stoichiometry from the same sequence (6, 7, 21). However, subfunctionalization by alternative decoding seems unlikely to explain the data presented here, because no novel gene functions are anticipated for the truncated forms of the frameshifted gene products.

Although poly(A) tracts themselves likely accumulate by a strong A + T mutation bias (22), the long-term stability of indels in some tracts may potentially reflect selection for reduced gene expression levels. It is also possible that frameshifts in poly(A) tracts while functionally rescued by transcriptional slippage currently represent the earliest stage in the degradation of these endosymbiont genes. In the *B. aphidicola* species analyzed here, this may lead to the loss of the endosymbiont cell wall and replacement with host cell membranes, as in the case of organelle cell membrane proteins that are mostly encoded by and derived from the host nuclear genome. Incidentally, the *mur* genes have been lost from the *B. aphidicola* (Cc) genome (15).

**Disruption of Gene Function.** Although transcriptional slippage restores disrupted reading frames, it also disrupts intact coding regions. We demonstrated slippage at in-frame poly(A) tracts in both aphid (Table 1) and ant (Table 2) endosymbiont genomes and observed an even distribution of in-frame poly(A) tracts across genes of all functional categories (Tables S4–S6). Given that up to several hundred sites per genome are prone to transcriptional slippage, truncated gene products must accumulate at high frequencies in the endosymbionts.

Assuming that disruption of the reading frame by transcriptional slippage is generally detrimental, the persistence of error-prone tracts in endosymbiont coding regions might be explained by reduced efficacy of selection in these species. Several lines of evidence point to reduced effective population size ($N_e$) in endosymbionts, a culmination of successive population bottlenecks upon transmission from host mother to offspring (23), low recombination rates, and demographic fluctuations in insect host populations (24, 25). Under predictions of nearly neutral evolution (26), reduced $N_e$ of endosymbionts is expected to accelerate the fixation of deleterious mutations by random genetic drift. Endosymbiont genomes and DNA sequence variation show many signs of deleterious evolution in small populations (22, 27, 28), and the persistence of slippery tracts may be another example of genetic degradation.

Transcriptional slippage is probably far more common in AT-rich endosymbiont genomes than in free living bacterial genomes that have moderate base compositions and experience stronger and/or more effective selection against poly(A) tracts. Moreover, the dynamic process of genome reduction and accumulation of poly(A) tracts may yield an increasing fraction of truncated protein products as the deterioration process continues. The chaperone GroEL is constitutively expressed at levels that correspond to 10% of the total protein in *B. aphidicola* (29). Notably, this gene does not contain any poly(A) tracts of 10 bp or longer in any of the *B. aphidicola*, *B. pennsylvanicus*, or *Wigglesworthia glossinidia* genomes. High levels of GroEL may mask the effects of destabilizing amino acid replacements on protein conformations (30, 31) and might therefore buffer against deleterious effects of truncated proteins, facilitating their degradation.

## Conclusions

Our results show that mRNA sequence heterogeneity at poly(A) tracts prevents frameshift mutations from destroying gene functions. We have demonstrated continued purifying selection on genes with such frameshifts, plus expression of full-length genes at both RNA and protein levels. These results contradict the hypothesis that transcriptional slippage does not occur in endosymbiont genomes (6). Although a majority of the slippage-prone poly(A) tracts may have accumulated despite an overall deleterious effect on the efficiency of expression, a few may potentially have been selected to reduce the frequency of mRNAs with intact ORFs. The result is a robust but inefficient system in which polymerase slippage during replication is modulated by subsequent errors during transcription and translation. Because factors such as robustness and accuracy of the information processing systems are crucial for cell functionality, knowledge about their dependencies in naturally small bacterial genomes may help in the design of synthetic small genomes.

## Materials and Methods

**DNA Sequencing and Transcription Analysis.** Aphid genomic DNA was prepared as described (32). The 11-kb fragment containing the *mur* genes was amplified and sequenced from *B. aphidicola* of the aphid species *Diuraphis noxia* (Dn), *Macrosiphoniella tanecetaria* (Mt), *R. padi* (Rp), and *Aphis sambuci* (As). The inference of the maximum-likelihood phylogeny of these *B. aphidicola* strains and those sequenced from *Acyrthosiphon pisum* (Ap) (9) and *Schizaphis graminum* (Sg) (10) was performed by using a concatenated dataset of the

aligned protein sequences for *murEFDGC*, *mraY*, *ftsW*, and *ddlB* from with the WAG+G substitution model in TREEFINDER (33).

