Hierarchical structure in the *Drosophila mojavensis* cluster (Diptera: Drosophilidae)

D. C. S. G. OLIVEIRA[1], E. H. WINTERMUTE[1], W. J. ETGES[2], W. B. HEED[3], P. M. O'GRADY[1,4] and R. DESALLE[1]

[1] *Division of Invertebrate Zoology, American Museum of Natural History, New York, NY, USA.*
[2] *Department of Biological Sciences, University of Arkansas, Fayetteville, AR, USA.*
[3] *Department of Ecology and Evolutionary Biology, The University of Arizona, Tucson, AZ, USA.*
[4] *Department of Biology, University of Vermont, Burlington, VT, USA.*

Oliveira, D. C. C. G., Wintermute, E. H., Etges, W. J., Heed, W. B., O'Grady, P. M. and DeSalle, R. 2003. Hierarchical structure in the *Drosophila mojavensis* cluster (Diptera: Drosophilidae). — *Hereditas 139*: 223–227. Lund, Sweden. ISSN 0018-0661. Received July 14, 2003. Accepted November 4, 2003

Sequences from three gene regions from the nuclear and mitochondrial genomes were used to examine the extent and nature of hierarchical structure in the *Drosophila mojavensis* cluster (*Drosophila arizonae*, *D. mojavensis* and *D. navojoa*) of the *D. repleta* species group. To determine the genetic divergence of these three species, sequence data were analyzed using maximum parsimony and population aggregation analysis. Individual and combined gene genealogies indicate that *D. arizonae* and *D. mojavensis* are neither diagnosable nor monophyletic with respect to one another. Although *D. navojoa* has differentiated from *D. arizonae* and *D. mojavensis*, as diagnosed by nuclear gene sequences, it may have undergone a reticulation event with *D. arizonae*. Our results suggest that either these taxa are still undergoing differentiation at the molecular level or have experienced gene flow in the recent past.

Rob DeSalle, Division of Invertebrate Zoology, American Museum of Natural History, Central Park West @ 79th Street, New York, NY 10024, USA. E-mail: desalle@amnh.org

The *Drosophila mojavensis* cluster is endemic to the arid lands and deserts of the southwest United States and Mexico and includes *D. mojavensis*, *D. navojoa* and *D. arizonae* (RUIZ et al. 1990). This cluster is part of the *D. mulleri* subgoup of the *D. repleta* species group (VILELA 1983; WASSERMAN 1992). These taxa form a well-supported monophyletic group with respect to other *D. mulleri* subgroup flies and are very closely related to each other. *Drosophila navojoa* is easily identified morphologically and is considered basal to the other two species (RUIZ et al. 1990). *Drosophila mojavensis* and *D. arizonae* are sibling species and the *D. arizonae* status as valid species has been maintained mainly based on cytological differences (RUIZ et al. 1990; WASSERMAN 1992). These two species produce fertile offspring in the laboratory (RUIZ et al. 1990), however hybrids of these taxa have not been collected in the wild (WASSERMAN and KOEPFER 1977; ETGES et al. 1999). The *mojavensis* cluster is, therefore, a model for speciation studies and has been the subject of chromosomal inversion, reproductive isolation and ecological studies (RUIZ et al. 1990; WASSERMAN 1992; ETGES and JACKSON 2001). Our goal in this communication is to use nucleotide sequences to characterize the genetic variation and hierarchical structure for these three currently recognized species.

MATERIAL AND METHODS

One mitochondrial (CoII) and two nuclear (*Marf* and *snf*) gene regions (Table 1) are used to examine the genetic changes that have occurred during the differentiation of the three *D. mojavensis* cluster species (RUIZ et al. 1990). In total (Table 2), we analyze 27 individuals of *D. mojavensis* (from 3 different localities), 6 individuals of *D. navojoa* (from 3 different localities) and 5 individuals of *D. arizonae* (from 2 different localities). Two strains of *D. mulleri*, one of *D. huaylasi* and the triad of sibling species in the *D.*

Table 1. *Gene regions used in this study.*

Gene name	Primers	Size (bp)	PI[a]	Functional class
CoII	O'GRADY et al. 1998	679	111	cytochrome oxidase subunit II, mitochondrial
Marf	BONACUM et al. 2001	621	53	mitochondrial assembly regulatory factor, nuclear
Snf	BONACUM et al. 2001	370	33	sans fille, nuclear

[a] Number of parsimony informative characters.

