The dynamics of speciation in the gall-forming aphid *Geoica* spp. within and among *Pistacia* host tree species

Rachel Ben-Shlomo1*, Stav Talal1# and Moshe Inbar2

1 Department of Biology and Environment, Faculty of Natural Sciences, University of Haifa-Oronim, Tivon, 36006, Israel
2 Department of Evolutionary and Environmental Biology, Faculty of Natural Sciences, University of Haifa, Haifa, 3498838, Israel

* Corresponding author. Email: ekly@research.haifa.ac.il
# Present address: School of Life Sciences, Arizona State University, 427 E Tyler Mall, Tempe, AZ 85281

**Keywords:**
*Geoica* spp., gall-forming aphids, haplotype, speciation

**ABSTRACT**
Trees of the genus *Pistacia* serve as obligate hosts for gall-forming aphids (Hemiptera, Aphididae, Fordini). Each aphid species induces a characteristic gall on a single *Pistacia* host species. The genus *Geoica* (Fordini) induces similar spherical closed galls on the lower side of the leaflet’s midvein, on different *Pistacia* species. Two species of *Pistacia* trees that harbor *Geoica* galls grow naturally in Israel: *P. palaestina* and *P. atlantica*. We analyzed the phylogeny and the genetic structure of the *Geoica* species complex in Israel, and assessed the genetic differentiation and the level of host plant specificity of the aphids between *P. atlantica* and *P. palaestina*. We found that the splitting of the genus between *P. atlantica* and *P. palaestina* is estimated to have occurred 24–25 Ma (the Oligocene/Miocene boundary). Five different haplotypes suggesting five different species have been further speciating among *Geoica* spp., galling on *P. atlantica*, and an additional three species, on *P. palaestina*.

**Accession Numbers (COI sequences):**
MW452548 - Pp-Q-7d-Gamla - Haplotype-VI
MW452549 - Pp-Q-10b-Gamla - Haplotype-VII
MW452550 - Pp-A-1a-Mezar - Haplotype-VII
MW452551 - Pp-S-6c-Sasa - Haplotype-VI
MW452552 - Pp-New Hermon-2-5 - Haplotype-VIII
MW452553 - Pa-Q-10a-II Gamla - Haplotype-IV
MW452554 - Pa-Q-5d Gamla - Haplotype-II
MW452555 - Pa-A-2f Mezar - Haplotype-IV
MW452556 - Pa-E-13 Anafa - Haplotype-III
MW452557 - Pa-K-10a Adolam - Haplotype-IV
MW452558 - Pa-K-3a Adolam - Haplotype-II
MW452559 - Pa-N-2a Nizana - Haplotype-I
MW452560 - Pa-Z-1-3 Hazeva - Haplotype-I
MW452561 - Pa-SL-3-3 Jordan-Sela - Haplotype-V
1. INTRODUCTION

Galls are deformed growths of plant tissues that are induced by numerous organisms, primarily insects. The formation of galls requires close associations between the inducers and the host plants. Indeed, most gall-forming insects are strictly specific to a given host plant species, organ and tissue. The mechanism of gall formation and maintenance, and the way insects modify plant developmental pathways is still unknown (Giron et al., 2016). However, typically, the morphology of the galls is associated with the inducing insect species.

Trees of the genus *Pistacia* serve as obligate hosts for gall-forming aphids (Hemiptera, Aphididae, Fordini) (Wool, 2004). The galls are formed early in the spring by a single aphid (the fundatrix), which feeds on the phloem sap. Within each gall several additional generations are produced parthenogenetically (females only). In the fall, winged aphids (the fall alatae) leave the galls and migrate to the roots of non-specific herbaceous host plants (Wool, 2004).

Each aphid species induces a characteristic gall on a single *Pistacia* host species. The galls have distinct shapes, colors and sizes (Inbar et al., 2004; Koach and Wool, 1977; Wool and Burstein, 1991). In most cases, the shape of the gall is characteristic of a particular species of aphid, and as such, can also be used as a "tool" in *Pistacia* systematics (Inbar, 2008), as in other hosts of gall-forming insects (Aguilar and Boecklen, 1992; Floate et al., 1996).

The genus *Geoica* (Fordini) represents an exceptional uncoupling between gall morphology and aphid-species identity. *Geoica* is a Palaeartic genus that forms galls on *Pistacia* trees and alters the host in the fall, to roots of Poaceae, and sometimes Cyperaceae (Blackman and Eastop, 2018). Most *Geoica* species (about 10 species) induce similar spherical (ball-shaped) closed galls, located on the lower (abaxial) side of the leaflet’s midvein, on different *Pistacia* species.

