Bacteriocyte-Associated Symbionts of Insects

A variety of insect groups harbor ancient prokaryotic endosymbionts

Nancy A. Moran and Aparna Telang

Endosymbiosis has been proposed as an evolutionary innovation that underlies the success and diversification of some groups of organisms by enabling them to exploit new ecological niches (e.g., Margulis and Fester 1991, Maynard Smith and Szathmary 1995). Three kinds of information are needed to evaluate this proposal. First, it is necessary to know something of the role played by symbionts in the lives of hosts: Are their effects ones that allow use of novel ecological niches? Second, information is needed on the evolutionary history of endosymbioses: How old are they and, in particular, do endosymbiotic infections predate diversification of major host groups? Finally, the evolutionary consequences for hosts of acquiring new capabilities in the form of symbiotic associations must be evaluated: Are these novel traits stable? Do they impose constraints as well as new capabilities on hosts?

Recent studies have begun to shed light on these questions for bacteriocyte-associated endosymbioses in insects. Bacteriocytes (or mycetocytes) are cells of animal hosts that are positioned at characteristic locations in the body and appear to be specialized for housing bacteria (Figure 1; Buchner 1965, Dasch et al. 1984). Bacteriocyte associates are found in many insect orders and families and have been estimated to occur in 10% of insect species (Douglas 1989).

Before the past decade, much of the knowledge of bacteriocyte associations was based on extensive microscopical studies by Paul Buchner and his associates. His fascinating compilation of this work (Buchner 1965) details astonishing diversity in animal-microbe associations. Most of this diversity is still unexplored, in large part because intracellular symbionts typically cannot be cultured outside of their hosts. Since 1990, however, molecular studies on a handful of endosymbioses have yielded considerable insight into their evolutionary histories and into the specific adaptations to symbiosis that are found in the bacteria themselves. Some of the most revealing of these studies concern the bacteriocyte-associated, mutualistic endosymbionts of insects.

The microscopical explorations performed by Buchner and his associates gave only hints regarding the relationships of insect endosymbionts to other bacterial groups and to other endosymbionts. The characterization of these bacteria and their phylogenetic placement relative to free-living prokaryotes remained almost a complete mystery until the development of molecular methods, in particular methods for determining nucleotide sequences. These methodological advances have revolutionized understanding of prokaryote evolution in general by providing a surplus of relatively reliable characters for phylogeny reconstruction. A major additional advantage of DNA sequences as phylogenetic characters is that they can be obtained for noncultivable organisms. In particular, sequences of the 16S ribosomal RNA (rRNA) genes have been used for the phylogenetic and taxonomic placement of many bacteria, including endosymbionts, that were previously unclassifiable (Maidak et al. 1996).

In this article, we provide a brief overview of recent work on the nature of the relationships between insect hosts and bacteriocyte-associated organisms (see Dasch et al. 1984, Dadd 1983, Douglas 1989 for detailed reviews). We then consider in more detail the emerging picture of endosymbiont evolution presented...
Finally, we consider how such studies bear on the following sets of questions: First, what is the distribution of endosymbiotic bacteria on the prokaryotic tree of life? Which bacterial groups have produced lineages that are endosymbiotic in insects? Are these related to one another as members of one or a few monophyletic groups specialized for endosymbiotic living? Or do endosymbionts of different insect lineages represent independent origins of endosymbiotic bacteria? Second, do the bacteria show long-term patterns of cospeciation with their hosts? For many insect symbioses, beneficial effects of the bacteria on their hosts are supported by experimental evidence: Antibiotic or heat treatments causing loss of symbionts are accompanied by reductions in performance, including lowered growth and survivorship and loss of reproductive abilities (e.g., Nogge 1976, Ishikawa and Yamaji 1983, Ohtaka and Ishikawa 1991, Prosser and Douglas 1991, Sasaki et al. 1991, Costa et al. 1993, Heddi et al. 1993; additional studies cited in Koch 1967 and Dasch et al. 1984). Many hosts, including aphids and other insect groups that feed exclusively on plant sap, can grow or reproduce only with intact symbionts. Others, including Sitophilus weevils, which feed on grain, can survive and reproduce without the symbionts, but they grow more slowly and are smaller than symbiotic hosts (Heddi et al. 1993). Although interpreting studies of symbiont-free hosts is complicated by the possibility that antibiotic or heat treatment may have direct effects on host metabolism, the preponderance of evidence points to a beneficial role of bacteriocyte associates for host insects.

