Numerous Transposed Sequences of Mitochondrial Cytochrome Oxidase I–II in Aphids of the Genus *Sitobion* (Hemiptera: Aphididae)

Paul Sunnucks and Dinah F. Hales
School of Biological Sciences, Macquarie University, Sydney

Polymerase chain reaction (PCR) products corresponding to 803 bp of the cytochrome oxidase subunits I and II region of mitochondrial DNA (mtDNA COI–II) were deduced to consist of multiple haplotypes in three *Sitobion* species. We investigated the molecular basis of these observations. PCR products were cloned, and six clones from one individual per species were sequenced. In each individual, one sequence was found commonly, but also two or three divergent sequences were seen. The divergent sequences were shown to be nonmitochondrial by sequencing from purified mtDNA and Southern blotting experiments.

All seven nonmitochondrial clones sequenced to completion were unique. Nonmitochondrial sequences have a high proportion of unique sites, and very few characters are shared between nonmitochondrial clones to the exclusion of mtDNA. From these data, we infer that fragments of mtDNA have been transposed separately (probably into aphid chromosomes), at a frequency only known to be equalled in humans. The transposition phenomenon appears to occur infrequently or not at all in closely related genera and other aphids investigated.

Patterns of nucleotide substitution in mtDNA inferred over a parsimony tree are very different from those in transposed sequences. Compared with mtDNA, nonmitochondrial sequences have less codon position bias, more even exchanges between A, G, C and T, and a higher proportion of nonsynonymous replacements. Although these data are consistent with the transposed sequences being under less constraint than mtDNA, changes in the nonmitochondrial sequences are not random: there remains significant position bias, and probable excesses of synonymous replacements and of conservative inferred amino acid replacements. We conclude that a proportion of the inferred change in the nonmitochondrial sequences occurred before transposition.

We believe that *Sitobion* aphids (and other species exhibiting mtDNA transposition) may be important for studying the molecular evolution of mtDNA and pseudogenes. However, our data highlight the need to establish the true evolutionary relationships between sequences in comparative investigations.

Downloaded from https://academic.oup.com/mbe/article/13/3/510/1053533 by guest on 16 August 2022

Introduction

Mitochondrial DNA (mtDNA) does not encode all of the proteins needed for organellar function: nuclear gene products are transported into the mitochondrion where they contribute to biochemical processes (Hartl and Neupert 1990). The serial endosymbiosis theory, now widely accepted, holds that mitochondria arose from symbiotic bacterial ancestors (Margulis 1970). Assuming this to be correct, function has transferred from mitochondrial to nuclear genomes, and movement of mtDNA sequences must be important in the evolution of the relationship between the two genomes. Several studies have shown transfer of DNA between mitochondrial and nuclear genomes in diverse organisms including plants, fungi, insects, sea urchins, birds, rodents, felids and humans (Gellisen et al. 1983; Jacobs et al. 1983; Tsuzuki et al. 1983; Zullo et al. 1991; Smith, Thomas, and Patton 1992; Lopez et al. 1994 and references). Transpositions may result in mtDNA-like sequences being dispersed through the nuclear genome (Gellisen et al. 1983; Zullo et al. 1991), or in a tandemly repeated array (Lopez et al. 1994). There are examples of movements of protein-coding, RNA-coding, and control regions of mtDNA. Sequence data have implicated various kinds of sequences in facilitating recombination, including long interspersed repetitive elements (LINEs), highly repeated inverted flanking sequences, or CA repeats. The only reported case of mtDNA transposition in insects is of two separate moderately repeated sequences showing homology to mitochondrial rRNA in *Locusta migratoria* (Gellisen et al. 1983).

Multiple mtDNA-like sequences have been shown to have nuclear and mtDNA locations by probing Southern blots of different purified DNA fractions, in situ hybridization of chromosomes, and sequencing through sites of insertion (Lopez et al. [1994] used all of these). Where sequence data are available, patterns of nucleotide substitution show that most differences between transposed and current mtDNA are due to the continuing evolution of the mtDNA. Typically, nuclear copies must be nonfunctional because of differences between mtDNA and nuclear genetic processes and the presence of frame-shifting insertions or deletions (Smith, Thomas, and Patton 1992; Lopez et al. 1994). Thus, copies of mtDNA transposed into nuclear genomes are expected to evolve as nuclear pseudogenes. Because patterns of mtDNA evolution are so different

Key words: mitochondrial DNA, cytochrome oxidase, transposition, transposed sequence, nuclear, phylogeny, *Sitobion*, aphid.

Address for correspondence and reprints: Paul Sunnucks, School of Biological Sciences, Macquarie University, Sydney NSW 2109, Australia. E-mail: psunnuck@rna.bio.mq.edu.au.

Mol. Biol. Evol. 13(3):510–524. 1996
© 1996 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

Table 1
Geographical Sources and Identity of Aphids

Species/Line	Collection Site	Host Plant	Date Collected
Species and lines from which PCR products were cloned			
S. miscanthi (2n = 18) 80 ..	Sydney, NSW	*Setaria* sp.	30.05.89
S. near *fragariae* 17 (1[a]) ...	Tasmania		14.10.86
S. "Smilax" 194	Sydney, NSW	*Smilax* sp.	10.06.94
Genomic sequence or Southern blotting			
S. miscanthi (2n = 18) 4 (26[a])	Sydney, NSW	*Paspalum dilatatum*	12.12.86
S. near *fragariae* SA2 (64[a])	Adelaide, SA		00.10.88
S. avenae 7	England		
S. fragariae 5	England		
S. rubiphila 94043	Utsunomiya, Japan	*Microstegium vimineum*	14.10.94
Macrosiphum euphorbiae 1	Sydney, NSW	cultivated *Helipterum* sp.	27.07.94
M. rosae Green 1	Sydney, NSW	cultivated roses	27.06.94
Metopolophium dirhodum 1	Sydney, NSW	cultivated wheat	10.01.95
Illinoia azaleae 1	Canberra, ACT	cultivated azaleas	10.10.94
Myzus ornatus 1	Sydney, NSW	cultivated mint	22.11.94

[a] Identity number given to an aphid line in Hales et al. (1990).

from those of nuclear pseudogenes, there arises the possibility of estimating evolutionary patterns of nuclear pseudogenes and functional mtDNA in very similar sequences. We are not aware that such comparisons have been proposed previously.

The region of mtDNA containing cytochrome oxidase I and II (COI–II) has been used in systematic and population genetic studies of insects, including aphids (Beckenbach, Wei, and Liu 1993; Brown et al. 1994; Sperling and Hickey 1994; Stern 1994). We investigated 803 base pairs (bp) of mtDNA sequence spanning the 3′ end of cytochrome oxidase subunit I (COI), $tRNA^{leu}$ (UUR), and most of cytochrome oxidase subunit II (COII) as a phylogenetic tool in chromosomal races and closely related species of aphids in the genus *Sitobion* (Hemiptera: Aphididae). We studied the molecular basis of initial observations that PCR products from individual aphids apparently contained more than one COI–II haplotype. To this end, we report sequence of PCR products and cloned single haplotypes in *Sitobion* aphids, and compare sequence obtained from outgroup genera. Our data indicate the presence of *many* separate transposed mtDNA sequences containing the mtDNA COI–II region in *Sitobion* species and some other aphids. We examine the significance and possible applications of this extraordinary transposition phenomenon.

Material and Methods

Aphids

We obtained DNA sequence using PCR primers designed from COI and COII of mtDNA, from six *Sitobion* aphid taxa: (1) *S. miscanthi* (Takahashi), (2) a species like but distinct from *S. fragariae* (Walker) and referred to as *S.* near *fragariae* (Hales et al. 1990), (3) an undescribed species feeding on *Smilax glyciphylla* referred to here as *Sitobion* "Smilax," (4) *S. avenae* (F.), (5) *S. fragariae,* and (6) *S. rubiphila* (Takahashi). We have used the term "lines" throughout this paper to describe parthenogenetic lineages derived from single aphids, rather than the usual term "clone," which we reserve for molecular clones. We have in culture a number of lines of *S. miscanthi,* and *S.* near *fragariae.* Data are reported from lines specified in table 1. We chose as outgroups two members of the closely related genus *Macrosiphum, M. rosae* (L.), and *M. euphorbiae* (Thomas), as well as *Metopolophium dirhodum* (Walker), *Illinoia azaleae* (Mason), and the more distantly related *Myzus ornatus* Laing. Aphids were free of hymenopteran or dipteran endoparasites.

DNA Extractions

DNA from all tissues within individual aphids ("whole-aphid DNA") was extracted by a "salting-out" protocol: individual wingless adults were crushed in a 1.5-mL microfuge tube and incubated at 37°C in 300 μL TNES (50 mM Tris, pH 7.5, 400 mM NaCl, 20 mM EDTA, 0.5% SDS) with 100 μg/mL Proteinase K. After 3–18 h, proteins were precipitated with 85 μL 5M NaCl and hard shaking for 15 sec. Proteins were pelleted in a microfuge at 14,000 rpm for 5 min. then DNA was precipitated from the decanted supernatant with 1 volume 100% ethanol. DNA was pelleted, washed in 70% ethanol, air-dried, and dissolved in 20–50 μL sterile water in rough proportion to the size of the aphid. DNA diluted 1:10 in sterile water was used as the template DNA in polymerase chain reactions (PCR). For samples received in ethanol, excess ethanol was allowed to evaporate and

Downloaded from https://academic.oup.com/mbe/article/13/3/500/1003331 by guest on 19 April 2024

the procedure above was followed, except that, because of lower DNA yield, samples were not diluted for PCR.

