Host Race Evolution in *Schizaphis graminum* (Hemiptera: Aphididae): Nuclear DNA Sequences

KEVIN A. SHUFRAN

USDA–ARS, Wheat, Peanut, and Other Field Crops Research Unit, 1301 N. Western Road, Stillwater, OK 74075

Environ. Entomol. 40(5): 1317–1322 (2011); DOI: http://dx.doi.org/10.1603/EN11103

**ABSTRACT**

The greenbug aphid, *Schizaphis graminum* (Rondani), was introduced into the United States in the late 1880s, and quickly was established as a pest of wheat, oat, and barley. Sorghum was also a host, but it was not until 1968 that greenbug became a serious pest of it as well. The most effective control method is the planting of resistant varieties; however, the occurrence of greenbug biotypes has hampered the development and use of plant resistance as a management technique. Until the 1990s, the evolutionary status of greenbug biotypes was obscure. Four mtDNA cytochrome oxidase subunit I (COI) haplotypes were previously identified, suggesting that *S. graminum* sensu lato was comprised of host-adapted races. To elucidate the current evolutionary and taxonomic status of the greenbug and its biotypes, two nuclear genes and introns were sequenced; cytochrome c (CytC) and elongation factor 1-α (EF1-α). Phylogenetic analysis of CytC sequences were in complete agreement with COI sequences and demonstrated three distinct evolutionary lineages in *S. graminum*. EF1-α DNA sequences were in partial agreement with COI and CytC sequences, and demonstrated two distinct evolutionary lineages. Host-adapted races in greenbug are sympatric and appear reproductively isolated. Agricultural biotypes in *S. graminum* likely arose by genetic recombination via meiosis during sexual reproduction within host-races. The 1968 greenbug outbreak on sorghum was the result of the introduction of a host race adapted to sorghum, and not selection by host resistance genes in crops.

**KEY WORDS** biotypes, cytochrome c, elongation factor 1-α, greenbug, host-plant resistance
**Table 1.** *Schizaphis graminum* biotypes and mtDNA haplotype associations from which nDNA was sequenced

<table>
<thead>
<tr>
<th>mtDNA Haplotype</th>
<th>Biotype</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>C</td>
<td>(Shufran et al. 2000)</td>
</tr>
<tr>
<td>E</td>
<td>(Shufran et al. 2000)</td>
<td></td>
</tr>
<tr>
<td>E-OK</td>
<td>(Shufran and Puterka 2011)</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>(Shufran et al. 2000)</td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>(Shufran et al. 2000)</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>(Shufran et al. 2000)</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>F</td>
<td>(Shufran et al. 2000)</td>
</tr>
<tr>
<td>G</td>
<td>(Shufran et al. 2000)</td>
<td></td>
</tr>
<tr>
<td>NY</td>
<td>(Shufran et al. 2000)</td>
<td></td>
</tr>
<tr>
<td>?-OK</td>
<td>(Shufran and Puterka 2011)</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>B</td>
<td>(Shufran et al. 2000)</td>
</tr>
<tr>
<td>B-OK</td>
<td>(Shufran and Puterka 2011)</td>
<td></td>
</tr>
<tr>
<td>Paspalum vaginatum</td>
<td>(Shufran and Puterka 2011)</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>H</td>
<td>(Shufran et al. 2000)</td>
</tr>
</tbody>
</table>

cytochrome oxidase subunit I (COI), which determined *S. graminum* populations had three mtDNA COI haplotypes (Shufran et al. 2000, Anstead et al. 2002). These mtDNA sequencing results were consistent with the results of restriction fragment length polymorphism (RFLP) analysis of mtDNA (Powers et al. 1989) and rDNA (Black 1993) and supported the theory that *S. graminum* was comprised of host-adapted races or subspecies. However, studies of intrabiotype variation demonstrated that biotypes existed independent of COI haplotypes (Anstead et al. 2002). Single biotypes were found in more than one COI haplotype.

The source of biotypic diversity was shown to be a product of recombination of virulence genes during meiosis, which occurs during the holocyclic life cycle phase of *S. graminum* (Puterka and Peters 1989, 1990, 1995). Crosses between and within biotypes resulted in new biotypic variation in the offspring. Multiple biotypes found among aphids with different COI haplotypes (Anstead et al. 2002) suggest there was gene flow between host races or homologous virulence genes existed between host races that were reproductively isolated. To answer this question, DNA sequence variation in the intron and codon regions of two nuclear genes was analyzed using the same *S. graminum* populations of Shufran et al. (2000), plus four additional clones.

**Materials and Methods**

Ten *S. graminum* biotypes (B, C, E, F, G, H, I, J, K, and NY), preserved at ~80°C from Shufran et al. (2000), three biotypes (E, B, and an unknown, designated as “?”) collected from wheat at Okeene, OK (Shufran and Puterka 2011), and the *Paspalum vaginatum* Sw. (AKA Florida or P) biotype (Nuressly et al. 2008) were used in the current study. The mtDNA COI haplotype associated with each is shown in Table 1. The prepGEM Insect DNA extraction kit (ZyGEM Corp. Ltd., Hamilton, New Zealand) was used according to the manufacturer’s instructions to extract genomic DNA from a single apterous individual of each biotype. Two nuclear DNA genes and their introns, *cytochrome c* (*CytC*) and *elongation factor 1-α* (*EF1-α*), were polymerase chain reaction (PCR) amplified and sequenced.