A nutritional dependence of hosts on their symbionts presumably explains why many insect taxa possessing bacteriocytes use narrow and nutritionally unbalanced diets, such as the phloem sap of plants or the blood of vertebrates. Because animals lack biosynthetic pathways that are typically present in prokaryotes, most animals must ingest a large number of essential nutrients. Alternatively, a microbial associate can provide the missing nutrients, allowing animal hosts to exploit resources that would otherwise be nutritionally deficient. Examples include members of the Hemiptera that use their sucking mouthparts to feed solely on plant phloem or xylem sap (Dasch et al. 1984). Plant phloem sap contains free amino acids but lacks some amino acids that are essential nutrients for animals. Another dietary habit associated with endosymbiosis is restriction to vertebrate blood, which is lacking in some vitamins. Endosymbiosis connected to blood feeding throughout the life cycle has evolved independently in a
number of insect orders, including Diptera, Hemiptera, and Mallophaga (Buchner 1965, Dasch et al. 1984). Some insects with more generalized diets also have endosymbionts; these include cockroaches, some ants, and some weevils. In ants, however, endosymbiosis may be associated with dependence on nutritionally deficient plant fluids.

The general conclusion of a nutritional role for symbionts has been accepted by almost all researchers who have studied bacteriocyte associates of insects (e.g., Buchner 1965, Koch 1967, Dasch et al. 1984, Dadd 1985). Yet definitive evidence that endosymbionts provide essential nutrients to insect hosts is rare. The best evidence pertains to the provision of tryptophan by Buchnera, the endosymbionts of aphids (Douglas and Prosser 1992, Lai et al. 1994). Also, some evidence suggests that tsetse symbionts provision hosts with B vitamins (Nogge 1981).

How are endosymbionts transmitted?

Bacteriocyte-associated endosymbionts are typically maternally inherited, often through complex developmental events that ensure transovarial transfer from the mother to the developing egg or embryo (Buchner 1965, Dasch et al. 1984). In the case of aphids and Buchnera, the bacteria are housed within bacteriocytes that are grouped into a loose organ within the body cavity, in the vicinity of the developing ovarioles (Figure 1b; Buchner 1965, Hinde 1971a, 1971b). The maternal bacteriocyte adjacent to an embryo near the blastoderm stage forms a small opening through which a bacterial inoculum passes. The inoculum then moves through the intervening hemolymph and enters a nearby opening on the oocyte surface. During early embryonic development, the presumptive bacteriocytes form, and the inoculating bacteria migrate into these cells.

This apparent fine tuning of the infection process is typical of bacteriocyte-associated symbionts in insects and suggests a long history of selection favoring host adaptations that help to maintain the association. However, the particulars of the infection process vary tremendously among insect groups (e.g., Sacchi et al. 1988, Hypsa 1993). For example, in whiteflies, an entire maternal bacteriocyte (or, in some species, several bacteriocytes) is transferred into each egg, with the maternal nucleus later degenerating (Buchner 1965, Costa et al. 1993).

Secondary endosymbionts

Some insects harbor so-called "secondary" endosymbionts that coexist in the same individual hosts with the bacteriocyte-inhabiting primary endosymbionts. These secondary symbionts are maternally inherited but do not appear to share a long evolutionary history with their hosts. They are usually not present in the bacteriocytes that house primary endosymbionts (Figure 1c; Buchner 1965). Tsetse fly secondary symbionts occur in midgut cells, by contrast to the primary symbionts, which are found in specialized bacteriocytes in the anterior gut (Pinnock and Hess 1974). Aphid secondary endosymbionts occur mostly in a sheathlike syncytium bordering the bacteriocytes that house Buchnera (Buchner 1965, Fukatsu and Ishikawa 1993, Chen and Purcell 1997). Whitefly secondary endosymbionts do inhabit the bacteriocytes in which primary endosymbionts are found but are clumped irregularly within vacuoles (Costa et al. 1993, 1996).