All female aphids except some "soldiers" have endosymbiotic bacteria (see Moran et al. 1993 and references). These are localized in the abdomen, especially in young adult winged individuals in which the thorax is packed with flight muscle. To obtain aphid DNA with low amounts of endosymbiotic bacteria ("low-bacterial" DNA), 20 thoraces of young adult winged *S. miscanthi* line 4 were treated with the same salting-out protocol as above.

Mitochondrial DNA was extracted by alkaline lysis (Tamura and Aotsuka 1988), with two modifications to increase purity of the DNA. First, extra rounds of centrifugation were carried out until the cell and organelle suspension was clear. Second, after pelleting, mitochondria were washed in "buffer A," resuspended, and recentrifuged. Mitochondrial DNA from about 200 mg of aphids was dissolved in 50 μL of TE (10 mM Tris, 1 mM EDTA). One microliter of a 1:100 dilution was used as PCR template. Five microliters of the 50 μL was used for each Southern-blotting lane.

DNA enriched for aphid chromosomal DNA relative to mtDNA ("chromosomal-enriched DNA") was obtained by keeping the upper band of a standard caesium chloride purification of mtDNA (similar to Powers et al. 1989) from 1 g of *S.* near *fragariae* 17 and 1 g of *S. miscanthi* 80.

PCR Protocols and Sequencing

We designed a PCR primer from cloned *S. miscanthi* mtDNA (unpublished data) in the COI region: 5′-GGAATACCTCGACGATATACAG-3′ (this would be called C1-J-2791 by the convention of Simon et al. [1994]). Coincidentally, this proved to be similar to primer mt2793+ used in the only previously published study of this region in aphids (Stern 1994). The second primer was a modification by one nucleotide of primer COII-Croz (Roehrdanz 1993): 5′-CCGCAAATTTCTGAACATTGACC-3′ (C2-N-3662), again similar to that of Stern (1994). PCR products were sequenced with these primers, and also with mt2993+ (Stern 1994; 5′-CATTCATATTCAGAATTACC-3′: C1-J-2993), plus a primer ("COf") designed from *D. yakuba* and *A. mellifera* sequence: 5′-CTGAATATTCATAACTTCAGTATCATTGATGACC-3′ (C2-N-3382). For some clones, COf was not very effective, and 3175+ was used instead (Stern 1994; 5′-CATGA-C/T-CATACAATTTTTATTAT-3′). Binding sites of these primers and the structure of the region sequenced are shown in figure 1. Mitochondrial cytochrome *b* PCR product of about 400 bp was amplified using primers cb1 and cb2 (Simon et al. 1994).

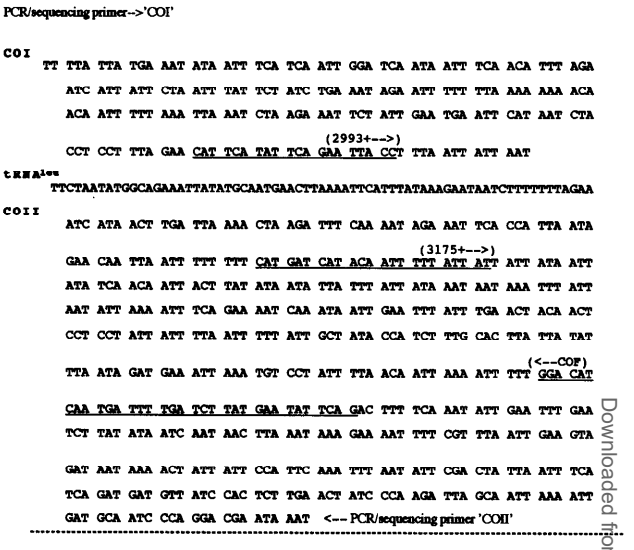


FIG. 1.—The nucleotide sequence of clone Sm80.4 (mtDNA COI–II). Locations of protein genes and $tRNA^{leu}$ are given, and protein-coding regions are shown as codons. Annealing sites of internal sequencing primers are underlined, and orientations of external primers are given.

PCR was carried out using Promega *Taq* polymerase (1 U) with Mg^{2+}-free reaction buffer, 1.5 mM $MgCl_2$, 200 μM each dNTP, 5–10 pm primers, 5–10 ng (1 μL as prepared above) cellular DNA, using the following program: 94°C for 5 min. (94°C 15 sec, 52°C 30 sec, 72°C 45 sec) 40 times, 72°C for 5 minutes. The whole 25 μL product was electrophoresed on a 1.5% agarose gel, and a band of about 900 bp was excised.

Double-stranded DNA was recovered with GeneClean (Bio 101) and eluted into 20 μL sterile water. Double-stranded sequencing was carried out with 4 μL of this DNA, using Sequenase version 2.0 (United States Biochemical), modified to use dimethyl sulphoxide (DMSO) as denaturant (Winship 1989). Reactions were run on 6% acrylamide with 90 mL 1 M sodium acetate in the anodal buffer (to 1.5l total) to improve separation of high molecular weight DNA. Gels were dried unfixed onto 3MM paper (Whatman), and exposed for 2–3 days on Hyperfilm (Amersham).

Cloning

We have found consistently that aphid mtDNA separated on standard agarose gels and purified with GeneClean cannot be cloned efficiently. Thus COI–II PCR products for cloning were separated on 2% low-melting-point agarose gels and stained with ethidium bromide. A band of approximately 900 bp was excised with minimal exposure to ultraviolet light. DNA was recovered by standard phenol extraction and ethanol precipitation and dissolved in TE for storage at 4°C (Sambrook, Fritsch, and Maniatis 1989). DNA was ligated into pGEM-T vector (Promega) following manu-

Downloaded from https://academic.oup.com/mbe/article/13/3/537/1053232 by guest on 19 April 2024

facturers' instructions, and transformed into DH5α electrocompetent cells (Sambrook, Fritsch, and Maniatis 1989, and Bio-Rad Gene Pulser protocol) before plating on standard color selection plates (Sambrook, Fritsch, and Maniatis 1989). Random recombinant colonies were investigated by PCR using primers COI and COII. To obtain DNA, a small piece of an individual colony was picked into 50 μL of TE and boiled for 3 min, then 3 μL of this was used as template in a PCR reaction under the conditions above. Products were processed for sequencing as above.

Southern Blotting

Samples were digested with *Hin*dIII, run for 40 h at 40 V on 0.8% agarose so that DNA of 1.5 kb ran about 18 cm. Southern blotting was carried out by standard capillary methods (Sambrook, Fritsch, and Maniatis 1989) onto Hybond-N (Amersham). Membranes were prehybridized for at least 3 h at 65°C in 1 mM EDTA, 0.25 M Na_2HPO_4 (pH 7.2), 7% SDS, and 1% w/v BSA, and hybridized in the same solution overnight with PCR product labeled with (^{32}P) αdATP using a random priming kit (Boehringer). Membranes were washed for 15 min. at 65°C in 0.26 M phosphate/1% SDS and then 2× SSC/0.1% SDS (Bruford et al. 1992).

Sequence Handling, Phylogenetics and Analysis of Rates of Evolution

Sequences were read manually, aligned by eye, and read into MacClade 3.01 for Macintosh (Maddison and Maddison 1992), for export to PAUP 3.1.1 (Swofford 1993). Support for groupings on the phylogenetic tree presented was investigated by carrying out 500 branch-and-bound bootstrap resampling replicates with the bootstrapping option in PAUP. Genetic distances are not corrected for multiple hits because of the low overall sequence divergence; the effect should be trivial underestimation of genetic distances.

We used the method of Gojobori, Li, and Graur (1982) to calculate the relative substitution frequencies (RSF) of the four types of nucleotide, inferred over an outgroup-rooted parsimony tree. This method accounts for the different proportions of nucleotides A, G, C, and T available for change, so that the rates of change can be compared directly. For each class of change from the *i*th type of nucleotide to the *j*th, the proportion of $i \rightarrow j$ substitutions in each sequence is calculated as P_{ij}: the observed number of $i \rightarrow j$ divided by the number of bases available for change in the inferred ancestral sequence. The RSF (F_{ij}) is the proportion of all the change in the sequence represented by $i \rightarrow j$, calculated as:

$$F_{ij} = P_{ij}/(\Sigma_i \Sigma_{j \neq i} P_{ij}).$$

We have tabulated F_{ij} as a percentage of all change in the sequence (Gojobori, Li, and Graur 1982). When combining data for different sequences, the mean is weighted in relation to the proportion of substitutions in each sequence.

To estimate expected ratios of synonymous/nonsynonymous changes, and proportions of amino acid substitutions of different degrees of chemical conservativeness, expected proportions in mammalian pseudogenes were used as an approximation (values in table 8 in Li, Wu, and Luo 1984). The number of substitutions at each codon position expected to cause amino acid replacement were approximated using proportions expected under random mutation in random protein-coding sequence: first position = 96% replacement, second position = 100%, and third position = 31% (Li and Graur 1991). The likely validity of these estimates is considered with respect to differences between mammalian and insect genomes.