Table 2. *Collection information and GenBank accession number for the taxa examined.*

Species	CoII	*Marf*	*snf*	Source
D. arizonae				
Ar1.1	AY437288	AY437333	AY437375*	Jalisco, Mexico[1]
Ar1.2	AY437289	AY437334	AY437376	
Ar1.4	AY437290	AY437335	AY437377	
Ar2.1	AY437291	N/A	AY437378	Mexico
Ar2.2	AY437292	AY437336	AY437379	
D. mojavensis				
A993.1	AY437255	AY437300	AY437343	Sonora, Mexico[2]
A993.2	AY437256	AY437301	AY437344	
A993.3	AY437257	AY437302	AY437345	
A993.4	AY437258	AY437303	AY437346	
A993.5	AY437259	AY437304	AY437347	
A993.6	AY437260	AY437305	AY437348	
A993.7	AY437261	AY437306	AY437349	
A993.8	AY437262	AY437307*	AY437350	
A993.9	AY437263	AY437308	AY437351	
A993.10	AY437264	AY437309*	AY437352	
A993.11	AY437265	AY437310	AY437353	
A993.12	AY437266	AY437311	AY437354	
A993.13	AY437267	AY437312	AY437355	
A993.14	AY437268	AY437313	AY437356	
A993.15	AY437269	AY437314	AY437357	
A993.16	AY437270	AY437315	AY437358	
A993.17	AY437271	AY437316*	AY437359	
A993.18	AY437272	AY437317	AY437360	
A993.19	AY437273	AY437318	AY437361	
A993.20	AY437274	AY437319	AY437362	
A993.21	AY437275	AY437320	AY437363	
A993.22	AY437276	AY437321*	AY437364	
A993.2a	AY437277	AY437322	AY437365	
A998.1	AY437278	AY437323	AY437366	California, USA[3]
A998.2	AY437279	AY437324	N/A	
A998.3	AY437280	AY437325	AY437367	
Moj2.2	AY437281	AY437326	AY437368	Mexico
D. navojoa				
Nav1	AY437282	AY437327	AY437369	Sonora, Mexico[4]
Nav6.4	AY437283	AY437328	AY437370	Sonora, Mexico
Nav6.5	AY437284	AY437329*	AY437371	
Nav7.1	AY437285	AY437330	AY437372	Sonora, Mexico
Nav7.3	AY437286	AY437331	AY437373	
Nav7.5	AY437287	AY437332	AY437374	
D. huaylasi				
Huaylasi	AY437299	N/A	AY437386	Peru[5]
D. mulleri				
Mu1.4	AY437293	AY437337*	AY437380	Cayman Islands[6]
Mu2.4	AY437294	AY437338	AY437381	Haiti[7]
D. mayaguana				
4.071	AY437298	AY437342	AY437385	Dominican Republic[8]
D. parisiena				
Pa 11.1	AY437295	AY437339	AY437382	Jamaica[9]
1.485	AY437296	AY437340	AY437383	Jamaica[10]
D. straubae				
St6.1	AY437297	AY437341	AY437384	Haiti[11]

Partial sequences (*) are included only in the combined analysis (Fig. 1).
1 – Tomatlan, Jalisco, Mexico; 26.vii.1981; A806; W. B. Heed. Ambrose Monell Cryo Collection barcode 102045 (American Museum of Natural History, NY). 2 – Rancho El Diamonte, Rt. 16, 82 km east of Hermosilo, Sonora, Mexico; 20.iv.1996; A993; W. Etges, G. Huckins and C. Durando. 3 – Power line Road, west of Havasu City, north of Whipple Mts., Arizona, Unites States; xi.1995; W. Etges, G. Huckins and P. O'Grady. 4 – Navojoa, Sonora, Mexico; E2.1. Ambrose Monell Cryo Collection barcode 102941 (American Museum of Natural History, NY). 5 – Quives, Peru; Hu-1. Ambrose Monell Cryo Collection barcode 109209 (American Museum of Natural History, NY). 6 – Cayman Brac, Cayman Islands, 26.xi.1983; A927, ORV27; W. Johnson and Benado. 7 – Gonaives, Haiti; A942. 8 – 27 km NE of Barahona Dominican Republic; A983. 9 – Port Henderson, Jamaica, 23.xi.1983; ORV24; W925; R. H. Thomas and W. B. Heed. Ambrose Monell Cryo Collection barcode 102332 (American Museum of Natural History, NY). 10 – Airport road, Kingston, Jamaica; A980. 11 – Fond Parisien, Haiti, 7.v.1982; ORV1, W901; M. Wasserman and W. B. Heed. Ambrose Monell Cryo Collection barcode 102352 (American Museum of Natural History, NY).

mayaguana subcluster, *D. mayaguana*, *D. straubae* and *D. parisiena* (2 individuals) are included in the analysis as outgroups (WASSERMAN 1992; DURANDO et al. 2000).