Effects of secondary endosymbionts on host fitness are not clear. Buchner (1965) described secondary endosymbionts of aphids as "recently acquired guests which are still in need of adaptation." They are more sporadic in their numbers and presence among species and even in their distribution among individuals of the same species, as has been shown for aphids (Fukatsu and Ishikawa 1993, Chen and Purcell 1997), whiteflies (Costa et al. 1993), and tsetse flies (Pinnock and Hess 1974). Chen and Purcell (1997) found that secondary endosymbionts are present in only some strains of pea aphid (Apositorphaspis rosae) and are identical in 16S rDNA sequence to secondary endosymbionts of rose aphid (Macrosiphum rosae). Vertical transmission from the time of the common ancestor of these two hosts would almost certainly have resulted in some sequence divergence; for example, the Buchnera 16S rDNA sequences for similarly closely related genera within Aphididae all differ by more than 2% (Munson et al. 1991).

The observed sequence identity of the secondary symbionts implies the occurrence of either horizontal transfer between species or recent, independent infections by a bacterium that is widely distributed in the environment. Furthermore, the symbionts could be experimentally transferred by injecting hemolymph from infected clones into uninfected clones of the same or different species. Thus, although maternal transmission was observed (Chen and Purcell 1997) and is presumably the usual route of infection, these secondary endosymbionts may also undergo some transfer among host lineages, including transfer between species. Their sporadic distribution within host species implies that unlike primary endosymbionts, they are not required for host development or reproduction. Currently, there is almost no evidence indicating whether their effects on host fitness are positive, negative, or neutral.

The phylogenetic distribution of insect endosymbionts

Molecular phylogenetic results based on 16S rDNA sequences show that within each of the insect groups examined so far, the primary endosymbionts descended from a single ancestor, as indicated by their forming a well-supported clade. This result has been obtained for the primary endosymbionts of aphids (Munson et al. 1991, Moran et al. 1993), mealbugs (Munson et al. 1993), whiteflies (Clark et al. 1992), carpenter ants (Candidatus camponotii; Schröder et al. 1996), Sitophilus weevils (Campbell et al. 1992), tsetse flies (Aksoy 1995), and cockroaches plus termites (Bandi et al. 1994, 1995). The endosymbionts of each of these insect groups form a distinct monophyletic group. The most obvious explanation for these findings is that the common ancestor to each of these clades was also endosymbiotic in an ancestor to the same group of hosts.

The positions of these endosymbiont clades on the bacterial phylo-
A number of insect endosymbionts fall within the gamma-3 subdivision of the Proteobacteria, a diverse group that includes *Escherichia coli* and other enteric bacteria. There are two main groups of insect symbionts within this subdivision. Clustered within the Enterobacteriaceae near *E. coli* are the secondary endosymbionts of pea aphids (Unterman et al. 1989), the secondary endosymbionts of tsetse flies (Aksoy et al. 1995), and the primary endosymbionts of *Sitophilus* weevils (Campbell et al. 1992). Together, these endosymbionts may form a clade, although variation in 16S rDNA sequences is not sufficient to resolve relationships within the gamma subdivision (Figure 2).

Also within the gamma-3 subdivision, but outside of the Enterobacteriaceae, are the secondary endosymbionts of pea aphids (Unterman et al. 1989), the secondary endosymbionts of tsetse flies (Aksoy et al. 1995), and the primary endosymbionts of *Sitophilus* weevils (Campbell et al. 1992). Together, these endosymbionts may form a clade, although variation in 16S rDNA sequences is not sufficient to resolve relationships within the gamma subdivision (Figure 2).
riaceae, is a clade consisting of primary endosymbionts of aphids, carpenter ants, and tsetse flies, with each of the three symbiont groups forming a well-supported clade. Current phylogenetic evidence raises the possibility that the shared ancestor for tsetse, ant, and aphid primary symbionts was endosymbiotic or possessed habits causing it to readily enter into endosymbiotic associations with insects. However, this possibility is not yet firmly established; indeed, it would be surprising in view of major differences among the host taxa and among the developmental patterns of the different symbioses. The 16S rDNA sequences of these endosymbionts have a low G+C content compared with free-living members of the gamma-3 subdivision (Table 1), raising the possibility that their apparent close phylogenetic relationship results in part from convergent evolution at some nucleotide positions. Also, undiscovered free-living bacteria may belong to the same clade. Thus, phylogenetic evidence from 16S rDNA is inconclusive but consistent with the independent derivation of aphid, ant, and tsetse endosymbionts from free-living ancestors.

The best-known exception to the concentration of insect bacteriocyte associates within the Proteobacteria are the intracellular symbionts of cockroaches and primitive termites (Bandi et al. 1994, 1995). These symbionts form a clade within the bacteroides–flavobacteria group and are, thus, distantly related to other known bacteriocyte associates.