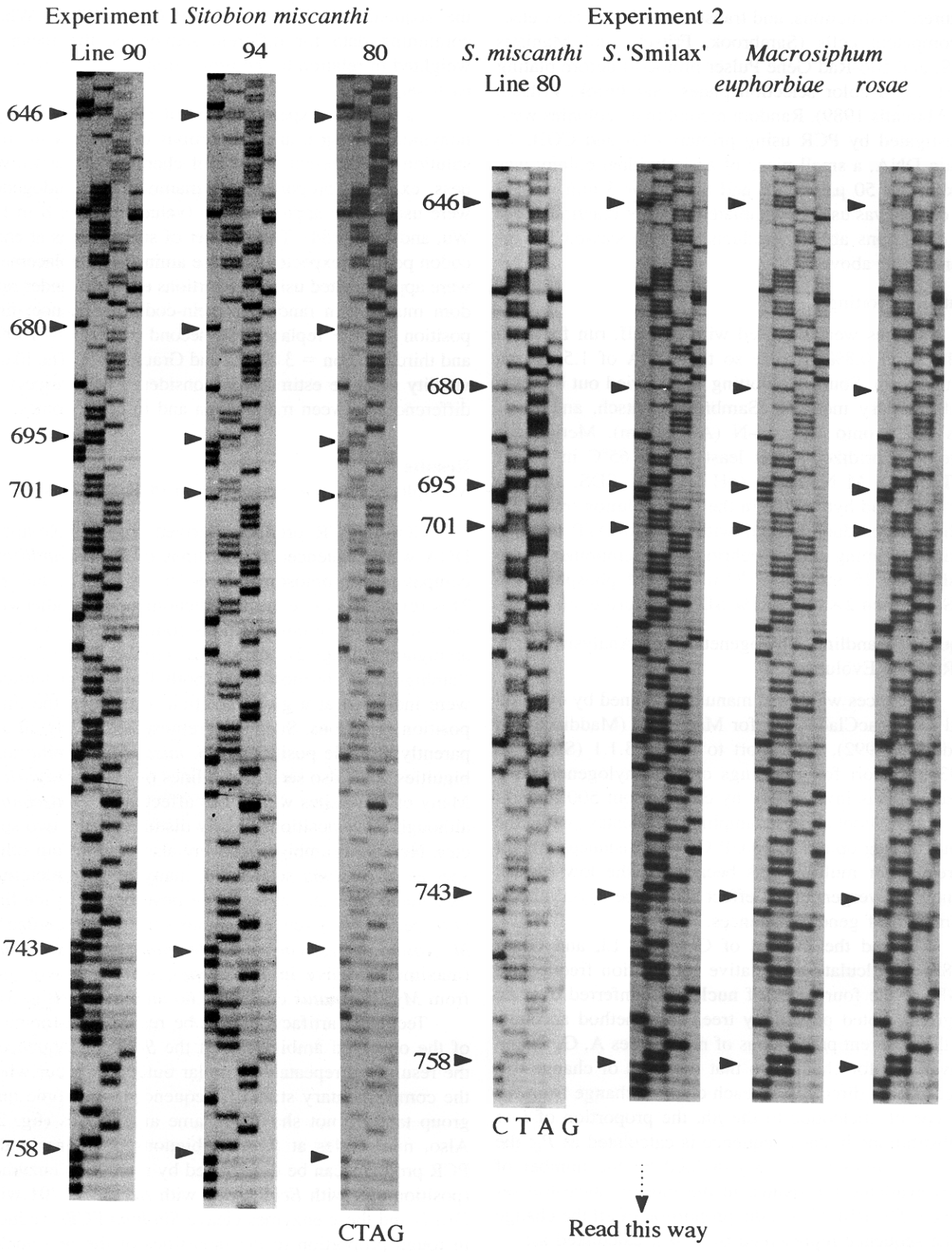
Results

Multiple Nucleotide Sequences in PCR Products

COI–II PCR products derived from whole-aphid DNA were sequenced for 13 lines of *S. miscanthi* encompassing chromosomal races 2n = 17, 18, and 20. This revealed an unexpected phenomenon. In otherwise clear sequence, certain positions could not be scored unambiguously (fig. 2): such sites will be referred to as "ambiguities." In most cases, both T and C nucleotides were indicated at a given position, usually at the third position of codons. Such sites represented nearly all apparently variable positions in *S. miscanthi.* Similar ambiguities were also seen in two lines of *S.* near *fragariae.* Many of these sites were those affected in *S. miscanthi,* although other positions clearly distinguish the two species. Numerous ambiguities were also seen in four other species of *Sitobion* sequenced; many of the indeterminate nucleotides are at the same positions as in the first two species. Very few ambiguous sites were recorded in *M. rosae, I. azaleae* and *Metopolophium dirhodum* (maximum of five in *I. azaleae*), while the sequence from *Myzus ornatus* contained no ambiguities (fig. 3).

Technical artifacts cannot be responsible for most of the observed ambiguities in the *Sitobion* sequences: the results are repeatable, similar outcomes occur when the complementary strand is sequenced, and some outgroup taxa do not show the same ambiguities (fig. 2). Also, nucleotides at three ambiguous positions in the PCR products can be determined by restriction enzymes (position 458 with *Eco*RI, 680 with *Acc* I and 701 with *Dra* I) and these enzymes cleave *Sitobion* PCR products in rough proportion to the intensities of the two nucleotides at ambiguous positions (data not shown). We hypothesize that PCR products from a single individual contained two or more sequences similar to COI–II.

Downloaded from https://academic.oup.com/mbe/article/13/3/509/1031231 by guest on 16 August 2022



Downloaded from https://academic.oup.com/mbe/article/13/3/510/1057321 by guest on 19 April 2024

FIG. 2.—Photographs of COI–II sequencing autoradiographs showing ambiguous sites. Sequence was obtained from PCR products derived from whole-aphid DNA. Six positions are indicated with arrows which show evidence of more than one haplotype within some individuals, i.e., they are "ambiguous sites." None of the sites is similarly affected in the *Macrosiphum* outgroups. Note that the ambiguities are consistent between the two separate experiments for *S. miscanthi* line 80. The sequences must be read in the direction indicated to match the presentation in later tables and figures.

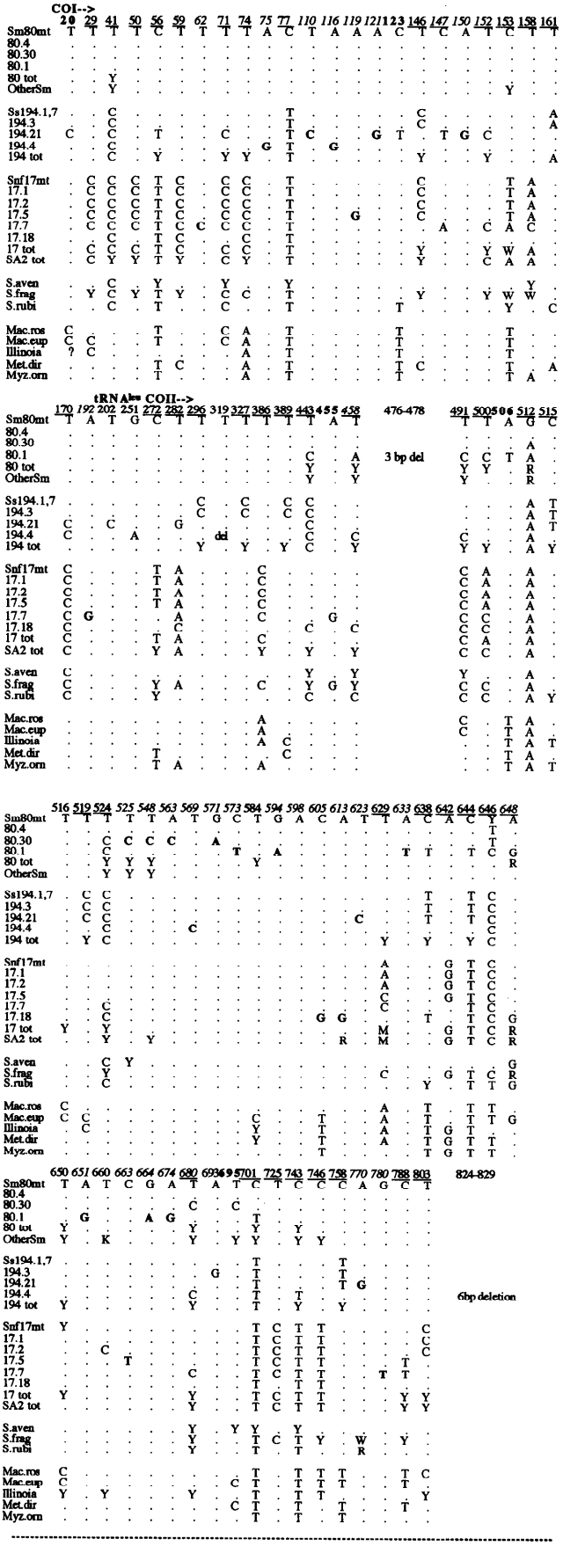


Table 2
Details of Cloned COI–II PCR Products Sequenced only Partially

Species/ Clone	Nucleotides Sequenced		Outcome
	Identity	Number	
S. miscanthi			
Sm80.13	316–803	488	Identical to Sm80.4
Sm80.14	316–803	488	Differs by G at 445
Sm80.35	534–803	270	Identical to Sm80.4
S. "Smilax"			
Ss194.18	556–803	248	Identical to 194.1
Ss194.17	605–803	199	Identical to 194.1
S. near *fragariae*			
Snf17.19	573–803	231	Identical to 17.1

COI–II PCR Products from Individuals are Heterogeneous in Sequence

A parthenogenetic line of each of three species was chosen: *S. miscanthi* (line 80), *S.* near *fragariae* (line 17), and *S.* "Smilax" (line 194) (table 1 for details). For each of these, COI–II PCR products from a single individual were cloned. Sequence was obtained for three clones from the *S. miscanthi* 80 individual, five from the *S.* near *fragariae* 17, and five from the *S.* "Smilax" 194. In addition, more clones were sequenced in part (table 2), to a total of six per individual. Clones are encoded by a species code + an aphid line code + a clone code. For example, clone 1 from the *S. miscanthi* 80 aphid is encoded Sm80.1. The code for *S.* "Smilax" is Ss and for *S.* near *fragariae,* Snf. Clones to be sequenced were taken at random, with the exception of Sm80.30, Ss194.4, and Snf17.7, which were chosen on the basis of having a C at position 680, as determined by restriction digestion.