A 640 bp PCR amplicon of the *CytC* was produced and direct sequenced using the primers *cytC*-C-5′ (5′-AAGTGTGCYCATGCACAC-3′) (Palumbi 1996) and *cytC*-B-3′ (5′-CATCTTGTTCCGGGGGTATATTCTT-3′) (Palumbi 1996). Amplicons were purified using the Wizard SV Gel and PCR Clean-Up System (Promega, Madison, WI). A 1.2-kbp amplicon of the *EF1-α* was produced using the primers *EF2-3′* (5′-ATGTGACCAGTGTGGCAATTCAA-3′) (Palumbi 1996) and EF515′ (5′-GGAAATGGGGAAAAGCTCTTGCAATGCG-3′) (Moran et al. 1999). The *EF1-α* amplicon was then cloned using the pGEM –T Easy Vector kit (Promega, Madison, WI) and sequenced with the T7 and SP6 primers. DNA sequencing was conducted with 5× to 6× coverage with BigDye-terminated reactions analyzed on an ABI model 3730 DNA Analyzer (Recombinant DNA/Protein Core Facility, Department of Biochemistry and Molecular Biology, Oklahoma State University, Stillwater, OK). Internal primers EF4 and EF5 were used to complete sequencing (von Dohlen et al. 2002) DNA sequences were assembled and aligned by the ClustalW method (Thompson et al. 1994) using the LaserGene Ver. 8.0.2 software suite (DNASTAR, Inc., Madison, WI). Default alignment parameters were used: gap penalty 15.0, gap length penalty 6.66, delay divergent sequences 30%, and DNA transition weight 0.5. Phylogenetic analyses were conducted using the MEGA5 statistical software package (Tamura et al. 2011). Maximum likelihood (ML) method with 1,000 bootstraps was used based on the Tamura and Nei (1993) model, with uniform substitution rates among sites, all sites (gaps and/or missing data) used, and the ML heuristic method Nearest-Neighbor-Interchange. Completed sequences were submitted to GenBank with accession numbers: *CytC*, JF719744 - JF719757; and *EF1-α*, JF968544 - JF968556. Voucher specimens of *S. graminum* biotypes (sisters of specimens used for DNA analysis) were deposited with the Museum of Entomology, FL State Collection of Arthropods, Gainesville, FL.

**Results**

In total, 518 nucleotides were sequenced from *S. graminum* *CytC*. No nucleotide variation was found in the coding region, i.e., the first 80 bases, and all substitutions, deletions, and insertions occurred in the 438-bp intron that followed. Phylogenetic analysis with maximum likelihood produced a dendrogram with three major clades (Fig. 1a), although the clade containing biotypes C, E, E-OK, I, K, and J was not well supported by bootstrap analysis. Neighbor-joining and maximum parsimony analysis produced essentially identical results. Each clade corresponded to one of the three mtDNA clades of Shufran et al. (2000), except biotype H fell into the clade containing biotypes F, G, NY, and the unknown OK biotype (Fig. 1a).
A sequence alignment of CytC from bases 240–310 shows differentiation of three clades corresponding to mtDNA haplotypes (Fig. 2a). Biotypes with mtDNA haplotype I contained an eight b insertion starting at position 310 and also differed by a SNP at position 278. Biotypes with mtDNA haplotypes II, III and H lacked this insertion, and mtDNA haplotypes II and III differed by two SNPs at positions 261 and 298.

In total, 1,165 nucleotides were sequenced from *S. graminum* EF1-α. Three introns of 63, 72, and 70 bp were found and all nucleotide substitutions, inserts, and deletions occurred in these. The exons of all biotypes had identical DNA sequences. Biotype J was not included because it is the rarest biotype and too few specimens remain in the collection. Biotype J has never been found a second time after its initial discovery. Phylogenetic analysis with maximum likelihood produced a dendrogram with three major clades (Fig. 1b). Neighbor-joining and maximum parsimony analysis produced essentially identical results (results not shown). One clade contained all biotypes having a mtDNA haplotype I (C, E, E-OK, I, and K); however, there were mtDNA haplotype II (F and NY) and haplotype III (Paspalum and B-OK) biotypes within. A second clade contained B-OK (Haplotype II and biotype H), and a third clade contained biotypes G and ?-OK (Haplotype II) and B (Haplotype III).

A sequence alignment of EF1-α from bases 81–160 shows differentiation of biotypes corresponding to...
mtDNA haplotypes (Fig. 2a.). All mtDNA Haplotype I biotypes (C, E, E-OK, I, and K had a three b deletion beginning at position 95. Biotypes F and the Paspalum biotype also had this deletion. Biotypes B, G, and ?-OK were differentiated from Haplotype I biotypes by having a three b insertion beginning at position 95 and SNPs at positions 94, 98, 100, 101, and 111. The Paspalum biotype was identical to biotypes C, E, I, and F. Biotypes B-OK and H had unique sequences. A second variable region of the EF1-α (data not shown) showed a two b deletion at positions 752–753 differentiation between biotypes C, E, E-OK, I, K, F, NY, and Paspalum from G, ?-OK, B, B-OK, and H.