### Ages and patterns of cospeciation

Another consistent pattern that has emerged for bacteriocyte-associated bacteria in insects is congruence of the phylogenies of the bacterial and host clades. For every case that has been studied sufficiently, the phylogeny for the bacterial associates matches the phylogeny for the corresponding insects. This congruence is strong evidence for parallel diversification arising from faithful and long-term vertical transmission of endosymbionts along host lineages. Support for congruence varies among the cases studied; it depends on the number of taxa included and on the robustness of the phylogenies. Varying levels of positive support have been found for aphids (Munson et al. 1991, Moran et al. 1993), tsetse flies (Aksoy et al. 1995), carpenter ants (Schröder et al. 1996), and cockroaches (Bandi et al. 1995). Even more important, no study has yet produced evidence contradicting the hypothesis of parallel phylogenesis for bacteriocyte-associated endosymbionts in insects. Thus, the syndrome of long-term associations that are strictly maternally inherited may be typical of these associations.

The support for long-term codiversification of insects and their endosymbionts is consistent with micropscopical and experimental observations of maternal transmission by Buchner (1965) and others. However, the phylogenetic studies extend these results by demonstrating that even rare instances of horizontal transmission are absent, a fact that could never be established based on experimental observations of transmission events alone. Even infrequent horizontal transmission events would scramble patterns of congruence of host–symbiont phylogenies when such phylogenies extend millions of years into the past.

For each insect bacteriocyte association, congruence of host and symbiont phylogenies provides strong evidence that the original infection predates the basal ancestor of the symbiont and of the corresponding hosts. In addition, congruence implies that those ancestors lived at the same time. Consequently, minimum ages of the infections can be inferred from the fossil records of the insects (Table 2). The general conclusion from thus incorporating paleontological evidence is that these infections are very old, dating back to perhaps 300 million years in the case of the clade containing cockroaches plus termites and to 200 million years in the case of the clade represented by aphids (Table 2). Schröder et al. (1996) suggest that the ancestor of all ants may have possessed an endosymbiont that was subsequently lost in most lineages but retained in carpenter ants and certain other species in the subfamily Formicinae. This hypothesis remains to be tested through examination of other formicines. If it is correct, the age of the infection would exceed 40 million years. In addition to the evidence based on congruent phylogenies and insect fossil dating, the antiquity of the infections is supported by levels of 16S rDNA sequence divergence within clades of endosymbionts.

The phylogenetic evidence for bacteriocyte associates of insects contrasts with that for some other endosymbioses. In Wolbachia pipiens, a bacterium that infects insect germ line cells and causes various reproductive abnormalities, molecular phylogenies give a clear indication of noncongruence, indicating horizontal movement of bacteria among lineages (Werren 1997,

---

**Table 1.** G + C composition of 16S rDNA of free-living and symbiotic bacteria.

<table>
<thead>
<tr>
<th>Taxon</th>
<th>G + C in 16S rDNA (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aeromonas trota</td>
<td>55.2</td>
</tr>
<tr>
<td>Alcaligenes sp.</td>
<td>54.9</td>
</tr>
<tr>
<td>Escherichia coli</td>
<td>54.4</td>
</tr>
<tr>
<td>Halomonas elongata</td>
<td>56.1</td>
</tr>
<tr>
<td>Haemophilus influenzae</td>
<td>52.2</td>
</tr>
<tr>
<td>Legionella sp.</td>
<td>53.1</td>
</tr>
<tr>
<td>Neisseria gonorrhoeae</td>
<td>55.0</td>
</tr>
<tr>
<td>Proteus vulgaris</td>
<td>52.8</td>
</tr>
<tr>
<td>Pseudomonas aeruginosa</td>
<td>54.1</td>
</tr>
<tr>
<td>Pseudomonas testosteroni</td>
<td>54.6</td>
</tr>
<tr>
<td>Vibrio parahaemolyticus</td>
<td>53.8</td>
</tr>
<tr>
<td>Yersinia pestis</td>
<td>54.1</td>
</tr>
<tr>
<td>Aphid primary symbiont (Buchnera aphidicola)</td>
<td>49.3</td>
</tr>
<tr>
<td>Ant primary symbiont (Candidatus componoti)</td>
<td>47.7</td>
</tr>
<tr>
<td>Tsetse primary symbiont (Wigglesworthia glossinidia)</td>
<td>47.9</td>
</tr>
<tr>
<td>Whitefly primary symbiont</td>
<td>47.8</td>
</tr>
<tr>
<td>Weevil symbiont</td>
<td>54.0</td>
</tr>
<tr>
<td>Mealybug symbiont</td>
<td>55.7</td>
</tr>
<tr>
<td>Pea aphid secondary symbiont</td>
<td>54.1</td>
</tr>
<tr>
<td>Whitefly secondary symbiont</td>
<td>54.4</td>
</tr>
</tbody>
</table>