In each individual, one sequence was found multiple times (or differed at only one position, possibly because of *Taq* polymerase misincorporation, analyzed later), but in addition there were two or three unique divergent sequences (fig. 3, table 2). For *S. miscanthi* 80 and *S.* near *fragariae* 17, a true mtDNA sequence

←

FIG. 3.—Nucleotide states at all positions which are variable in cloned COI–II sequences. Cloned haplotypes are identified by species code (Sm = *S. miscanthi,* Ss = *S.* "Smilax," Snf = *S.* near *fragariae*) + line identity + clone number. Sequence from whole-aphid DNA PCR products from these species is suffixed "tot," and from mtDNA-only PCR products "mt." "OtherSm" refers to a panel of 12 diverse lines of *S. miscanthi* which were also sequenced from whole-aphid DNA PCR products. The various classes of site mentioned in the text are specified: underline = mtDNA-like site, italics (and nucleotide in bold) = *unique site,* underlined italic = *multiple-hit site,* and bold = **outgroup site**. Other species abbreviations are obvious with reference to table 1. Nucleotide codes: Y = C + T, R = A + G, W = A + T, M = A + C.

Downloaded from https://academic.oup.com/mbe/article/3/6/515/1053335 by guest on 16 August 2022

Table 3
Pairwise Distances between Cloned Sequences

	S. miscanthi (Sm)			*S.* "Smilax" (Ss)			*S.* near *fragariae* (Snf)			
	80.4	80.30	80.1	194.7	194.21	194.4	17.1	17.5	17.7	17.18
Sm										
80.4	—	1.0	2.1	2.1	3.1	2.1	3.4	3.5	4.0	3.2
80.30	8	—	2.6	2.6	3.6	2.4	4.1	4.2	4.2	3.7
80.1	17	21	—	2.5	3.5	2.6	4.2	4.3	4.4	2.6
Ss										
194.7	17	21	20	—	2.2	2.5	3.7	3.9	4.4	3.0
194.21 ..	25	29	28	18	—	3.2	4.1	4.2	4.1	3.0
194.4	17	19	21	20	26	—	3.5	3.6	3.6	2.5
Snf										
17.1	27	33	33	30	33	28	—	0.6	2.2	2.9
17.5	28	34	34	31	34	29	5	—	2.1	3.2
17.7	32	34	35	35	33	29	18	17	—	3.0
17.18	23	27	19	22	22	18	21	23	22	—

NOTE.—Below the diagonal are absolute distances: above it are uncorrected % divergences adjusted for missing data. Codes of presumed mtDNA haplotypes are underlined.

was determined from PCR products amplified from purified mtDNA. In both cases this corresponds to the sequence found multiple times in single specimens. MtDNA was not available for *S.* "Smilax" 194, but by analogy the common sequence (seen in five clones including Ss194.7) is assumed to be mtDNA in origin. The mtDNA-derived sequences are confirmed as mtDNA by comparison with aphid sequence (Stern 1994), and because they appear to encode functional products when translated with the *Drosophila* mtDNA code. All sequences reported here have an inferred phenylalanine at the fifth amino acid position in the highly conserved electron-binding region of COII. This is shared with sequences in Stern (1994), but differs from species in 10 orders of insects including another hemipteran (Liu and Beckenbach 1992). Thus the presence of this residue is good evidence that a sequence is derived from aphid mtDNA and not a contaminant.

Patterns of Evolution in mtDNA

Uncorrected sequence divergence between mtDNA COI–II in the *S. miscanthi* 80 individual, the *S.* near *fragariae* 17, and the *S.* "Smilax" 194 is 0.021 to 0.037 (table 3). Patterns of difference are typical of those observed in other closely related insect COI–II mtDNAs, including high transition/transversion ratios and a predominance of third-position substitutions (Beckenbach, Wei, and Liu 1993; Brown et al. 1994). Of 38 positions variable between the mtDNA of the three *Sitobion* species, 13.2% are first positions in codons, 2.6% second, and 84.2% third. Thirteen are in COI (6.5% of COI positions), 25 in COII (4.7%), and none in $tRNA^{leu}$.

Estimate of Maximum Polymerase Incorporation Error

Three clones differed by only one nucleotide from mtDNA sequence in the same individual (Sm80.14 was G at 445, Ss194.3 was G at 693, and Snf17.2 was C at 660). If these sequences are from functional mtDNA (and not nonmitochondrial sequences) we believe that the three unique states may represent *Taq* polymerase misincorporation, because all three substitutions would result in amino acid replacements which are very rare in mtDNA sequences (data below), in 2/3 cases at residues conserved between all our aphid mtDNA sequences and *Drosophila* (Liu and Beckenbach 1992). If these changes are indeed from polymerase misincorporation, our detected error rate is 3/6742 (bases sequenced in clones thought to be mtDNA) = 0.45 per 1,000 = 0.36 per complete sequence of 803 bp.

Patterns of Evolution in "Nonmitochondrial" Haplotypes

There is much evidence that clones differing from mtDNA sequence in each individual are not of contemporary mitochondrial origin. First, sequences showing nucleotides seen only in PCR products from whole-aphid DNA (and not from purified mtDNA) cannot be mitochondrial. Second, two clones (Sm80.1 and Ss194.4) have frame-shifting deletions. All but Ss194.21 can be positively identified as nonmitochondrial by these criteria. However, Ss194.21 is one of only two clones having a point substitution within the highly conserved $tRNA^{leu}$ (Sperling and Hickey 1994; present data); the other is nonmitochondrial Ss194.4. Also, like

Downloaded from https://academic.oup.com/mbe/article/13/3/510/1052322 by guest on 19 April 2024

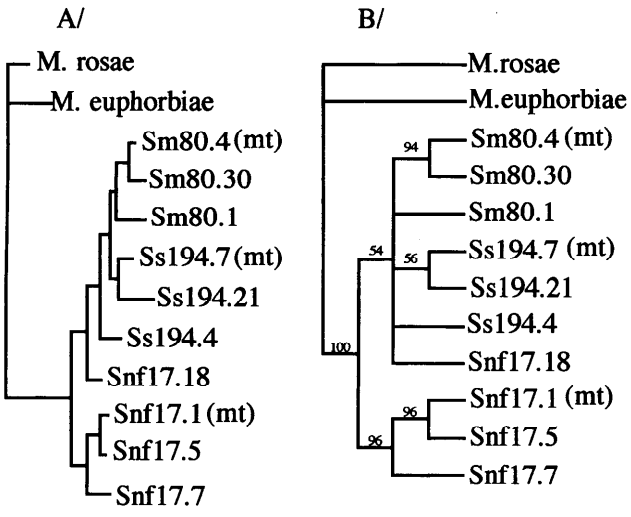


FIG. 4.—*A,* Parsimony phylogeny based on variation at all sites, with *Macrosiphum* species as outgroups. The tree was the shortest (149 steps) found in a branch-and-bound PAUP search. Consistency index (CI) excluding uninformative characters = 0.66. Skewness of the distribution of lengths of random tree indicates that the data contain significant phylogenetic signal ($g_1 = -1.12$, Hillis and Huelsenbeck 1992). Mitochondrial sequences are denoted "(mt)." *B,* The consensus tree resulting from 500 branch-and-bound bootstrap resamplings of the tree in *a*: percentage support is given at the nodes.

the other sequences not sharing the proven or inferred mtDNA haplotype from the same individual, Ss194.21 diverges substantially (at 18 positions) from that sequence. In light of the above, all clones not sharing mtDNA sequences, including Ss194.21, are referred to as "nonmitochondrial."

For analysis of evolution of the sequences, it was necessary to estimate phylogenetic relationships between clones. Accordingly, a branch-and-bound parsimony search was carried out in PAUP using all variable sites and all the cloned sequences, with *Macrosiphum* sequences as outgroups. One shortest tree of 149 steps was found, and was investigated by 500 bootstrapping replicates, again using branch-and-bound searches (fig. 4).

MtDNA in each individual is in the most derived grouping, while nonmitochondrial clones branch from earlier nodes (fig. 4). Nonmitochondrial sequences are clearly not derived from each other by duplication: they have a high proportion of unique characters, and share very few derived characters. We propose that the cloned nonmitochondrial sequences evolved from ancestral mtDNAs in separate transpositions.

We investigated nucleotide evolution in the seven nonmitochondrial sequences by mapping unambiguous state changes over a phylogenetic tree and determining rates and patterns of nucleotide substitution over branches leading to mitochondrial and nonmitochondrial sequences. In the most parsimonious tree (fig. 4), one clone (Snf17.18) did not group with the other clones from its species, but with *S. miscanthi/S.* "Smilax" clones. There is little resolution where the *S. miscanthi* + *S.* "Smilax" grouping diverges from *S.* near *fragariae,* and it appears that homoplasies influence the placement of Snf17.18 away from the other *S.* near *fragariae* sequences. Constraining Snf17.18 to group with other *S.* near *fragariae* sequences increases the tree length by only 2 (to 151). Analysis of patterns of change was carried out with Snf17.18 placed as the most basal sequence in the *S.* near *fragariae* grouping—the most parsimonious tree consistent with the hypothesis that nonmitochondrial copies of mtDNA sequence arose from transposed mtDNA ancestral sequences.