DNA was isolated from individual wild caught flies using the DNeasy Tissue Kit (Qiagen). PCR amplification and DNA sequencing were performed as in REMSEN and O'GRADY (2002). Sequences were edited using Sequencher software (Gene Codes Corporation). The alignment generated in Clustal X 1.8 (THOMPSON et al. 1997) was adjusted by eye using MacClade 4.0 (MADDISON and MADDISON 2000). One gene region, *Marf*, is largely non coding and has several indels. The gaps were coded as presence/absence characters by the "simple indel coding" method of SIMMONS and OCHOTERENA (2000). This procedure added 23 characters, 19 of them parsimony-informative, to the *Marf* matrix and a single parsimony-informative character to the *snf* matrix (not included in Table 1). Complete DNA matrices and the gaps coded as characters can be seen at http://research.amnh.org/molecular/sequence1.html.

We performed both individual and combined analyses in a maximum parsimony (MP) framework using PAUP* 4.0 (SWOFFORD 2002). Settings for MP analyses were as follows: search type = heuristic, addition sequences = random, number of replicates = 500, branch swapping = TBR. Support at each node was assessed using bootstrap proportions (FELSENSTEIN 1985, 1988) and the jackknife (FARRIS et al. 1996) with 200 bootstrap or jackknife replicates (other settings as above). Jackknife analyses were done with 25% character deletion. Population aggregation analysis (PAA) was used as outlined in DAVIS and NIXON (1992) to search for diagnostic nucleotides. The sequences were viewed in MacClade with the matchchar option, which facilitates visualization of characters in PAA. Sequences divergence were assessed using the pairwise distance command in PAUP*4.0.

RESULTS AND DISCUSSION

Tree based patterns at the species boundary

A single *D. arizonae* individual (marked with an * in Fig. 1 and 2) appears as sister group to *D. navojoa* in the combined analysis. This result is due entirely to the CoII sequence. The same individual is not closely related to *D. navojoa* in any other analyses (Fig. 2). This observation suggests either directional hybridization or incomplete lineage sorting of mitochondrial DNA. Since the two nuclear loci show that *D. navojoa* is monophyletic, this result implies that these loci have coalesced. The mitochondrial genome should have coalesced approximately four times faster than nuclear genes because of smaller effective population sizes (HUDSON 1990). Taken together, the data and theory suggest that the placement of this enigmatic *D. arizonae* mtDNA haplotype within *D. navajoa* is due to introgression between the two species. This is surprising because crosses between these two species in the laboratory resulted in sterile offspring (RUIZ et al. 1990).

Both individual (Fig. 2) and combined (Fig. 1) analyses indicate that *D. arizonae* and *D. mojavensis* are not monophyletic with respect to each other. Interestingly, two *D. mojavensis* individuals (a, b; Fig. 1 and 2) are connected with four *D. arizonae* in the CoII analysis, but these individuals are not related to *D. arizonae* in any nuclear analysis. Similarly, one *D. mojavensis* (c, Fig. 1 and 2) has the same *snf* haplotype