What is the basis for this difference? That is, why might bacteriocyte associates of insects lack the capacity for horizontal transfer? When symbionts are maternally transmitted and beneficial to hosts, hosts are selected to provide a continually favorable environment to ensure the persistence of the association. Consequently, selection on bacteria for the ability to withstand conditions outside bacteriocytes will be relaxed. The resulting intolerance of conditions outside hosts could effectively eliminate horizontal transfer, facilitating long-term cospeciation of host and symbiont.

In the case of Wolbachia, by contrast, the typical infection route is maternal but the interaction appears to be more parasitic than mutualistic (Werren 1997, Bourtzis and O’Neill 1998). The bacteria do not reside in specialized bacteriocytes, and their maintenance and transmission do not appear to be facilitated by host adaptations. Rather, they inhabit a variety of cell types and invade host eggs using their own devices (Werren 1997). As a result, they may retain a more generalized ability to live under different conditions and thus have greater propensity for rare horizontal transmission. In some marine invertebrates, such as bivalves and marine annelids, the associations are

<table>
<thead>
<tr>
<th>Host clade</th>
<th>Minimum age of symbiotic association (i.e., age of common ancestor of hosts)</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aphids (Hemiptera: Aphidoidea)</td>
<td>150–250 million years</td>
<td>Munson et al. 1991, Moran et al. 1993</td>
</tr>
<tr>
<td>Cockroaches + termites (Isoptera + Blattaria)</td>
<td>135–300 million years</td>
<td>Bandi et al. 1995</td>
</tr>
<tr>
<td>Tsetse flies (Diptera: Glossinidae)</td>
<td>40 million years</td>
<td>Aksoy et al. 1995</td>
</tr>
</tbody>
</table>

Adaptations of endosymbiotic bacteria

Because bacteriocyte associates of insects cannot be cultured outside of hosts, it has been difficult to show how they differ from free-living bacteria. However, genetic characterization of Buchnera, the primary endosymbiont of aphids, has demonstrated some striking adaptations to endosymbiotic life (P. Baumann et al. 1997).

The best-documented contribution of aphid endosymbionts to their hosts is the provision of required amino acids that are rare or absent from the phloem sap diet. In particular, several lines of evidence support the role of Buchnera in provisioning hosts with tryptophan, an amino acid that is required by animals but is rare in phloem sap (Douglas and Prosser 1992, Lai et al. 1994). Tryptophan biosynthesis is regulated by feedback inhibition acting on anthranilate synthase, which is encoded by trpEG. In Buchnera that are endosymbiotic with members of the Aphididae, trpEG is excised from the chromosome and amplified on plasmids (Lai et al. 1994). This amplification, which involves both the tandem duplication of the operon on each plasmid and the presence of multiple plasmids, appears to function in the overproduction of tryptophan through enhanced production of the limiting biosynthetic enzyme. Comparisons of phylogenies based on plasmid-borne and chromosomal genes indicate that trpEG on plasmids originated from the chromosomal genes within the same bacterium and that the plasmids are transmitted strictly vertically (Rouhbaksh et al. 1996, 1997). The trpEG plasmid presumably was fixed through favorable selection at the host level; that is, host insects bearing endosymbionts with this feature experienced greater growth rates and became established at the expense of other aphid lineages of the same species in which endosymbionts possessed only a chromosomal trpEG copy.

A similar instance of plasmid amplification enabling the overproduction by Buchnera of an essential aphid nutrient has been discovered for leucine production (Bracho et al. 1995). The four genes that underlie production of leucine are excised from the chromosome and transferred to plasmids in some Buchnera species. However, only a single set of the leucine genes is present on each plasmid. The presence of the replicating unit of the plasmid appears to be ancestral in Buchnera of all aphids, but its acquisition of the leucine biosynthetic genes appears to have occurred independently in different lineages (van Ham et al. 1997).