Total aphid and ant RNA was extracted by using TRIzol (Invitrogen) from an *R. padi* culture maintained on barley and using RNAqueous or TRI reagent (Ambion) from field-collected *C. pennsylvanicus* pupae, respectively. First-strand cDNA was synthesized with the reverse transcriptase M-MLV using random primers and DNase-treated total RNA as a template (Invitrogen or Ambion for aphids and ants, respectively). Products were amplified from this cDNA template with specific PCR primers, cloned (TOPO TA cloning kit, Invitrogen), and sequenced by using an ABI3730xl DNA Analyzer (Applied Biosystems).

**Protein Expression.** *R. padi* aphids were maintained as a clonal line of parthenogenic females on young barley plants, at a constant temperature of 20°C with a light regime of 16 h light and 8 h dark. Mature adult aphids where collected and dissected to remove embryos. Bacterial preparations were isolated by disrupting aphids in ice-cold Buffer A (50 mM Tris·HCl, pH 7.5, 0.25 M sucrose) in a hand-held glass homogenizer until insect cells were broken, followed by centrifugation at $2,000 \times g$ for 5 min to remove large insect debris. The supernatant was then centrifuged at $14,000 \times g$ for 15 min to pellet the bacterial cells, and the pellet was resuspended by flicking into the fixative comprising ethanol formaldehyde (3:1, vol/vol). Approximately 10-$\mu$l volumes were spotted onto clean glass slides and air-dried at 60°C.

Total whole-aphid protein extracts were obtained from 30 to 40 individuals of the summer parthenogenetic form of *R. padi*. Embryos were collected by dissection of 50 adults. *B. aphidicola* cells were extracted from adult *R. padi* aphids as described in ref. 34. Proteins were separated by SDS/PAGE by using the MiniProtean II system (BioRad) and transferred to PVDF membranes (Millipore) probed with anti-*murE* and *murF* antibodies (Thermo) followed by anti-rabbit secondary antibody conjugated with alkaline phosphatase. Immunostaining was performed by using anti-*murF* antibodies with Alexa Fluor 488 (bacterial preparations) and Alexa Fluor 633 (embryo preparations). Preparations were viewed by using a Zeiss fluorescent microscope DMRXE with a Hamamatsu OrcaIII camera and OpenLab image capturing software for the bacterial preparations or Zeiss confocal TCS-SP with Ar-laser for 488 and 514 nm excitation and HeNe-laser for 633-nm excitation (embryo preparations).