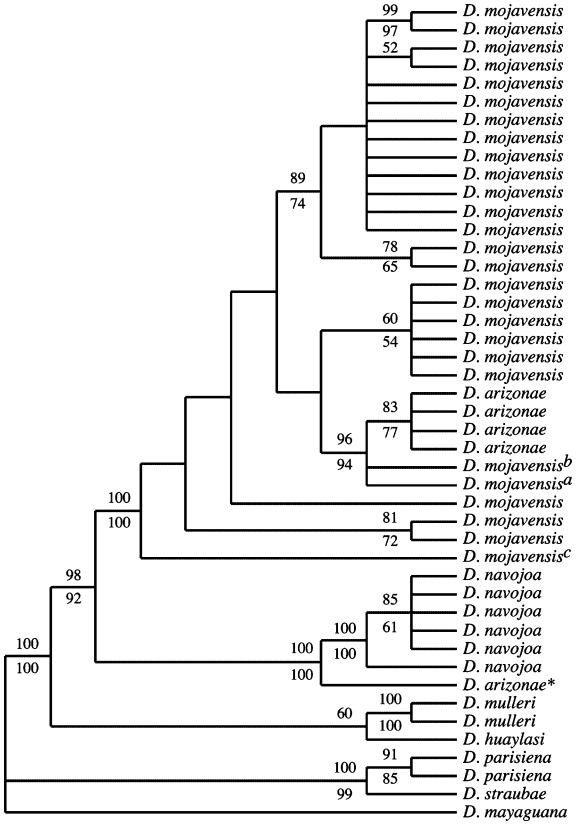

Fig. 1. Strict consensus of the most parsimonious trees of the combined analysis. Some individuals are marked, *, a, b or c, in order to indicate their position in the single gene trees in Fig. 2, see text for details. Values of support (when > 50%) for each node are indicated, jackknife with 25% deletion (above) and bootstrap (below). Most parsimonius trees = 64,073, score = 477, consistency index = 0.6897, retention index = 0.8849. The maxtrees command in Paup*4.0 was set to 100,000 for bootstrap and Jackknife analysis.

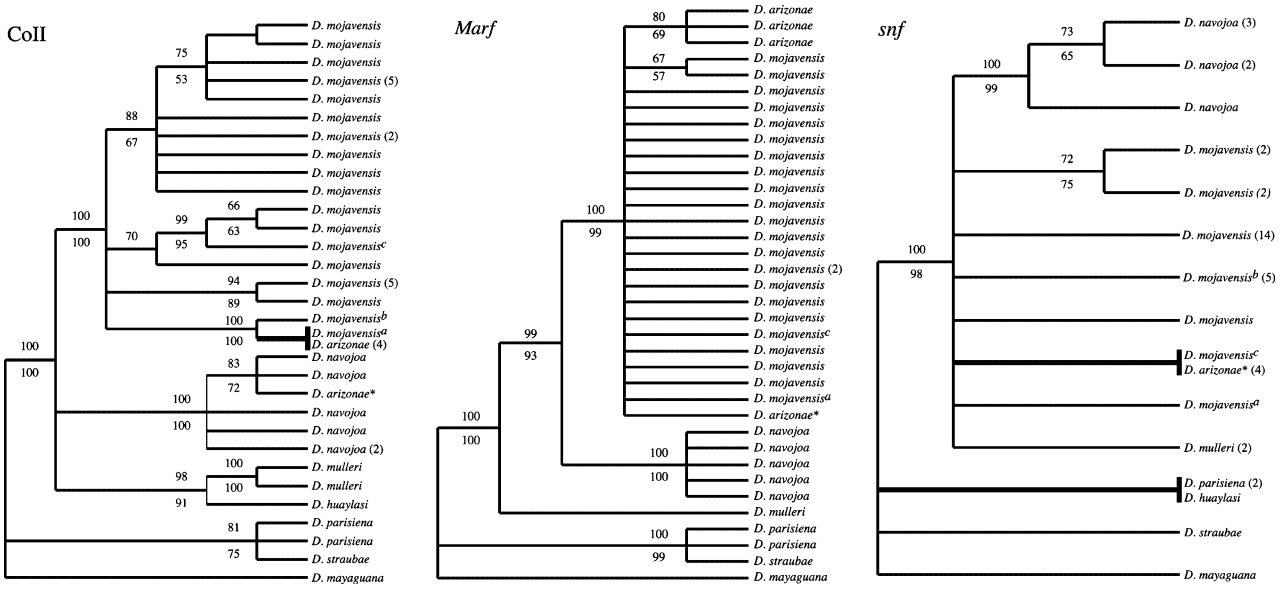

Fig. 2. Strict consensus of the most parsimonious trees for the three gene partitions used in this study. The terminals represent a single haplotype, the number of individuals that posses certain haplotype is shown in parenthesis. The thicker lines highlight when a haplotype is present in more than one species. Some individuals are marked, *, A, B or C, in order to indicate their position in the other single gene threes and in the combined analysis in Fig. 1, see text for details. Values of support (when > 50%) for each node are indicated, jackknife with 25% deletion (above) and bootstrap (below). The maxtrees command in Paup*4.0 was set to 100,000 for bootstrap and jackknife analysis. CoII: most parsimonious trees (MPT) = 148, score = 220, consistency index (CI) = 0.7318, retention index (RI) = 0.9135; *Marf*: MPT > 250,000, score = 144, CI = 0.7986, RI = 0.8914; *snf*: MPT = 12, score = 51, CI = 0.9020, RI = 0.9000.

as *D. arizonae*. Again, the placement of these individuals is indicative of either incomplete lineage sorting for the gene regions we have examined or gene flow among these taxa in the wild. While both scenarios are possible, these species are of recent origin, as evidenced by morphology and polytene chromosome similarities and these species maintain large population sizes. Such conditions could place the coalescent time for many genes prior to the speciation event (HUDSON 1990). This explanation seems particularly likely for the nuclear gene *snf*, due to the small number of haplotypes (Fig. 2) and low nucleotide variation. Although incomplete sorting of the ancestral polymorphism can explain the observed pattern of DNA sequence variation for the *D. mojavensis* – *D. arizonae*, the possibility of reticulation cannot be excluded.