Other unusual aspects of the Buchnera genome may represent adaptations to endosymbiotic life. Because symbionts are confined to host cells, their growth rates must be depressed and coordinated with the development of the host. For Buchnera, there is a tight coupling of bacterial cell number and aphid growth, with Buchnera showing a doubling time of approximately two days, a growth rate that is much lower than the maximum exhibited by many free-living bacteria (Baumann and Baumann 1994). Probably in connection with this low maximal growth rate, both Buchnera and Wigglesworthia, the bacteriocyte associate of tsetse flies, possess only
a single copy of the rRNA genes, in contrast to free-living bacteria, which typically have several copies (Aksoy 1995, Baumann et al. 1995). Reduction in the number of copies of rRNA genes in prokaryotes is associated with lowered maximum growth rates.

Endosymbiont adaptations for improved function as mutualists present the clear possibility of conflict between selection on individual cell lineages within a host aphid and selection on individual aphids. For example, the excess production of tryptophan must be beneficial to hosts but is probably detrimental to the bacterial cell lineage producing the nutrient. In related free-living bacteria, production of anthranilate synthase has been shown to be costly; indeed, it lowers fitness of bacteria when tryptophan is not limiting (Dykhuizen 1978), suggesting that a mutant Buchnera lacking the excess copies would enjoy a short-term advantage. However, host lineages in which all symbionts were lacking the excess copies would suffer a nutritional handicap and would thus be at a selective disadvantage within aphid populations. The high mutation rate for loss of tandem repeats in bacteria (10^{-4}–10^{-3} per generation; Roth et al. 1996), in combination with the immediate selective advantage for mutant cells, would be expected to destabilize these host-level adaptations. In the long term, however, selection at the host level would favor adaptations that stabilize beneficial gene amplification.

Some findings suggest that such host-level selection has resulted in reduced homologous recombination in Buchnera. Curiously, in certain Buchnera of the family Aphididae, reduction in plasmid-borne functional trpEG has occurred through the transformation of some of the repeats into pseudogenes (Lai et al. 1996, L. Baumann et al. 1997). Typically, however, such unneeded repeats would instead be eliminated through recombination. A possible explanation for the transformation into pseudogenes is selection on hosts to stabilize beneficial gene amplification, resulting in loss of some recombination capabilities (P. Baumann et al. 1997). If certain aphid lineages subsequently evolved feeding habits or life cycles that reduced their need for tryptophan provision by Buchnera, selection then might have favored reduced trpEG expression due to the energy expense of making unneeded anthranilate synthase. If recombinational mechanisms for eliminating repeat copies had been lost earlier, gene silencing would have had to occur through the fixation of stop codons and frameshifts in the DNA sequence, resulting in transformation of the functional genes into pseudogenes. If this scenario is correct, then the loss of certain recombination functions would appear to be an adaptation to stabilize mutualistic traits of the endosymbiont that might otherwise be eliminated through a combination of high mutation rates and selection at the bacterial cell level.

**Evolutionary consequences of endosymbiosis**

Endosymbiosis is sometimes considered to be an initial step in the evolution of organelles (e.g., Margulis 1993). However, organelles have arisen only a few times from free-living prokaryotes, either once or twice each for mitochondria and chloroplasts (Cavalier-Smith 1992, Gray 1992). By contrast, intracellular symbionts have arisen far more often, as a glimpse at Buchner's (1965) volume confirms. Thus, organelle characteristics are not an inevitable consequence of endosymbiosis combined with maternal transmission.

One feature of mitochondrial and chloroplast genomes is the loss of most genes present in the ancestral prokaryotic genome and the dependence on host gene products for many of the basic functions required for replication. The only endosymbiont for which substantial genome characterization has been carried out is Buchnera, for which 85 kb has been sequenced in the Baumann laboratory (summary in P. Baumann et al. 1997). Despite substantial overall sequence divergence from related free-living bacteria, Buchnera coding gene sequences indicate selective constraint of the functioning polypeptides; the reading frame is maintained and nucleotide base substitutions are concentrated at silent, or synonymous, sites (i.e., sites that do not affect amino acid identity). Genes with a wide range of "housekeeping" and biosynthetic functions have been identified in Buchnera (summaries in Baumann et al. 1995, P. Baumann et al. 1997, Clark and Baumann 1997, Clark et al. 1997). This work provides strong evidence that Buchnera contains a complement of genes enabling a wide range of processes involved in cell growth and reproduction. This endosymbiotic association arose at least 200 million years ago, so the retention of most genes is not readily explained as the result of the association being too young for Buchnera to have developed characteristics of organelles. A more important limitation on the loss of genes might be that endosymbionts in animals reside primarily in somatic cells (the bacteriocytes), eliminating most opportunities for gene transfer between symbiont and host nuclear genome.