The patterns of evolution along branches leading to mtDNA sequences and along ones leading to nonmitochondrial sequences were strikingly different, in a number of ways explored in the rest of this section. There were not enough changes in individual clones for clone-by-clone analysis to be statistically robust. The cloned nonmitochondrial sequences show similar characteristics (table 4), so data were pooled for statistical analysis. Change along the three longest mtDNA paths (from the nodes Sm80.4–Sm80.1, Ss194.7–Ss194.4, and Snf17.1–Snf17.18) was calculated for comparison. MtDNA state changes are predominantly at third positions (86%), sometimes at first positions (10%), and rarely at second positions (3%). The nonmitochondrial sequences show significantly less bias ($\chi^2 = 8.7$, 2 df, $P < 0.025$), with 59% of changes at third positions, 33% at first positions, and 8% at second positions. However, the deficit of second positions and excess of third positions still represents a pattern significantly different from random ($\chi^2 = 24.3$, 2 df, $P < 0.001$). As the pattern of position bias in nonmitochondrial clones is more similar to that seen in mtDNA than to random, a proportion of change in nonmitochondrial clones seems to have occurred under mtDNA constraints before transposition. Other evidence for constraint and possible causes of this phenomenon are investigated later.

RSFs were calculated for the three longest paths between mtDNA sequences (specified earlier), and for the weighted average of the seven nonmitochondrial branches. RSFs show considerable differences between mtDNA and nonmitochondrial sequences. In particular, the RSFs in nonmitochondrial sequences tend to be more equal than those in mtDNA, and there are six types of change (five are transversions) seen in the former but not in the latter (table 5). While mtDNA and nonmitochondrial sequence evolution are both marked by a prevalence of transitions, the proportion of these is higher for mtDNA (>85%) than for nonmitochondrial DNA (>73%). Also, there are differences between the two types of sequence in the likelihood that a given nucleotide changes, or that a given nucleotide results (re-

Downloaded from https://academic.oup.com/mbe/article/3/5/507/1053337 by guest on 19 April 2021

Table 4
Summary of Unambiguous State Changes Compared in mtDNA and Non-mitochondrial (non-mt) Clones

Change in/Since Ancestor With	Position COI, COII			Total	$tRNA^{leu}$	Dels	Unique
	1	2	3				
S. miscanthi							
80.4 mtDNA/80.30 non-mt	0	0	2	2	0	0	
80.30 non-mt/80.4 mtDNA	1	1	4	6	0	0	4
80.4 mtDNA/80.1 non-mt	0	1	6	7	0	0	
80.1 non-mt/80.4 mtDNA	5	1	4	10	0	1	6
S. "Smilax"							
194.7 mtDNA/194.21 non-mt	1	0	4	5	0	0	
194.21 non-mt/194.7 mtDNA	4	1	6	11	1	0	5
194.7 mtDNA/194.4 non-mt	2	0	7	9	0	0	
194.4 non-mt/194.7 mtDNA	1	0	7	8	1	2	5
S. near *fragariae*							
17.1 mtDNA/17.5 non-mt	0	0	2	2	0	0	
17.5 non-mt/17.1 mtDNA	1	0	1	2	0	0	2
17.1 mtDNA/17.7 non-mt	1	0	6	7	0	0	
17.7 non-mt/17.1 mtDNA	3	0	4	7	0	0	3
17.1 mtDNA/17.18 non-mt	1	0	12	13	0	0	
17.18 non-mt/17.1 mtDNA	1	1	3	5	0	0	2
Summary of mtDNA							
Total	5	1	39	45	0	0	
Three longest paths	3	1	25	29	0	0	
Three shortest branches	1	0	8	9	0	0	
Summary of nonmitochondrial							
Total	16	4	29	49	2	3	
Three longest paths	7	2	14	23	—	—	
Three shortest branches	6	2	11	19	—	—	
Unique sites	10	4	11	25	—	—	

NOTE.—Change was measured over the most parsimonious tree consistent with the hypothesis that nonmitochondrial clones were transposed from mtDNA from the same species. Unique changes are defined in the text. The "three shortest branches" are those linking the mtDNA to the most similar transposition: Sm80.4/Sm80.30, Ss194.7/Ss194.21, and Snf17.1/Snf17.5.

spectively, row and column totals in table 5). The differences in row and column totals for A, G, C, and T are generally smaller in nonmitochondrial DNA than mtDNA. For both types of sequence, patterns of replacement indicate that DNA will tend to become richer in A and T.

If a mutation is equally likely on both DNA strands, then RSFs of the following complementary nucleotide exchanges are expected to be equivalent: G→A and C→T; A→G and T→C; A→T and T→A; A→C and T→G; C→A and G→T; C→G and G→C (Li, Wu, and Luo 1984). We note that the strand asymmetry for the mtDNA sequences is quite extreme (e.g., G→A = 0% while C→T = 51.1%; A→G = 4.2% while T→C = 32.9%), as reported for other insect mtDNA (Tamura 1992). In contrast, these equalities hold much better in the overall nonmitochondrial data, as would be expected if more changes were free to occur on both strands.

Nonmitochondrial Sequences Show Some Evolutionary Constraint

In addition to codon position bias in nonmitochondrial sequences (above), there was other evidence that they did not evolve under selective neutrality. DNA which evolves under functional constraint is expected to show a lower proportion of amino acid replacements than sequence which is less constrained, and replacements in the former should tend to be between amino acids with similar properties (Li, Wu, and Luo 1984). The ratio of nonsynonymous to synonymous substitution is an order of magnitude higher in the nonmitochondrial than in the mtDNA sequences; overall 0.75:1 and 0.07:1, respectively (table 6). These data are consistent with a lower degree of constraint in nonmitochondrial sequences than in mtDNA. However, the nonmitochondrial sequences still show some signs of constraint: there are significantly more synonymous substitutions ($\chi^2 = 20.7$, 1 df, $P < 0.001$) and significantly more of the most conservative amino acid replacements than expected under random models ($\chi^2 = 6.1$, 1 df, $P < 0.025$; table 7). We checked whether the excess of synonymous substitutions in nonmitochondrial sequences can be explained simply by codon composition and transition bias, and conclude that it cannot. In *S. miscanthi* mtDNA sequence, 27% of codons are four-fold degenerate, 44% are synonymous for C/T interchanges, and 29% are synonymous for A/G interchanges. Under evolution with 75% transitions we expect $(1 \times 0.27) + (0.6 \times 0.44) + (0.6 \times 0.29) = 0.708$ of third-position substitutions to be synonymous. This gives an expected number of 20.5, which is very similar to the 20 predicted from the codon table, but differs significantly from the observation of 26 (table 6; $\chi^2 = 5.04$, 1 df, $P < 0.025$). The excess of third-position synonymous substitutions is associated with an excess of C/T interchanges, indicating strand asymmetry similar to that seen in the mtDNA data (table 5).

Many Changes Inferred in Nonmitochondrial Clones may have Occurred before Transposition

Attempts were made to identify characters which arose before transposition, possibly in an as yet unchar-

Downloaded from https://academic.oup.com/mbe/article/13/3/510/... by guest on 19 April 2024

Table 5
Relative Substitution Frequencies (RSF, %) for mtDNA and Nonmitochondrial Sequences

	Nonmitochondrial, codon position 1 (n = 16, 78.2%)						Nonmitochondrial, "unique" sites (n = 25, 73.0%)				
	A	C	G	T	Total		A	C	G	T	Total
A		0.0	19.7	1.4	21.1	A		1.7	19.0	1.7	22.4
C	8.6		0.0	42.9	51.5	C	0.0		13.1	17.6	30.7
G	5.5	0.0		9.2	14.7	G	25.2	0.0		10.8	36.0
T	0.0	10.1	2.6		12.7	T	0.0	11.5	0.0		11.5
Total	14.1	10.1	22.3	53.5		Total	25.2	13.2	32.1	30.1	

	Nonmitochondrial, codon position 3 (n = 29, 73.8%)						MtDNA codon position 3 (n = 25, 85.3%)				
	A	C	G	T	Total		A	C	G	T	Total
A		1.9	13.1	1.9	16.9	A		0.0	1.6	0.0	1.6
C	0.0		20.5	33.6	54.1	C	13.3		0.0	66.8	80.1
G	0.0	0.0		0.0	0.0	G	0.0	0.0		0.0	0.0
T	1.9	27.1	0.0		29.0	T	1.4	16.9	0.0		18.3
Total	1.9	29.0	33.6	35.5		Total	14.7	16.9	1.6	66.8	

	Nonmitochondrial, all changes (n = 49, 75.2)						MtDNA, all changes (n = 29, 88.2%)				
	A	C	G	T	Total		A	C	G	T	Total
A		1.4	15.9	5.7	23.0	A		0.0	4.2	0.0	4.2
C	3.6		5.2	14.8	23.6	C	9.6		0.0	51.1	60.7
G	24.3	0.0		7.2	31.5	G	0.0	0.0		0.0	0.0
T	0.9	20.2	0.7		21.8	T	2.1	32.9	0.0		35.0
Total	28.8	21.6	21.8	27.7		Total	11.7	32.9	4.2	51.1	

NOTE.—RSFs for mtDNA summarize only the three longest paths on the parsimony tree, to maintain independence of data. Number of unambiguous changes and percent of transitions are given (n, %). The y-axis indicates changes from, and the x-axis changes to, the base indicated.