The taxonomic status of the genus *Geoica* is challenging and unclear. Based on the morphometry of fall alatae, Brown and Blackman (1994) analyzed the so-called *G. urticaria* complex. They divided the complex into five species inhabiting the Mediterranean parts of Europe, North Africa, the Levant and Iran: *G. rungsi* and *G. harpazi* on *P. atlantica*; *G. muticae* on *P. mutica*; *G. urticaria* on *P. terebinthus*; and *G. wertheimae* on *P. palaestina*. These species were not associated with any given gall trait. Limited molecular analyses suggested there were three species of the genus *Geoica* in Israel, *G. wertheimae* on *P. palaestina*, and two putative species on *P. atlantica*, both collected in northern Israel (Inbar et al., 2004). These two species, which may possibly be *G. rungsi* and/or *G. harpazi* (Brown and Blackman, 1994), were marked by Inbar et al. (2004) as *Geoica* sp. An additional species producing spherical galls, found on the leaf petiole (not on the leaflet midvein) of *P. atlantica* trees, which was identified at the time as *G. swirskii* (Remaudiere et al., 2004), has been assigned to a new genus, *Inbaria* (Barjadze et al., 2018). Blackman and Eastop (2018) recently revised the status of the *G. urticaria* complex. Considering the substantial variation in the morphology of the galls and of the emigrant alatae and their embryos, they suggested additional species, as yet unnamed, in the genus *Geoica*; these species are closely related to *G. urticaria*, but associated with other *Pistacia* spp.

Two species of *Pistacia* trees that harbor *Geoica* galls grow naturally in Israel: *P. palaestina* and *P. atlantica*. The former is common in Mediterranean habitats, while the latter is found in the Mediterranean and Irano-Turanian distribution zone. Both species inhabit diverse habitats along climatic gradients that vary in solar radiation, temperature and precipitation. The origin and evolution of the genus *Pistacia* are controversial (Kozhoridze et al., 2015; and references therein); hence, most studies suggested that the genus evolved more than 80 M years ago. Within the genus *Pistacia*, *P. palaestina* and *P. atlantica* are placed in different phylogenetic lineages (e.g., Golan-Goldhirsh, 2009; Kafkas and Perl-Treves, 2002; Zohary, 1972; Xie et al., 2014). Nevertheless, natural crossing between the two species does exist (Golan-Goldhirsh, 2009). Natural populations of *P. palaestina* and *P. atlantica* are found sympatrically in several regions in northern and central Israel, while other populations are clearly allopatric. Galls of the genus *Geoica* can be readily found throughout the distribution zone of both *Pistacia* tree species.
in Israel, including on hybrid trees of \textit{P. atlantica} x \textit{P. palaestina} (Ben-Shlomo and Inbar, personal observations).

The East Mediterranean region, at the margin of the Eurasian and African plates, went through massive geological changes during the late Tertiary (66 to 2.6 Ma) and the succeeding Quaternary (Pleistocene) glaciation. These changes involved multiple events of land contacts and withdrawals (Lymberakis et al., 2007; and references therein). Such infrequent connections possibly facilitated occasional openings for species dispersal between the Irano-Turanian and the Mediterranean distribution zones.

The complex systematics and taxonomy of the genus \textit{Geoica}, and its unique and fragmented distribution on its \textit{Pistacia} host species (and hybrids), suggest that it may have speciated in the Middle East, in general, and in Israel, in particular, where we find allopatric and sympatric populations across different climatic regions. The main goal of this study was to trace the speciation processes of the genus \textit{Geoica} in the Levant by analyzing the phylogeny and genetic structure of the species complex in Israel, using several molecular markers (nuclear microsatellite loci and sequences of the mitochondrial COI and 12S genes). Specifically we assessed: 1. the genetic differentiation of the aphids (both between and within populations) between \textit{P. atlantica} and \textit{P. palaestina}, as well as the level of host plant specificity (i.e., are particular \textit{Geoica} aphids in sympatric populations found solely on a specific host \textit{Pistacia} species?); 2. the genetic diversity of the \textit{Geoica} aphids between populations within each species; 3. the association between gall morphology and aphid species on both \textit{Pistacia} hosts; and 4. the timeline of speciation according to the molecular clock, if speciation came about.

2. MATERIAL AND METHODS

2.1 Sampling:  
At the end of the summer (September–October) \textit{Geoica} galls were collected throughout their distribution range of \textit{P. atlantica} and \textit{P. palaestina} in Israel. In addition, we had several samples collected from \textit{P. atlantica} trees in various localities in Moab and Edom in Jordan (Fig. 1 and Table 1). The sampling (in total 141 galls) included allopatric and sympatric populations: 59 galls from \textit{P. palaestina}, 77 from \textit{P. atlantica} and 5 galls from a single hybrid tree (\textit{P. atlantica} x \textit{P. palaestina}).