One apparent long-term consequence of endosymbiosis is increased rates of substitution in DNA sequences of endosymbionts relative to non-endosymbiotic relatives (Moran 1996). This increase is seen in the 16S rDNA of endosymbionts of aphids, whiteflies, mealybugs, tsetse flies, and carpenter ants (Moran 1996, Lambert and Moran in press). In Buchnera, in which many more genes have been sequenced than in any other endosymbiont, the increased rate extends to all genes examined (Moran 1996, unpublished analyses).

Positive natural selection for base substitutions seems unlikely to explain the faster evolution at all loci, because the effects of positive selection are expected to be locus and site specific within genes. In fact, it appears that the base substitutions are deleterious. Evidence for this hypothesis is provided by the changes in protein-coding genes in Buchnera. The net effect of the faster evolution in Buchnera is the accumulation, throughout the genome, of A and T at the expense of representation of G and C. This shift in nucleotide base composition is strong enough that most amino acid differences between polypeptides of Buchnera and those of E. coli are ones that allow increased A + T in the corresponding codon and thus in the DNA sequence.
It is hard to imagine that the accumulation, in all polypeptides, of amino acids with $A + T$-rich codon families is adaptive at the level of protein function. A more plausible explanation is mutational bias favoring $A + T$ combined with an increased rate of base substitution within lineages.

If positive selection is excluded as the basis for the faster substitution rates of endosymbionts, we are left with two other categories of explanation. One is that the increased rate of substitution (fixed changes occurring in a lineage) is due to an increased rate of new mutations. This increased mutation rate could result from either an increased rate per generation or a faster average generation time. But if increased incidence of mutation were the sole explanation, the increase would have equivalent effects on substitution rates at sites that are subject to selection and at sites that are effectively neutral with respect to selection, and comparative sequence analysis indicates stronger effects on selected sites. A convenient way to categorize sites according to the intensity of selection is to separate silent (or synonymous) from replacement (or nonsynonymous) sites in protein coding sequences. At silent sites, where nucleotide changes do not affect the corresponding amino acid and thus do not affect the gene product, base substitutions have little consequence for fitness. At replacement sites, where a nucleotide change results in a change in the amino acid and thus in the selected phenotype, base substitutions are more likely to affect fitness. The increase in substitution rate in coding genes of Buchnera is heavily concentrated at replacement sites (Moran 1996). This observation rules out increased mutation as the primary basis for the acceleration in substitution rates.

This result instead supports the second explanation, that the excess changes represent deleterious substitutions. According to this view, purifying selection is less effective at purging deleterious mutations. This reduction in the effects of selection could arise in either of two ways: from relaxed selection or from lower effectiveness of selection as a result of population structure. The first possibility, that selection is uniformly relaxed at all genes of Buchnera, regardless of function, is doubtful. Because endosymbionts show reduced effective population sizes than free-living bacteria, with a consequently greater susceptibility to genetic drift, the second possibility may apply. This proposal resembles the arguments made by Lynch (1996, 1997) for the accumulation of deleterious mutations in mitochondria and other organelles. Organelles have similar population structure to endosymbionts because they are asexual, intracellular, and maternally inherited.

**Current mysteries and future problems**

The recent molecular studies of insect bacteriocyte-associated endosymbioses have given glimpses into the evolution of some of these associations. But most remain entirely unstudied, particularly the complex multiple infections found in some insects, such as the planthoppers and treehoppers (Hemiptera; Buchner 1965). Furthermore, this work has raised some new questions about the evolution of bacteriocyte associates in insects. For example, both Buchnera and tsetse endosymbionts exhibit unusually high levels of GroEL protein (Ishikawa 1989, Aksoy 1995, Baumann et al. 1996), a chaperonin that normally functions in refolding other polypeptides. The reason for this overexpression is not yet clear, although speculative explanations have been put forward (Ishikawa 1989, Moran 1996, P. Baumann et al. 1997). The finding of plasmid-borne pseudogenes in some Buchnera presents another mystery: Why are nonfunctional copies not lost by recombination, as is typical for bacteria? Some basic difference involving recombination pathways may, as already discussed, distinguish Buchnera from free-living bacteria. Whether other endosymbionts also have atypical recombination pathways is not yet known. The basis for the nucleotide base compositional bias against G + C, a distinctive feature of numerous bacteriocyte associates (Table 1), is also unknown.