Table 6
Ratios of Nonsynonymous : Synonymous Substitutions in mtDNA and Nonmitochondrial Clones

	CODON POSITION COI, COII			
TYPE OF SEQUENCE	1	2	3	Total
MtDNA (three longest paths)				
Sm80.4	0:0	1:0	0:6	1:6
Ss194.7	0:2	0:0	0:7	0:9
Snf17.1	1:0	0:0	0:12	1:12
Total mtDNA	1:2	1:0	0:25	2:27
Expected[a]	2.9:0.1	1.0:0.0	5.0:20.0	21.8:7.3
Nonmitochondrial DNA				
Sm80.30	0:1	1:0	1:3	2:4
Sm80.1	5:0	1:0	1:3	7:3
Ss194.21	3:1	1:0	0:6	4:7
Ss194.4	1:0	0:0	0:7	1:7
Snf17.5	1:0	0:0	0:1	1:1
Snf17.7	3:0	0:0	0:4	3:4
Snf17.18	1:0	1:0	1:2	3:2
Total nonmitochondrial DNA	14:2	4:0	3:26	21:28
Expected[a]	15.4:0.6	4.0:0.0	9.0:20.0	36.7:12.3

[a] Expectation under random change in random protein-coding sequence, Li and Graur (1991).

Downloaded from https://academic.oup.com/mbe/article/13/3/510/1057321 by guest on 19 April 2024

Table 7
Nature of Nucleotide Changes and Inferred Effects on Amino Acids in Nonmitochondrial Sequences

		NONSYNONYMOUS[a]			
	SYNONYMOUS	Conservative	Moderately Conservative	Moderately Radical	Radical
Expected proportion[b]	0.279	0.201	0.282	0.129	0.11
Expected number ...	13.7	9.8	13.8	6.3	4.9
Observed number ...	28	11	7	1	2

[a] Categories as used by Li, Wu, and Luo (1984) defined in Grantham (1974).
[b] From Table 8 of Li, Wu, and Luo (1984).

acterized mtDNA, in the seven nonmitochondrial clones. We hypothesized that 38 sites seen to change under mtDNA evolution ("mtDNA-like" sites in fig. 3) may be under low constraint and thus disproportionately likely to have evolved in the nonmitochondrial clones prior to transposition. Indeed, the nonmitochondrial clones have a significant excess ($\chi^2_{adj} = 6.7$, 1 df, $P < 0.01$) of changes at these sites. (As there are 38 mtDNA-like sites and 734 protein-coding positions, approximately $38/734 \times 49 = 2.5$ of the 49 substitutions in nonmitochondrial sequences are expected by chance to fall at these sites; seven were observed. The probability that these changes occurred under pseudogene evolution is further reduced because nearly all have an alternative state common in outgroup sequences (e.g. Ss194.21, position 71). We conclude that most or all of these seven changes occurred prior to transposition.

If all inferred state changes in the nonmitochondrial clones really did occur after transposition, it might be expected that the 734 protein-coding sites would be approximately equally likely to change. The average number of changes per clone is 49/7 (table 3), so the mean chance of any particular site being changed is $7/734 = 0.0095$ per clone. Given that one clone is already changed at a given site, the chance of any other clone being changed there is $0.0095 \times 6 = 0.057$. The chance of a third clone also being changed there is $0.057 \times 0.0095 \times 5 = 0.003$. The likelihoods of these events are further reduced (by about 90% for nucleotides C and G) if nucleotide states are shared between clones. Despite these low probabilities, three sites were changed in pairs of nonmitochondrial clones, and two in trios (these five sites are marked as "multiple-hit" sites in fig. 3): in most cases the same nucleotide was shared. We conclude that most or all of the 12 state changes at multiple-hit sites occurred prior to transposition.

Five other changes in the nonmitochondrial sequences (marked "outgroup sites" in fig. 3) were also found to be extremely unlikely to have occurred under pseudogene evolution. These characters were unique among the cloned sequences, but shared with outgroup taxa, suggesting the possibility that they arose in mtDNA. For example, Ss194.21 at position 20 has a C, a state shared with both *Macrosiphum* outgroups. There are only 12 sites where two outgroups share a state not seen in any other sequence. Thus the chance of the event at position 20 under a random model is $13/734 = 0.018$, reduced further ($\times 0.12$) because C is shared with outgroups but is rare (about 12%). Given that there are 11 state changes in Ss194.21, we would expect only 0.023 changes in the clone to be of the type seen at position 20, so we regard the observed change as significantly unlikely. Similar calculations can be made for the other "outgroup" sites, and we conclude that these five sites occurred prior to transposition.

Adding together the "mtDNA-like," "multiple-hit," and "outgroup" sites, we have identified 24 state changes observed in nonmitochondrial clones, most of which probably occurred under constraint before transposition. These changes affect too few second positions (first = 6, second = 0, third = 18), 87.5% are transitions, and a high proportion are synonymous (18 of 24, 75%) including at third positions (17 of 18, 94%). All of these characteristics are consistent with mtDNA evolution.

The remaining 25 changes in the nonmitochondrial clones are unique characters (not seen in any other sequence, marked "unique" in fig. 3) for which there is no a priori reason to suspect that they did not arise under pseudogene evolution. Their evolution is quite consistent with expectations of pseudogene evolution. Unique changes have no significant position bias ($\chi^2 = 3.4$ 2 df, $P > 0.1$), only 73.3% transitions, and there are four types of transversion not seen in the mtDNA data. RSFs indicate little strand asymmetry, and C+G are replaced more than A+T (table 5). The proportion of unique changes which are synonymous (10 of 25, 40%) is close to the 33% expected from the codon table.

There are two factors which reduce confidence that *all* unique sites occurred after transposition. First, position bias is still somewhat against second positions (table 4). Second, applying the observed maximum polymerase incorporation error of 0.36 sites per sequence (above), up to 2.5 of the 25 changes might be due to this cause.

We have not made a direct comparison of the rates of mtDNA and pseudogene evolution, because of the high proportion of apparent pretransposition changes in nonmitochondrial sequences. Possible causes of this finding are discussed later.

Southern Blot Analysis

If many separate copies of mtDNA regions have integrated into the chromosomes of *Sitobion* aphids,

Downloaded from https://academic.oup.com/mbe/article/13/7/1004/1071104 by guest on 16 August 2022

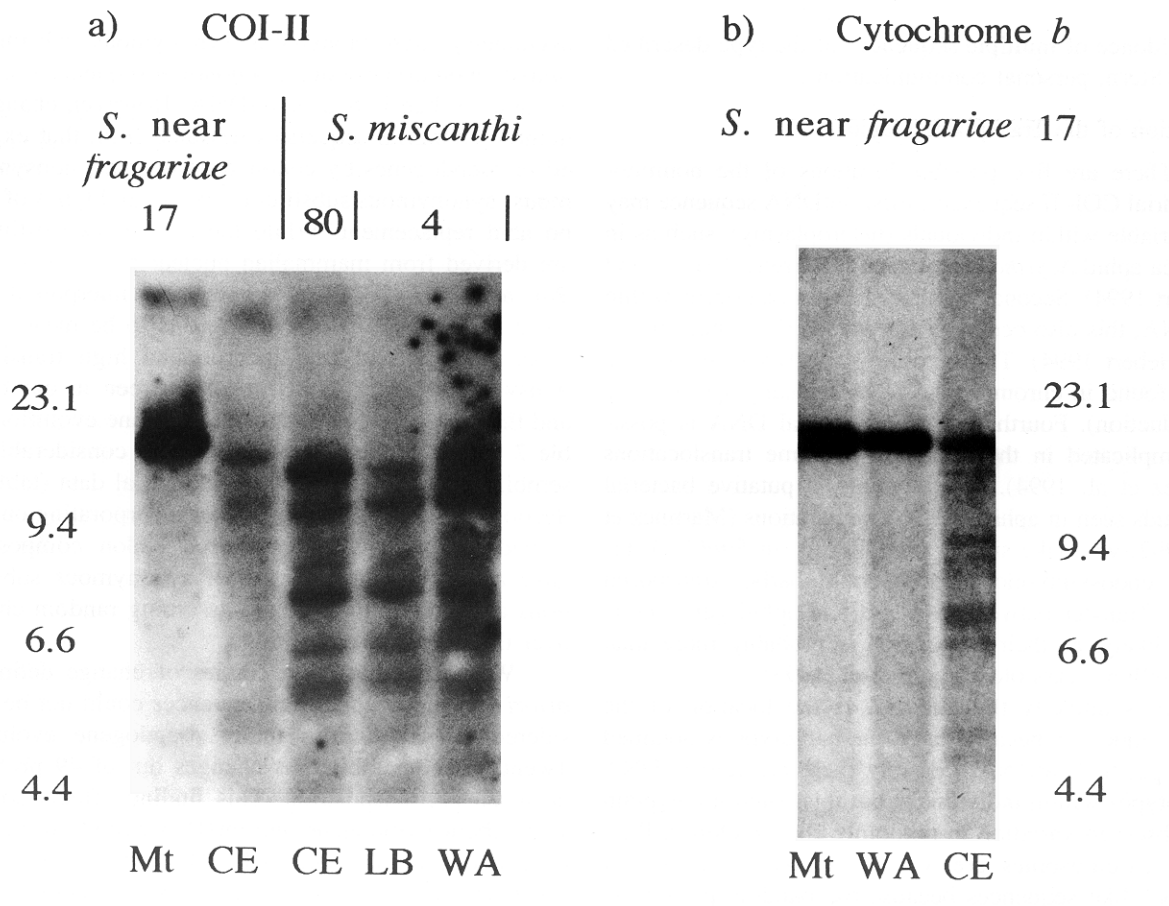


FIG. 5.—Photographs of autoradiographs showing different types of DNA digested with *Hin*dIII and probed with (*a*) COI–II PCR product and (*b*) cytochrome *b* PCR product. Mt = purified mtDNA, CE = chromosomal-enriched DNA, LB = low-bacterial DNA, and WA = whole-aphid DNA. Molecular weights are marked in kilobases.