![Figure 1: Distribution map and sampling sites in Israel and Jordan of \textit{P. atlantica} (square), \textit{P. palaestina} (circle) and sympatric populations of \textit{P. atlantica} and \textit{P. palaestina} (triangle).](image-url)
Table 1: Sampling of *Geoica* galls throughout the range of *Pistacia atlantica* and *P. palaestina* in Israel and Jordan. Numbers are following Fig 1.

<table>
<thead>
<tr>
<th>Country</th>
<th>Number</th>
<th>Population</th>
<th><em>Pistacia atlantica</em></th>
<th><em>Pistacia palaestina</em></th>
<th>Hybrid <em>palaestina x atlantica</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Israel</td>
<td>1</td>
<td>Mt Hermon</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Hazeka</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Gamla</td>
<td>8</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Meizar</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Tel Anafa</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Ein Avazim</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>Ramat Naftali</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>Sasa</td>
<td></td>
<td>13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>Dishon</td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>Nof Carmel</td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>Ada</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>Park Canada</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>Jerusalem</td>
<td>7</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>Adolam</td>
<td>10</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>Nizana</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>Lotz</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>Hazeva</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jordan</td>
<td>18</td>
<td>Bir Dabarat</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>Ja'far al-Tayyar</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>Hemed</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>Sela</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td><strong>77</strong></td>
<td><strong>59</strong></td>
<td><strong>5</strong></td>
</tr>
</tbody>
</table>

2.2 Gall morphology:
All galls are globular, completely sealed, and located on the abaxial side of the leaflet’s midvein. The size, shape and surface configuration of the galls vary considerably both within and among their host tree populations. The morphology of the galls was categorized according to the gall’s surface and shape characteristics: A. surface: smooth, partially crumpled and coarse; B. shape: ball-like, flattened discus-like, pear-like, “bumpy” spheres and coral-like (see photos in Table 2A). Gall characteristics were then associated with host plant species and the aphid’s genetic identity.

2.3 Genetic analysis of the aphids:
Due to their parthenogenetic reproduction, by the fall each gall may contain up to several hundred genetically identical aphids (Wool and Ben-Zvi 1998). DNA was extracted from the pool of aphids in each gall using the DNeasy Blood and Tissue Kit (Qiagen). The sequence of fragments of the mitochondrial genes *Cytochrome Oxidase I* (COI) and the ribosomal 12S subunits, as well as analysis of seven polymorphic microsatellite loci (Ivens et al., 2010), were used as molecular markers.

2.3.1 Mitochondrial COI and 12S sequencing: A fragment of 630bp of the mitochondrial COI gene (n=114 galls) and a fragment of 548bp of the mitochondrial 12S gene (n=78 galls) were
amplified by PCR in a total volume of 13 μl and at an annealing temperature of 55°C. The design of primers followed COI sequences found in the GenBank (AY227075; F- GTTGGGATAATGGTCTTC, and R- AGGCCAATT GTATTATAGC), and arthropods-conserved 12S primers (Kocher et al., 1989; 12SaI- AAACTAGATTAGATACCTATTAT and 12Sbi- AAGAGCGACGGCGATGT GT). The single PCR products of the amplified fragments were purified using the BigDye Terminator sequencing procedure (Applied Biosystems).

The sequences’ chromatograms were visualized, checked manually, and the least clear ends of each sequence (up to 40bp from each end) were excluded from the analysis. Sequences were aligned using the software MEGA 5.1 (Tamura et al., 2011). The chromatographs of all Single Nucleotide Polymorphism (SNPs) were re-checked visually, and the resulting phylogenetic dendrograms were clustered by the Maximum Likelihood method (1,000 bootstrap).

The COI phylogeny tree was tested for molecular clocks throughout the whole data set, with the published *Baizonga pistaic* (AY227079; Inbar et al. 2004) and *Slavum wertheimae* COI sequences (northern populations — GU391571; Negev Mountains — GU391572; Avrani et al., 2012) as outgroups. The Maximum Likelihood value for the given topology, with and without the molecular clock constraints, were compared using the Tamura-Nei (1993) model method (conducted in MEGA5.1: Tamura et al., 2011), where the null hypothesis tested was an equal evolutionary rate throughout the tree.

2.3.2 Microsatellites Analysis: 123 individuals from various populations and species of *Pistacia* were analyzed by seven known nuclear microsatellite loci (Gu-1, 3, 4, 6, 8, 11 and 13; Ivens et al., 2010). The F- primer of each microsatellite was labeled with fluorescent dye (6-Fam, Vic, Ned or Pet) and the amplification products were read by ABI 3130xl Fluorescence Reader (Applied Biosystems).