Endosymbiotic interactions are shaped by forces acting at multiple levels, primarily that of the host individual and the bacterial cell lineage. Some mutations in endosymbionts will favor the individual symbiont lineage at the expense of the host. Such "selfish" mutations would include any that result in higher replication of the symbiont at the host's expense. If higher replication results in disproportionate transmission to progeny, these selfish symbionts could increase in frequency within a host population, at least in the short term. Michod (1997) presents a theoretical framework for evaluating this very situation that characterizes bacteriocyte associations in insects: a balance between frequent mutation in the direction of "selfish" traits favoring individual cell lineages and selection on the overall group (in this case, the host individual and its resident bacterial population). Buchner (1965) mentions cases in whiteflies in which symbionts spread beyond their usual limits within individual host insects, indicating the possibility of such selfish behavior that harms the host.

Of course, selection favoring better-nourished (and faster-growing) host individuals will counter the spread of selfish bacterial lineages. One host adaptation that would limit the spread of selfish symbionts would be the sequestration of "germ line" symbionts from more rapidly dividing "somatic" symbionts that function in provisioning hosts with nutrients. Early microscopical studies suggest this kind of system in cicadas (Koch 1967). The infection of progeny with entire maternal bacteriocytes, as in whiteflies (Buchner 1965, Costa et al. 1996), would also have the effect of limiting the spread of "selfish" bacteria to single host progeny, provided that bacteria are confined to single bacteriocytes early in the development of the host. Frank (1996) proposes that the segregation of germ line symbionts may be a "policing" mechanism for the containment of selfish cell lineages within hosts of endosymbions.

**Role of endosymbiosis in host evolution**

Endosymbiosis has been argued to be a route for evolutionary innovation that underlies the origin and diversification of many groups of
organisms that would otherwise not exist. Within the realm of the insects, this hypothesis appears to be true: Essentially all strict phloem-feeders, strict blood-feeders, and other groups with restricted diets possess endosymbionts. Recent molecular phylogenetic evidence indicates that endosymbiotic infections date to the origins of a variety of major insect clades, reinforcing the idea that the origin of endosymbiosis was an integral factor spurring diversification of some insect taxa. Because some of these groups are large adaptive radiations, endosymbiosis has probably had a massive effect on patterns of insect diversity.

Although endosymbiosis appears to facilitate diversification within ecological niches that would otherwise be inadequate, dependence on an endosymbiont may simultaneously enforce some restrictions on host evolution. Genes encoded within the endosymbiont are subject to a different set of population genetic parameters than genes encoded in the insect nuclear genome. As noted above, endosymbionts are subject to within-host selection, which often may counter selection between hosts. In addition, mutation rates and generation times differ between endosymbiont and host genomes. The distinct population genetics of endosymbiont genes may affect the speed of the response to selection. For example, adaptation may occur faster in symbiont genes than host genes due to the shorter generation time and consequent greater supply of mutations of the bacteria. Finally, symbiont genes are subject to a population structure in which genetic drift is more likely to limit the effectiveness of selection, resulting in increased fixation of mildly deleterious mutations. This increased rate of accumulation of deleterious mutations will be countered, to some extent, by selection at the level of host individuals. The adequacy of this counterbalancing will depend on the impact of selection at the host level, which is dependent in turn on the effective population size of the host.

It is not yet clear what factors are most important in determining the net effect of these conflicting evolutionary forces. One intriguing possibility is that the accumulation of deleterious mutations would over-run purifying selection at the host level, except in cases in which the host population size is large. This possibility may help to explain why maternally transmitted endosymbionts are rare in animals such as vertebrates, which have small populations, as compared with insects, many of which have unusually large populations.

Acknowledgments

Research on insect endosymbionts in the Moran lab is supported by grants from the National Science Foundation (DEB-9527635) and the US Department of Agriculture, National Research Initiative (96-35302-3817). Aparna Telang was supported by a fellowship from the Plant-Insect Interactions Training Grant to the University of Arizona from NSF, USDA, and the US Department of Energy. We thank Paul Baumann for comments on an early draft of this review and Leah Kenaga and David Bentley of the Imaging Facility (Arizona Research Laboratories) for assistance with electron microscopy.

References cited