Discussion

*S. graminum* in the United States appears to be comprised of two to three evolutionary distinct entities, which are most likely host-adapted races. This divergence is made evident by the complete agreement of phylogenetic analysis of mtDNA COI (Shufran et al. 2000) and nDNA CytC sequences. The EF1-α sequence data were not in complete agreement with the COI and CytC sequences; however it showed that all the mtDNA haplotype I biotypes were in a single major clade, indicating a distinct evolutionary divergence from most of the other biotypes and their representative mtDNA haplotypes. The EF1-α gene and its introns probably did not evolve at the same rates as COI or CytC, which could account for the discrepancies.

Mitochondrial and nuclear DNA sequence data support the existence of a sorghum host-race that first occurred in the United States during 1968 and has thereafter been predominant on wheat and sorghum. The biotypes E, I, and K can be considered as members of the sorghum biotype, which initially was named biotype C. This grouping can be extended to greenbugs containing the mtDNA haplotype I, which is predominant on crops. Another designation other than biotype can now be assigned for greenbugs that occur on and are pests of sorghum. I propose that these populations and biotypes now be recognized as a true sorghum host race that was introduced into the United States during 1968. This hypothesis was originally suggested by Blackman (1981) and further discussed by Blackman and Eastop (2007). Besides the data presented herein, most other molecular DNA data support this hypothesis (Powers et al. 1989, Black 1993, Zhu-Salzman et al. 2003).

Anstead et al. (2002) and J. D. Burd and K. A. Shufran (J.D.B. and K.A.S., unpublished data), indicated that *S. graminum* host races, as identified by mtDNA haplotypes, are sympatric in nature, occurring on both crop and wild grasses. In addition to the data presented herein, most other molecular DNA data support the hypothesis that *S. graminum* host races were reproductively isolated (Powers et al. 1989, Black 1993, Zhu-Salzman et al. 2003). Little gene flow appears to have occurred between two or three tentative host races. Host race specific sex pheromones produced by oviparae during the holocycle may be one isolating mechanism by which this occurs. Eisenbach and Mittler (1987b) demonstrated sex pheromone discrimination by males in two greenbug biotypes, C and E. Because there is no relation between biotype and mtDNA haplotype (Anstead et al. 2002), the authors could have had two different host races in their experiment. Unfortunately, no voucher specimens were kept by Eisenbach and Mittler (1987b) so it is impossible to determine if this was the case. Further investigation of sex pheromone discrimination in *S. graminum* is warranted.

Puterka and Peters (1989, 1990, 1995) made interbiotype crosses (C, E, and F) in the laboratory. The results demonstrated that recombination of virulence genes resulted in biotypic diversity in *S. graminum*. Based on the molecular genetic data presented in the past and herein, it appears that biotypic diversity may result from sexual reproduction within host-races, not between them. Virulence genes responsible for producing the same biotype in more than one host race (Anstead et al. 2002) would be a result of homologous genes occurring in different host races. It is not known if there was forced mating between mtDNA haplotypes, but it is likely because Puterka and Peters (1989, 1990, 1995) used the same biotype F (Haplotype II) as this study, and because their C and E were collected from wheat. Crosses between biotype F and biotypes C and E made by Puterka and Peters (1989, 1990, 1995) would not likely occur in nature, because biotype F is associated with turf grasses and has not been found on wheat or sorghum.

Biotypic diversity in *S. graminum* now appears to be because of sexual reproduction within two or three host races (or perhaps subspecies), even though populations are sympatric. The evolutionary status of other aphid biotypes also has been determined (e.g., *Acyrthosiphon pisum* (Via 1999, Hawthorne and Via 2001, Via and Hawthorne 2002); *Aphis gossypii* (Vanlerberghe-Masutti and Chavigny 1998, Najar-Rodriguez et al. 2009); and *Therioaphis trifoli* (Sunnucks et al. 1997). In these cases, the term biotype no longer applies because specific causes of intraspecific variation were determined by molecular genetics techniques, host utilization studies, and the development of quantitative trait loci (QTL) genomic maps. The production of a QTL genomic map in *S. graminum* would further elucidate the evolutionary and biochemical basis of biotype, and perhaps lead to host-race distinction. The subsequent identification of specific aphid genes virulent to wheat and sorghum resistance genes would further aid in the development and use of greenbug resistant crops.

Acknowledgments

I thank John Burd for supplying specimens of the *Paspalum vaginatum* biotype, and for determining the biotype of the greenbugs collected near Okeene, OK. Barbara Driskel completed the DNA extraction, PCR, cloning, and preparation of samples for DNA sequencing. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the United States Department of Agriculture.
References Cited


Received 19 April 2011; accepted 27 June 2011.