Microsatellites’ allele identification, genotyping and observed heterozygosity (Ho) were determined directly from the chromatographs using 500-Liz standard marker and Peak Scanner software (Applied Biosystems). Expected heterozygosity (He) and genetic identity (I) were calculated following Nei (1978). Data were analyzed by GenAlEx version 6.5b4 software (Peakall and Smouse, 2006). The analysis of molecular variance (AMOVA) procedure followed the methods of Michalakis and Excoffier (1996) using GenAlEx 6.5b4 software (999 permutations: Peakall and Smouse, 2006). Population clustering was performed using the Bayesian clustering method (STRUCTURE), which divided the samples into possible homogenous groups (subpopulations) according to their degree of genetic similarity (STRUCTURE 2.3.4 admixture model; burn-in of 100,000 steps and 100,000 iterations; five replicates for each K) (Pritchard et al., 2000, 2010). The inference of the probable number of clusters was extracted by the log likelihood for each putative number of populations (K), Ln P(D) = L(K), and by the delta K methods (Evanno et al., 2005), using the program Structure Harvester (Earl and vonHoldt, 2012).

3. RESULTS
3.1 Gall morphology:
Gall morphology was typified for 43 galls collected from *P. palaestina* and 26 from *P. atlantica* (Table 2). Ball-, flattened discus- and pear-like galls were found on both *Pistacia* species, whereas the exceptional coral-like galls, and the bumpy spheres ones were restricted only to *P. palaestina* and *P. atlantica*, respectively. The surface of the galls on *P. palaestina* were smoother (no coarse galls) than those found on *P. atlantica* (Table 2A). On *P. palaestina*, 74.4% of the galls had smooth skin and 25.6% were partially crumpled. On *P. atlantica*, 42.3% were smooth, 23% were partially crumpled and 34.6% were coarse. The shape of the galls on *P. palaestina* varied, with 30.2% shaped like balls, 23.3% like flattened discuses, 27.9% like pears, and eight
additional galls (18.6%), found only in one population, shaped like small “bushy corals” (see below). The shape of the galls on P. atlantica also varied, with 50% shaped like balls, 11.5% like flattened discuses, 34.6% like pears and 3.8% like “bumpy” spheres. Hence, neither the shape of the galls nor the surface were strictly associated with a Pistacia host plant species.

3.2 COI sequencing:
Cluster analyses (Maximum Likelihood, as well as Neighbor Joining) of 630bp of the mitochondrial COI gene clearly discriminated between the Geoica samples from the two Pistacia species (Fig. 2; at least 41 SNPs; >6.5%).

On P. atlantica, five Geoica clusters were noticeable: one haplotype (I) was found only in the xeric southern region of Israel and Jordan, and three haplotypes (II–IV) were found in the mesic central and northern regions of Israel. The fifth haplotype (V) was found only in two galls collected from Sela (Jordan). The geographic distribution of the southern haplotype (I) in Israel did not overlap with the others.

The distinction between the southern haplotype and the other three haplotypes was >26 SNPs (>4.1%). Among the three central-northern clusters, there was no indication of gall morphology, geographical or ecological differentiation, and the aphids of the various haplotypes (II–IV) were found in sympathy. The distinction between the three haplotypes was of at least 25 SNPs (>3.9%), with no indication of intermediate sequential substitutions. The Jordanian populations of Sela and Ja'far al-Tayyar harbored individuals from both the southern and the central-northern haplotypes. All five of the galls collected from the hybrid tree (P. atlantica x P. palaestina) presented haplotype IV of Geoica sp. of P. atlantica. No association was found between the various haplotypes and the morphology of the galls (Table 2B).

On P. palaestina, three major Geoica haplotypes were found (Fig. 2A). Haplotypes VI and VII differed by at least 16 SNPs (>2.5%). Both haplotypes were found sympatrically (also on the same tree) in most sampled populations, with the exception of the two northern populations, Mt. Hermon and Sasa, whose data pointed to only one haplotype of COI sequence (haplotype VI; Fig. 2B, C). No morphological association was found between the type of surface or shape of the galls and haplotypes VI or VII (Table 2B). A third haplotype (VIII) presented a distinct coral-like gall morph, differed from the other two haplotypes by at least 41 SNPs (>6.5%); it was found only in the remote area of Mt. Hermon.

3.3 12S sequencing:
The homologies between 78 sequenced samples were generally much higher than those found for COI, and the genetic distances between samples were lower in order of magnitude (overall mean genetic distance estimation was 0.015±0.006). Nonetheless, comparable to COI division (with the same individuals expressing the equivalent haplotypes; Fig. S1), Maximum Likelihood cluster analysis of 549bp of the mitochondrial 12S gene indicated the same major four branches (haplotypes I–IV) of Geoica on P. atlantica, the two clusters on P. palaestina (haplotypes VI–VII) and the distinct lineage of the coral-like morph (haplotype VIII) on Mt. Hermon. The sequence differences between the Geoica groups on P. atlantica were four SNPs (0.7%), and the two subgroups of P. palaestina differed by only one SNP.
Table 2:

A. The distribution of the various gall morphologies on *Pistacia atlantica* and *P. palaestina*

The shape of the gall

<table>
<thead>
<tr>
<th>Type of skin</th>
<th>Ball-like</th>
<th>Flatten “Discus”</th>
<th>Pear-like</th>
<th>“Bumpy” spheres</th>
<th>Coral-like</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>P. atlantica</em> smooth</td>
<td>4</td>
<td>---</td>
<td>6</td>
<td>1</td>
<td>---</td>
</tr>
<tr>
<td>partially crumple</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>coarse</td>
<td>6</td>
<td>1</td>
<td>2</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td><em>P. palaestina</em> smooth</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>---</td>
<td>8</td>
</tr>
<tr>
<td>partially crumple</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>coarse</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>

B. The COI haplotypes (I-VIII) of each gall morph

<table>
<thead>
<tr>
<th>Type of skin</th>
<th>Ball-like</th>
<th>Flatten “Discus”</th>
<th>Pear-like</th>
<th>“Bumpy” spheres</th>
<th>Coral-like</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>P. atlantica</em> smooth</td>
<td>I, II, IV</td>
<td>---</td>
<td>I, II, IV</td>
<td>III</td>
<td>---</td>
</tr>
<tr>
<td>partially crumple</td>
<td>I</td>
<td>1</td>
<td>1</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>coarse</td>
<td>II, IV</td>
<td>1</td>
<td>II, IV</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td><em>P. palaestina</em> smooth</td>
<td>VI, VII</td>
<td>V, VI</td>
<td>VII</td>
<td>V, VII</td>
<td>VIII</td>
</tr>
<tr>
<td>partially crumple</td>
<td>VI</td>
<td>VI</td>
<td>VII</td>
<td>V, VII</td>
<td>---</td>
</tr>
<tr>
<td>coarse</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>

A.
3.4 Microsatellites Analysis:
A preliminary run of all tested populations and host trees included clustering possibilities for 24 potential assemblies (K values from 1 to 24; the different groups x various localities; STRUCTURE 2.3.4 admixture model; burn-in of 100,000 steps and 100,000 iterations; five replicates for each K). The results suggested differentiation between Geoica on P. atlantica and P. palaestina (Fig. 3A). These results were in accordance with the sequencing results of the mitochondrial COI and 12S genes. Therefore, we further considered each of these two groups on different Pistacia species separately.

3.4.1 Geoica sp. from P. atlantica: Mean genetic diversity parameters indicated observed heterozygosity (Ho) of 0.272±0.035 and expected heterozygosity of 0.339±0.036. The inbreeding coefficient (Fis) was positive (0.203±0.198; Table S1), but did not significantly deviate from the Hardy-Weinberg equilibrium. AMOVA analysis indicated that 35% of the variance was within individuals, 53% between individuals and 12% between populations (Table S1). The average number of migrants between populations (Nm) per generation was 1.847.

STRUCTURE and Structure Harvester analyses (100,000 Burn-in period; and 100,000 iterations; five replicates for each K; K=1-16) of 74 samples indicated clustering to four Geoica major subgroups (Fig. 3B). While the northern populations (1–6) showed mixed genetic types, unique profiles characterized population 7 (Adolam), the Negev Mountains populations (9 and 10), and the Negev population of Hazeva (11). The Jordanian populations (12–15) mostly showed either the genetic structure of the northern populations, or the profile of Hazeva (11). The population of Jerusalem (8) had a mixed microsatellite genetic profile resembling either the northern population or the Negev populations.
Following the COI clustering results, we further analyzed the microsatellite differentiation between the four major COI haplotypes. AMOVA showed that 10% of the molecular variance was between haplotypes ($\PhiPT=0.103; P=0.02$). A Structure Harvester analysis (Earl and vonHoldt, 2012) indicated three major groups: one belonging to haplotype I, one to haplotype II and an additional one to haplotypes III and IV (Fig. 3C).

3.4.2 Geoica sp. from P. palaestina: The mean genetic diversity parameters of samples collected throughout the P. palaestina distribution range in Israel revealed relatively high levels of observed and expected heterozygosity ($Ho=0.461\pm0.054$; $He=0.562\pm0.053$), with no significant deviation from the Hardy-Weinberg equilibrium (Table S2). AMOVA indicated that 42% of the variance was within individuals, 47% between individuals and 11% between populations. The average number of migrants between populations (Nm) per generation was 2.126.

STRUCTURE and Structure Harvester analyses (100,000 Burn-in period; 100,000 iterations; five replicates for each K; K=1-7) suggested three major groups (Fig. 3D). However, the STRUCTURE profile within each population was mixed. The microsatellite differentiation did not indicate any association with the two COI haplotypes.