1. Koch AL (2004) Catastrophe and what to do about it if you are a bacterium: The importance of frameshift mutants. *Crit Rev Microbiol* 30:1–6.
2. Wagner LA, Weiss RB, Driscoll R, Dunn DS, Gesteland RF (1990) Transcriptional slippage occurs during elongation at runs of adenine or thymine in *Escherichia coli*. *Nucleic Acids Res* 18:3529–3535.
3. Gurvich OL, *et al.* (2003) Sequences that direct significant levels of frameshifting are frequent in coding regions of *Escherichia coli*. *EMBO J* 22:5941–5950.
4. Moxon R, Bayliss C, Hood D (2006) Bacterial contingency loci: The role of simple sequence DNA repeats in bacterial adaptation. *Annu Rev Genet* 40:307–333.
5. Dunbar HE, Wilson AC, Ferguson NR, Moran NA (2007) Aphid thermal tolerance is governed by a point mutation in bacterial symbionts. *PLoS Biol* 5:e-96.
6. Baranov PV, Hammer AW, Zhou J, Gesteland RF, Atkins JF (2005) Transcriptional slippage in bacteria: distribution in sequenced genomes and utilization in IS element gene expression. *Genome Biol* 6:R25.
7. Larsen B, Wills NM, Nelson C, Atkins JF, Gesteland RF (2000) Nonlinearity in genetic decoding: Homologous DNA replicase genes use alternatives of transcriptional slippage or translational frameshifting. *Proc Natl Acad Sci USA* 97:1683–1688.
8. Dale C, Moran NA (2006) Molecular interactions between bacterial symbionts and their hosts. *Cell* 126:453–465.
9. 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.
10. Tamas I, *et al.* (2002) 50 million years of genomic stasis in endosymbiotic bacteria. *Science* 296:2376–2379.
11. Akman L, *et al.* (2002) Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia, Nat Genet* 32:402–407.
12. van Ham RC, *et al.* (2003) Reductive genome evolution in. *Buchnera aphidicola. Proc Natl Acad Sci USA* 100:581–586.
13. Gil R, *et al.* (2003) The genome sequence of. *Blochmannia floridanus*: Comparative analysis of reduced genomes. *Proc Natl Acad Sci USA* 100:9388–9393.
14. 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.
15. Pérez-Brocal V, *et al.* (2006) A small microbial genome: The end of a long symbiotic relationship? *Science* 314:312–313.
16. Nakabachi A, *et al.* (2006) The 160 kb genome of the bacterial endosymbiont *Carsonella*. *Science* 314:267.
17. Blattner FR, *et al.* (1997) The complete genome sequence of. *Escherichia coli* K-12. *Science* 277:1453–1474.
18. Glass JI, *et al.* (2000) The complete sequence of the mucosal pathogen *Ureaplasma urealyticum*. *Nature* 407:757–762.
19. Komaki K, Ishikawas H (2000) Genomic copy number of intracellular bacterial symbionts of aphids varies in response to developmental stage and morph of their host. *Insect Biochem Mol Biol* 30:253–258.
20. Woude MW (2007) Re-examining the role and random nature of phase variation. *FEMS Microbiol Lett* 254:190–197.
21. Baranov PV, Fayet O, Hendrix RW, Atkins JF (2006) Recoding in bacteriophages and bacterial IS elements. *Trends Genet* 22:174–181.
22. Moran NA (1996) Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proc Natl Acad Sci USA* 93:2873–2878.
23. Mira A, Moran NA (2002) Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. *Microb Ecol* 44:137–143.
24. Funk DJ, Wernegreen JJ, Moran NA (2001) Intraspecific variation in symbiont genomes: Bottlenecks and the aphid-*Buchnera* association. *Genetics* 157:477–489.
25. Abbot P, Moran NA (2002) Extremely low levels of genetic polymorphism in endosymbionts (*Buchnera*) of aphids (*Pemphigus*). *Mol Ecol* 11:2649–2660.
26. Ohta T (1973) Slightly deleterious mutant substitutions in evolution. *Nature* 246:96–98.
27. Moran NA, Baumann P (2000) Bacterial endosymbionts in animals. *Curr Opin Microbiol* 3:270–275.
28. Lambert JD, Moran NA (1998) Deleterious mutations destabilize ribosomal RNA in endosymbiotic bacteria. *Proc Natl Acad Sci USA* 95:4458–4462.
29. Baumann P, *et al.* (1996) Levels of. *Buchnera aphidicola* chaperonin GroEL during growth of the aphid *Schizaphis graminum* Curr Microbiol 32:279–285.
30. Fares MA, Ruiz-Gonzalez MX, Moya A, Elena S, Barrio E (2002) Endosymbiotic bacteria: GroEL buffers against deleterious mutations. *Nature* 417:398.
31. Fares MA, Moya A, Barrio E (2004) GroEL and the maintenance of bacterial endosymbiosis. *Trends Genet* 20:413–416.
32. Wernegreen JJ, Ochman H, Jones IB, Moran NA (2000) Decoupling of genome size and sequence divergence in a symbiotic bacterium. *J Bacteriol* 182:3867–3869.
33. Jobb G, von Haeseler A, Strimmer K (2004) TREEFINDER: A powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol* 4:18.
34. Harrison CP, Douglas AE, Dixon AFG (1998) A rapid method to isolate symbiotic bacteria from aphids. *J Inv Pathol* 53:427–428.

EVOLUTION