Southern blots of DNA restricted with endonuclease and probed with mtDNA PCR products should show multibanded patterns. Such multibanded patterns are indeed seen. The autoradiograph of the Southern blot probed with COI–II shows several high-molecular-weight bands in all samples except the *S.* near *fragariae* mtDNA, which has a single large fragment (fig. 5*a*). In *S. miscanthi* samples, the largest strong band is consistent with its being *S. miscanthi* mtDNA, as this is known to be just smaller than the *S.* near *fragariae* fragment (unpublished data). In addition, there are at least six bands between about 5 and 15 kb in the *S. miscanthi* samples. There is no apparent difference in banding pattern between whole-aphid and low-bacterial DNA in *S. miscanthi* 4: from this we infer that endosymbiotic bacteria contain few or no transposed mtDNA-like sequences. The banding patterns of the two *S. miscanthi* samples are similar, but differ from those of *S.* near *fragariae.* On an autoradiograph of a different membrane probed with cytochrome *b,* there are at least 11 clear bands in the *S.* near *fragariae* 17 chromosomal-enriched DNA lane, but only one strong one in whole-aphid and mtDNA lanes (fig. 5*b*).

Discussion

Occurrence in Aphids of Multiple Transposed COI–II Sequences

Our results demonstrate that sequences have been transposed on multiple occasions from mtDNA to a nonmitochondrial location in at least three species of *Sitobion* aphids. In their new position, sequences undergo a less constrained pattern of evolution. PCR products from the COI–II region are heterogeneous in sequence to a degree unprecedented in invertebrates. The phenomenon may extend to some closely related genera including *Macrosiphum, Illinoia,* and *Metopolophium,* but we do not detect ambiguities in the more distantly related *Myzus ornatus.* Stern (1994) reported sequence of cloned COI–II from aphids phylogenetically very distant from *Sitobion* (70–140 million years; Heie 1987) and found

Downloaded from https://academic.oup.com/mbe/article/3/5/510/1057241 by guest on 19 April 2024

no evidence of multiple sequences of the type described here (Stern, personal communication).

Location of the Transposed Copies

There are five possible locations of the nonmitochondrial COI–II sequences. First, mtDNA sequence may be variable within individuals (heteroplasmy), such as in the pea aphid *Acyrthosiphon pisum* (Barrette, Crease, and Hebert 1994). Second, duplicated genes can occur within mtDNA; this also occurs in pea aphids (Barrette, Crease, and Hebert 1994). Third, copies of mtDNA genes have been found in chromosomal DNA in many species (see Introduction). Fourth, extrachromosomal DNA is possibly implicated in the evolution of some translocations (Lopez et al. 1994). One example is putative bacterial plasmids seen in aphid mtDNA preparations (Martinez et al. 1992; we make similar observations in *Sitobion*). Finally, endosymbiotic bacteria could harbor transposed DNA. Transfer between bacterial and aphid genomes is quite likely, as their association is probably more than 160 million years old (Moran et al. 1993).

It is unlikely that mtDNA is the location of the extra sequences because only one haplotype is obtained from purified mtDNA. Possibly there are rare mtDNA haplotypes within individuals, but this could not explain the observed variation in randomly chosen clones. Bacterial chromosomes are also an unlikely location of mtDNA-like sequences because the banding pattern on Southern blots probed with COI–II is similar with or without bacterial DNA (fig. 5*a*). Similarly, plasmid-like structures would be largely excluded from "chromosomal-enriched DNA" but this shows the same pattern as whole-aphid DNA (fig. 5*a*). The 3.5-kb putative plasmid is ruled out, because it is present in purified mtDNA fractions but these show only a single sequence. The likeliest location of the extra mtDNA-like sequences is aphid chromosomes.

Pattern of Evolution of COI–II Sequences: Change in mtDNA versus Pseudogene Branches

Nonmitochondrial sequences would not be functional if translated with the nuclear genetic code because of differences between nuclear and mtDNA codes (Jacobs et al. 1983; Smith, Thomas, and Patton 1992; Lopez et al. 1994). Given the evidence for their nuclear location, it is reasonable to assume that after transposition they retain many characters of mtDNA but start to evolve as nuclear pseudogenes (see also Jacobs et al. 1983; Fukuda et al. 1985; Smith, Thomas, and Patton 1992; Lopez et al. 1994).

We assessed the pattern of sequence evolution over unambiguous state changes over an outgroup-rooted parsimony tree. In all parameters (position bias, relative rates of interchange between the four nucleotides, strand asymmetry, nonsynonymous : synonymous substitution ratios), evolution of nonmitochondrial sequence is closer to random than is that of mtDNA. However, change in nonmitochondrial sequences deviated from that expected in pseudogenes by codon position bias, nonsynonymous : synonymous substitution ratios, and types of amino acid replacement. While these latter two estimates are derived from mammalian nuclear pseudogenes (Li, Wu, and Luo 1984) and so must be interpreted with some caution, we believe the results to be meaningful. First, Li, Wu, and Luo incorporated high transition : transversion bias in their model (as seen in our data), and their estimated pattern of pseudogene evolution (table 2 of Li, Wu, and Luo 1984) bears considerable resemblance to our total nonmitochondrial data (table 5). Second, we showed explicitly that incorporating our observed transition bias and initial codon composition gave a very similar estimate of synonymous substitutions at third positions as did assuming random change over the universal codon table.

We found that three classes of change defined *a priori* in nonmitochondrial sequences could not be considered to have arisen under pseudogene evolution. Twenty-four unambiguous changes out of 49 probably occurred pretransposition. This finding affects conclusions about pseudogene and mtDNA evolution. First, if the pseudogenes and terminal mtDNAs truly share direct ancestors, then extra steps in mtDNA branches of the parsimony tree are implied. mtDNA evolution would be more rapid than inferred from state changes on the tree, and pseudogene evolution would be slower. Second, it is possible that our mtDNAs and pseudogenes are not directly related. For example, through past sexual reproduction, ancestral mtDNAs or transposed copies from different lineages may have been mixed together within individuals. Alternatively, we may have uncovered pseudogenes of divergent ancestral haplotypes transposed in heteroplasmic individuals. Any of the above may affect estimation of rates of evolution in mtDNA and/or transposed sequences. Consequently, we need to know which of these factors affected transposed sequences before we can confidently compare rates of pseudogene and mtDNA evolution.

Numerous Transposition Events Inferred in *Sitobion* Aphids

We have presented evidence which convinces us that the sequenced pseudogenes derive from separate transpositions rather than from duplication and modification of other transposed sequences. The many unique sites and the low number of derived sites shared between transposed sequences indicate separate origins. Only one transposed sequence is more closely related to another pseudogene than to contemporary mtDNA from the same

Downloaded from https://academic.oup.com/mbe/article/13/3/510/1023431 by guest on 16 April 2024

species (Snf17.7–Snf17.5) and even this relies on only 1 change out of 18 (table 3). There is also direct evidence for several transposed copies which have not yet been sequenced: there are ambiguous sites in the data set which have not yet been explained by cloned sequences.

Even in the unlikely case that we have detected all separate transpositions, there have been seven within the three very closely related *Sitobion* species. There is only one published precedent for such large numbers of mtDNA transposition events: in humans (although cicadas in the genus *Magicicada* also may show multiple transpositions, James Tang, Chris Simon, and Grey Staley, personal communication). Fukuda et al. (1985) inferred 10 to 130 copies of each of four mtDNA regions to be dispersed through human chromosomes. In other species, there have been only small numbers of events (e.g., sea urchin, Jacobs et al. 1983). We do not know at present why transpositions should be common in *Sitobion* aphids. Possibly some process makes fragments of mtDNA more common in cells (such as the human autosomal locus which favors mtDNA deletions, Suomalainen et al. 1995). It is also possible that a mechanism leading to high frequency of transposition has arisen recently within the *Sitobion* lineage.

Applications of Nuclear Transposed Copies of mtDNA Sequences

The present results raise some exciting possibilities for exploration of evolution of mtDNA regions and for testing some molecular and phylogenetic assumptions. Far from being technical obstacles to avoid, transposed sequences appear to be "fossilized" mtDNAs. Transposed sequences may be useful to estimate comparative rates of evolution of nuclear pseudogenes and mtDNA. Much attention has been paid to this important topic (Britten 1986; Vawter and Brown 1986: DeSalle et al. 1987; Tamura 1992). For such applications, the *Sitobion* group offers a very unusual testing ground: there are many accessible transpositions, and the genus is quite speciose. In some species reproduction is usually or always parthenogenetic, so patterns of genetic divergence should be simplified. For example, transfer of nuclear or mtDNA haplotypes through sexual reproduction should occur rarely or not at all. A major potential advantage of studying transposed mtDNA sequences rather than mtDNA alone is that the former should be overwritten by changes only at random sites, leaving a truer record of mtDNA evolution. However, before such applications can be realized in the *Sitobion* system, we need to know more about the nature of the transpositions.