A. Geoica sp.
B. Geoica spp. on Pistacia atlantica

K=4

C. Geoica spp. on Pistacia atlantica

COI Haplotypes

K=3

Haplotypes I

Haplotype III+IV

Haplotype II

D. Geoica spp. on Pistacia palaestina

K=3


C. STRUCTURE differentiation of the Geoica microsatellite profile on P. atlantica, according to their COI haplotypes.


3.5 The molecular clock:

We tested for molecular clocks throughout the COI tree of the whole data set using the Maximum Likelihood method (Tamura and Nei, 1993; conducted in MEGA5.1), with B. pistaciae and S. wertheimae, sampled from the same trees as the Geoica samples, as outgroups. The analysis indicated the possibility of an equal evolutionary rate (the null hypothesis of an equal evolutionary rate throughout the tree was not rejected; P=1). We compared the two sequences that sowed the lowest genetic distance, one from Geoica found on P. atlantica and another on P. palaestina (Pa-Sela–4.1, haplotype V vs. Pp-Adolam-10c, haplotype VII), and between both sequences and the sequence of B. pistaciae. We found 57–58 SNPs between Geoica sp. and B. pistaciae (which parted ~34 Ma; Ren et al., 2013), and 41 SNPs between the two Geoica lineages. On this basis, we estimated the divergence of Geoica ancestries between the two Pistacia species as ~24-25 Ma (the Oligocene/Miocene boundary). Likewise, the divergence between the distinct coral-like gall haplotype (VIII) and the other two haplotypes on P. palaestina (41 SNPs) was also estimated as ~24-25 Ma. The divergence of the various haplotypes on P. atlantica (25–26 SNPs) was estimated as ~15 Ma (mid-Miocene), and between the two haplotypes of P. palaestina (16 SNPs), ~9.5 Ma.

4. DISCUSSION

4.1 Gall morphology:

Gall morphological traits, although of plant tissues, are induced and largely associated with the identity of the inducing insects. The generally globular and sealed morphological appearance on the leaflet midveins of various species of Pistacia trees is a clear characteristic of the galls induced by Geoica aphids. Blackman and Eastop (2018) implied that the color, type of surface and shape of the gall slightly differ between Pistacia host species. We noted that the gall’s color as well as the type of skin surface are less indicative, as they may have been altered among habitats, by the gall developmental stage, and with sun exposure (Inbar and Ben-Shlomo, personal observation). As shown here (Table 2), the shape of the gall was also not a distinctive trait distinguishing between and within P. atlantica and P. palaestina Geoica galls. The delicate morphological traits (i.e., surface and spherical characteristics of the galls) most likely depend upon and are controlled by the individual Pistacia tree genotype or physiological machinery. Interestingly, out of eight discrete haplotypes, only one Geoica haplotype (P. palaestina — haplotype VIII) can be clearly recognized according to its distinct coral-shaped morphology (Table 2). Indeed, these narrowly distributed galling aphids in northern Israel have distinct
morphological and molecular (see below) traits, which recently qualified it as a new species named *Geoica inbari* (Barjadze et al., 2022).

### 4.2 Speciation of the genus *Geoica* in the Levant

#### 4.2.1 The specificity of host-aphid interaction: In recent decades "DNA taxonomy" has become widespread in delimiting cryptic species (Fiser et al., 2018; Despres, 2019). Mitochondrial DNA, particularly a fragment of the *Cytochrome C Oxidase I* (COI) gene, is widely used as a "barcode marker" for identifying divergent lineages of insects. Attempts had been made to adopt an "objective" approach and to define species according to a sequence-divergent threshold (reviewed in Hebert et al., 2010; Krishnamurthy and Francis, 2012). The suggested rate of COI sequence-divergent threshold for insects is 3% divergent (Stoeckle and Hebert, 2008; Krishnamurthy and Francis, 2012), although there are reports of lower thresholds (e.g., several identified Lepidoptera species diverged as little as 1.3%; Hebert et al., 2010). Hence, to better assess the status of potentially cryptic species, multiple independent molecular markers are desirable (Fiser et al., 2018; Despres, 2019). Therefore, in addition to an analysis of the barcode COI marker, we analyzed a conservative mtDNA locus (12S), as well as seven known nuclear microsatellites.

The *Geoica-Pistacia* host-associated lineages clustered to entirely different clades in each and every genetic marker analyzed, both mitochondrial and nuclear. Even when both *Pistacia* species grow in sympatry, each consists solely of its own *Geoica* haplotypes. The explicit discrimination between *P. atlantica* and *P. palaestina* in both allopatric and sympatric populations suggest that the splitting of the genus *Geoica* between the two *Pistacia* host species was a single and early event. Despite the frequent landings of the sexupare on the "wrong" *Pistacia* (Wool et al., 1997), our results indicate that successful host shifts that require gall formation within this aphid complex is extremely rare, if at all.

Considering numerous molecular markers, Ren et al. (2013, 2017, 2019) estimated the divergent time of the Melaphidina, a related group of aphids (Fordini). Focusing on the gall-forming aphids on the genera *Rhus* and *Pistacia*, they suggested that the mean age of the common ancestor of the Fordiniae subtribe is 74–77 Ma. They estimated that within the *Pistacia* lineage, the divergence of the Baizongiini tribe occurred no later than ~65 Ma, and the split between the genera *Baizongia* and *Geoica* took place ~34 Ma (Eocene/Oligocene boundary). Using the Ren et al. (2013) chronogram as a base time, and assuming an equal evolutionary rate, we estimated the divergent time of *Geoica* sp., lineages between the two *Pistacia* host species as 24–25 Ma (the Oligocene/Miocene boundary).

#### 4.2.2 The co-evolution *Pistacia-Geoica*: The origin and evolution of the genus *Pistacia* are controversial. Zohary (1952) suggested that the genus evolved ~80 Ma; Kozhoidze et al. (2015) proposed a somewhat longer evolution of ~84 Ma based on global occurrence, geostatistical analyses, and environment-based probability distribution; while Xie et al. (2014), who sequenced two nuclear and seven plastid loci of the *Pistacia* species, estimated the divergence of *Pistacia* from its closest relative at 37–38 Ma, and the divergence between *P. atlantica* and *P. palaestina* at 9.42 Ma. Our finding indicating the divergence of *Geoica* ancestries between the two *Pistacia* species at 24–25 Ma supports the longer timescale for the evolution of *Pistacia*, allowing for diversification and/or co-diversification of the host-specific *Geoica* aphids and their *Pistacia* host ancestor(s). Following the divergence between *P. atlantica* and *P. palaestina*, a further monophyletic speciation of the genus *Geoica* occurred within each *Pistacia* host lineage. The succeeding speciation of *Geoica* on *P. atlantica* was more extensive (see below); hence, considering its complex biogeographical distribution, we cannot yet determine on which *Pistacia* species the speciation of the *Geoica* genus originated.

#### 4.2.3 Genetic differentiation of *Geoica* spp. among and within *P. atlantica* populations: *P. atlantica* is distributed along the Irano-Turanian region, one of the largest floristic regions of the world. However, its evolutionary history and biogeography is less understood (reviewed in
Manafzadeh et al., 2017). The collision and the deformation of Central Asian geology, as well as the complex and heterogeneous geological history, most likely vastly affect the speciation of various species inhabiting this region (Manafzadeh et al., 2017). Our results of five discrete COI haplotypes (differing by at least 25 SNPs; >3.9%), supported by the divergence of 12S, as well as the microsatellite profiles, suggest that at least five distinct Geoica species radiated on *P. atlantica* in the Levant. The high level of sequence divergence and the corresponding branching out of both mitochondrial and nuclear loci suggest a long-standing splitting to different varieties, estimated to take place during the mid-Miocene (~15 Ma) — a phase (Kurschner et al., 2008; Manafzadeh et al., 2017) of markedly lower the CO$_2$ level, associated with both a temperature drop and major glaciation (Kurschner et al., 2008).

The heterogeneous geological history of the Irano-Turanian zone (Manafzadeh et al., 2017), together with long-lasting climatic fluctuations, triggered repeated changes in the distribution of *P. atlantica*, and consequently of *Geoica* aphids, with possible isolated populations left in refugia islands of suitable habitats (Yom-Tov and Tchernov, 1988; Danin, 1999). The lasting patchy distribution may well facilitate the accumulation of mutations, leading to further speciation of the *Geoica* complex on *P. atlantica*.

Nowadays the distribution of *P. atlantica* in Israel reflects a discontinuous spread from a mesic climate in northern and central regions to a xeric environment in the south. One of the five *Geoica* haplotypes, haplotype I, is restricted to xeric habitats. The Israeli Negev Desert is particularly dry with annual precipitation lower than 100 mm of rain (Amiran et al., 1970), and has a geographically isolated *P. atlantica* population, perhaps a relict. The distribution of the *P. atlantica* population in the Negev is currently shrinking due to increasing aridity over the last thousand years. Consequently, the populations of aphids (similar to the population of trees) are smaller and more isolated. The *Geoica* haplotype I was found only in this xeric ecoregion and no other haplotype was found in the Negev Desert. The extreme ecological conditions, in concert with the locally isolated population, are promoting rapid speciation in the Negev. Similarly, distinct allopatric genetic haplotypes and cryptic speciation were detected in the closely related gall-forming aphid *Slavum wertheimae* (Avrani et al., 2012) and *Forda riccobonii* (Ben-Shlomo and Inbar, unpublished data) on the same *P. atlantica* populations in the Negev Desert. Interestingly, *P. atlantica* does not show an equivalent genetic separation between these two ecogeographic regions (Avrani et al., 2012).