Investigation of transposed DNA in molecular evolutionary studies may be applicable in many species, because mtDNA transpositions are probably much more common and widespread than the general rate of reporting suggests. Many independent transpositions have been found in humans where research effort is greatest (Tsuzuki et al. 1983; Fukuda et al. 1985; Kamimura et al. 1989; Wenger and Gassmann 1995), and many transpositions are being found during sequencing for other purposes (see Introduction). Our results highlight the need for establishing the origin of sequence under examination.

Sequence Availability

Aligned sequences are available from GenBank: accession numbers U41114–U41132.

Acknowledgments

We thank Simon Easteal, Craig Moritz, an anonymous referee, Andrea Taylor, Phillip England, and James Cook for commenting on the manuscript, and David Stern and Martin Slade for informative discussions. David Stern was also very generous in providing primers. We are indebted to all in lab E8C230: especially Cathy Watson, David Loebel, Karen Humphrey, and Peter Johnston. Thanks are due to Dr. Mary Carver, Dr. Masahisa Miyazaki, and Cliff Brookes for samples and species determinations, and to Dr. R. L. Blackman and Professor O. E. Heie for discussions about outgroups. This work was supported by ARC grant A19130882 to DFH, and by Macquarie University funds.

Downloaded from https://academic.oup.com/mbe/article/13/3/0/1027892 by guest on 09 April 2024

LITERATURE CITED

BARRETTE, R. J., T. J. CREASE, and P. D. N. HEBERT. 1994. Mitochondrial DNA diversity in the pea aphid *Acyrthosiphon pisum*. Genome **37**:858–865.

BECKENBACH, A., Y. W. WEI, and H. LIU. 1993. Relationships in the *Drosophila obscura* species group, inferred from mitochondrial cytochrome oxidase II sequences. Mol. Biol. Evol. **10**:619–634.

BRITTEN, R. J. 1986. Rates of DNA sequence evolution differ between taxonomic groups. Science **231**:1393–1398.

BROWN, J. M., O. PELLMYR, J. N. THOMPSON, and R. G. HARRISON. 1994. Phylogeny of *Greya* (Lepidoptera: Prodoxidae), based on nucleotide sequence variation in mitochondrial cytochrome oxidase I and II: congruence with morphological data. Mol. Biol. Evol. **11**:128–141.

BRUFORD, M. W., O. HANOTTE, J. F. Y. BROOKFIELD, and T. BURKE. 1992. Single-locus and multilocus DNA fingerprinting. Pp. 225–269 *in* A. R. HOELZEL, ed. Molecular genetic analysis of populations: a practical approach. Oxford University Press, Oxford.

DESALLE, R., T. FREEDMAN, E. M. PRAGER, and A. C. WILSON. 1987. Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian *Drosophila*. J. Mol. Evol. **26**:157–164.

FUKUDA, M., S. WAKASUGI, T. TSUZUKI, H. NOMIYAMA, and K. SHIMADA. 1985. Mitochondrial DNA-like sequences in the human nuclear genome: characterization and implications in the evolution of mitochondrial DNA. J. Mol. Biol. **186**:257–266.

GELLISEN, G., J. Y. BRADFIELD, B. N. WHITE, and G. R. WYATT. 1983. Mitochondrial DNA sequences in the nuclear genome of a locust. Nature **301**:631–634.

GOJOBORI, T., W.-H. LI, and D. GRAUR. 1982. Patterns of nucleotide substitution in pseudogenes and functional genes. J. Mol. Evol. **18**:360–369.

GRANTHAM, R. 1974. Amino acid difference formula to help explain protein evolution. Science **185**:862–864.

HALES, D. F., R. L. CHAPMAN, R. M. LARDNER, R. COWEN, and E. TURAK. 1990. Aphids of the genus *Sitobion* occurring on grasses in southern Australia. J. Aust. Entomol. Soc. **29**:19–25.

HARTL, F.-U., and W. NEUPERT. 1990. Protein sorting to mitochondria: evolutionary conservations of folding and assembly. Science **247**:930–938.

HEIE, O. E. 1987. Evolution. Pp. 367–391 *in* A. K. MINKS and P. HARREWIJN, eds. Aphids, their biology, natural enemies and control. Elsevier, Amsterdam.

HILLIS, D. M., and J. P. HUELSENBECK. 1992. Signal, noise, and reliability in molecular phylogenetic analyses. J. Hered. **83**:189–195.

JACOBS, H. T., J. W. POSAKONY, J. W. GRULA, J. W. ROBERTS, J.-H. XIN, R. J. BRITTEN, and E. H. DAVIDSON. 1983. Mitochondrial DNA sequences in the nuclear genome of *Strongylocentrotus purpuratus*. J. Mol. Biol. **165**:609–632.

KAMIMURA, N., S. ISHII, M. LIANDONG, and J. W. SHAY. 1989. Three separate mitochondrial DNA sequences are contiguous in human genomic DNA. J. Mol. Biol. **210**:703–707.

LI, W.-H., and D. GRAUR. 1991. Fundamentals of molecular evolution. Sinauer, Sunderland, Mass.

LI, W.-H., C.-I. WU, and C.-C. LUO. 1984. Nonrandomess of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. J. Mol. Evol. **21**:58–71.

LIU, H., and A. T. BECKENBACH. 1992. Evolution of the mitochondrial cytochrome oxidase II gene among 10 orders of insects. Mol. Phylogenet. Evol. **1**:41–52.

LOPEZ, J. V., N. YUHKI, R. MASUDA, W. MODI, and S. J. O'BRIEN. 1994. *Numt*, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. J. Mol. Evol. **39**:174–190.

MADDISON, W. P., and D. R. MADDISON. 1992. MacClade: analysis of phylogeny and character evolution, version 3.0. Sinauer Associates, Sunderland, Mass.

MARGULIS, L. 1970. Origin of eukaryotic cells. Yale University Press, New Haven, Conn.

MARTINEZ, D., A. MOYA, A. LATORRE, and A. FERERES. 1992. Mitochondrial DNA variation in *Rhopalosiphum padi* (Homoptera: Aphididae) populations from four Spanish localities. Ann. Entomol. Soc. Am. **85**:241–246.

MORAN, N. A., M. A. MUNSON, P. BAUMANN, and H. ISHIKAWA. 1993. A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. Proc. R. Soc. Lond. B **253**: 167–171.

POWERS, T. O., S. G. JENSEN, S. D. KINDLER, C. J. STRYKER, and L. J. SANDALL. 1989. Mitochondrial DNA divergence among greenbug (Homoptera: Aphididae) biotypes. Ann. Entomol. Soc. Am. **82**:298–302.

ROEHRDANZ, R. L. 1993. An improved primer for PCR amplification of mitochondrial DNA in a variety of insect species. Insect Mol. Biol. **2**:89–91.

SAMBROOK, J., E. F. FRITSCH, and T. MANIATIS. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbour Laboratory Press, New York.

SIMON, C., F. FRATI, A. BECKENBACH, B. CRESPI, H. LIU, and P. FLOOK. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann. Entomol. Soc. Am. **87**:651–701.

SMITH, M. F., W. K. THOMAS, and J. L. PATTON. 1992. Mitochondrial DNA-like sequence in the nuclear genome of an akodontine rodent. Mol. Biol. Evol. **9**:204–215.

SPERLING, F. A. H., and D. A. HICKEY. 1994. Mitochondrial DNA sequence variation in the spruce budworm species complex (*Choristoneura*: Lepidoptera). Mol. Biol. Evol. **11**: 656–665.

STERN, D. L. 1994. A phylogenetic analysis of soldier evolution in the aphid family Hormaphididae. Proc. R. Soc. Lond. B **256**:203–209.

SUOMALAINEN, A., J. KAUKONEN, P. AMATI, R. TIMONEN, M. HALTIA, J. WEISSENBACK, M. ZEVIANI, H. SOMER, and L. PELTONEN. 1995. An autosomal locus predisposing to deletions of mitochondrial DNA. Nat. Genet. **9**:146–151.

SWOFFORD, D. L. 1993. PAUP: phylogenetic analyisis using parsimony. Version 3.1. Illinois Natural History Survey, Champaign, Ill.

TAMURA, K. 1992. The rate and pattern of nucleotide substitution in *Drosophila* mitochondrial DNA. Mol. Biol. Evol. **9**:814–825.

TAMURA, K., and T. AOTSUKA. 1988. Rapid isolation method of animal mitochondrial DNA by the alkaline lysis procedure. Biochem. Genet. **26**:815–819.

TSUZUKI, T., H. NOMIYAMA, C. SETOYAJA, S. MAEDA, and K. SHIMADA. 1983. Presence of mitochondrial-DNA-like sequences in the human nuclear DNA. Gene **25**:223–229.

VAWTER, L., and W. M. BROWN. 1986. Nuclear and mitochondrial DNA comparisons reveal extreme rate variation in the molecular clock. Science **234**:194–196.

WENGER, R. H., and M. GASSMANN. 1995. Mitochondria contaminate databases. Trends Genet. **11**:167–168.

WINSHIP, P. R. 1989. An improved method for directly sequencing PCR amplified material using DMSO. Nucleic Acids Res. **17**:1266.

ZULLO, S., L. C. SIEU, J. L. SLIGHTOM, H. I. HADLER, and J. M. EISENSTADT. 1991. Mitochondrial d-loop sequences are integrated in the rat nuclear genome. J. Mol. Biol. **221**: 1223–1235.

SIMON EASTEAL, reviewing editor

Accepted November 23, 1995

Downloaded from https://academic.oup.com/mbe/article/13/3/510/1055821 by guest on 16 April 2024