Three additional *Geoica* haplotypes were found sympatrically, also on the same tree, in the central and northern mesic regions of Israel (with over 600 mm mean annual precipitation). The existence of two *Geoica* species in northern-central Israel was noticed by Brown and Blackman (1994), who described *G. rungsi* and *G. harpazi*, which differ in their alatae female morphology. Based on the sequences of the COI and COII genes, Inbar et al. (2004) also suggested two *Geoica* species on *P. atlantica*. Unfortunately, we cannot match our results to the reported species and defined haplotypes of Brown and Blackman (1994). Notably, the two published COI sequences (AY227075 and AY227085; Inbar et al. 2004), which differ by 12 SNPs, are both clustered within the species that represents haplotype III, suggesting the possibility of yet more complex radiation. The last haplotype, V, is rare and was found only in two individuals from Sela, Jordan (population number 21, Fig. 1).

The evidently current sympatric allocation of at least three different species in northern and central Israel is probably a result of allopatric speciation, with secondary sympathy, that may have been pushed further by competition (Inbar 1998). Following the last glaciation cycle, when the climate warmed once more, the *P. atlantica* population expanded and became continuous again, bringing about the *Geoica*-galling aphids to present extensive “sympatric” allocation. The extreme and long (15 Ma) separation, together with the extensive genetic differentiation, suggest that these five haplotypes should be named as new species. Analyses of the genetic structure of *Geoica* spp. on a wider geographic scale in Central Asia and North Africa should shed light on the evolutionary history and origin of the various haplotypes on *P. atlantica*.
4.2.4 Genetic variation between and within P. palaestina–Geoica populations: P. palaestina is widely distributed in the Mediterranean zones of Israel (Ben-Shlomo and Inbar, 2012) and is much less fragmented than the populations of P. atlantica. Furthermore, the distribution of this species is restricted to mesic habitats. This factor may explain why P. palaestina hosts fewer Geoica haplotypes than P. atlantica. Nonetheless, further sympatric speciation of Geoica sp. with three distinct COI haplotypes was found on P. palaestina. One unique Geoica species (G. inbari, haplotype VIII), which differs from the other two haplotypes by at least 41 SNPs (>6.5%), presents a clear altered form of the gall with a coral-like morphology, rather than a globular one. G. inbari (Haplotype VIII) puts forward an extensive ancient separation (~24–25 Ma). This distinct species is rare, and found only in the far and high Mt. Hermon.

The other two COI haplotypes differ by more than 2.8% (at least 20 SNPs) of the sequence. The 12S sequencing results showed differentiation of only one substitution, hence the allies in each group were in accordance with the COI results. While finding clear differentiation in the mitochondrial genome, the nuclear microsatellite profile of the two haplotypes looked alike. We suggest that the two haplotypes represent two cryptic species of Geoica. Comparable discordance between mitochondrial and nuclear diversity patterns between cryptic species are commonly found (Despres, 2019; Hinojosa et al., 2019), as the two genomes respond in a different manner to the demographic and spatial event resulting from the maternally transmitted monoploid mtDNA and the meiotic segregation and recombination of nuDNA. One of these two species may be G. wertheimae, reported as the Geoica species found on P. palaestina. Three sequences of Geoica wertheimae are deposited in the Genbank: Accessions AH012643, AY227080, AY227090 - "M Wink 18450" (Inbar et al., 2004); EU701670 - voucher CNC#HEM012658 (Foottit et al., 2008); and DQ499622 (Zhang and Qiao, 2007). The first two cluster within the lineage of Geoica on P. palaestina presented here (Fig. S2). As the sequence of Foottit et al. (2008) was identical to most samples of haplotype VI, we regard haplotype VI as Geoica wertheimae and suggest that haplotype VII should be named as yet another Geoica species.

5. CONCLUSIONS
The galling-aphid genus Geoica went through extensive radiation in the Levant. The splitting of the genus between P. atlantica and P. palaestina is estimated to have occurred 24–25 Ma (the Oligocene/Miocene boundary). Five different haplotypes suggesting five different species have been further speciating among Geoica spp., galling on P. atlantica, and an additional three species, on P. palaestina.

ACKNOWLEDGEMENTS
We thank MOFET Institute for financial support and the students of Molecular Ecology workshop for their help.
REFERENCES


Giron, D., Huguet, E., Stone, G.N., Body, M., 2016. Insect-induced effects on plants and possible effectors used by galling and leaf-mining insects to manipulate their host-plant. J. Insect Physiol. 84, 70–89.


https://doi.org/10.1007/s12686-010-9293-3


https://doi.org/10.1111/ecog.01496.