Sequences divergence

Pairwise distances show that the intraspecific variation is always lower than 1.5% for *Marf* and lower than 1.0% for *snf*. The same is true for the divergence between *D. mojavensis* and *D. arizonae*. For the CoII region the *D. mojavensis* haplotypes can be up to 2.5% divergent from the haplotypes found in *D. arizonae* and also in *D. mojavensis*[a] or *D. mojavensis*[b]. The sequences of *D. navojoa* vary from the sequences of *D. mojavensis* and *D. arizonae* by 3% for *Marf* and *snf*. Divergence between *D. navojoa* (*Drosophila arizonae** as well) and the other two species reaches 9% for CoII.

Table 3. *PAA generated nuclear diagnostic sites for D. navojoa. Numbers at top of table refer to the aligned nucleotide positions (for complete matrices see http://research.amnh.org/molecular/sequence1.html). Only diagnostic positions are shown. Gap regions were not included.*

	Marf	*snf*
	1 2 3 3 4 4 5 5 5 5 5	1 1 2 2 2 2 2
	5 3 8 0 0 1 2 1 4 6 7 9	3 6 4 7 0 1 2 3 9
	7 1 3 0 3 2 8 8 9 0 9 3	3 9 7 7 6 0 7 3 4
D. mojavensis/D. arizonae	C C C T A K C C A K A T	C T K G C G T A G
D. navajoa	T A T C T C G T C C T C	T C A A T T G G C

Species diagnostics in nuclear gene sequences

The results of population aggregation analysis (DAVIS and NIXON 1992) are shown in Table 3 and demonstrate a number of sites in the two nuclear genes that diagnose *D. navojoa* as distinct from the other two species in the cluster. There are no sites in the two nuclear genes or in CoII gene sequences that diagnose *D. mojavensis* or *D. arizonae* as distinct, consistent with the maximum parsimony analysis (Fig. 2). The lack of diagnosis for the mtDNA sequences can be traced to the mtDNA haplotype of a *D. arizonae* individual with nucleotides changes shared with *D. navojoa*.

CONCLUSIONS

Analysis of the patterns of hierarchy at the species boundary for the species triad of *D. navajoa*, *D. arizonae* and *D. mojavensis*, suggests that two major factors have, in the past and will continue in the future, to influence the differentiation of these three species. First, if *D. arizonae* and *D. mojavensis* are indeed good species, then incomplete lineage sorting appears to be a strong factor in confounding hierarchy at the DNA level. Second, the detection of a *D. arizonae* individual that has a *D. navajoa* mtDNA haplotype, even in the small sample sizes we use here, indicates that hybridization in the wild might be playing a role in the evolution of these two species.

REFERENCES

Davis, J. I. and Nixon, K. C. 1992. Populations, genetic variation and the delineation of phylogenetic species. – Syst. Biol. 41: 421–435.

Durando, C. M., Baker, R. H., Etges, W. J., Heed, W. B., Wasserman, M. and DeSalle, R. 2000. Phylogenetic analysis of the repleta species group of the genus Drosophila using multiple sources of characters. – Mol. Phylogenet. Evol. 16: 296–307.

Etges, W. J., Johnson, W. R., Duncan, G. A., Huckins, G. and Heed, W. B. 1999. Ecological genetics of cactophilic *Drosophila*. – In: Robichaux, R. (ed.), Ecology of Sonoran Desert plants and plants communities. Univ. of Arizona Press, Tucson, pp. 164–214.

Etges, W. J. and Jackson, L. L. 2001. Premating isolation is determined by larval rearing substrates in cactophilic *Drosophila mojavensis*. VI. Epicuticular hydrocarbon variation in *Drosophila mojavensis* cluster species. – J. Chem. Ecology 27: 2125–2149.

Farris, J. S., Albert, V., Kallersjo, M. et al. 1996. Parsimony jackknifing outperforms neighbor-joining. – Cladistics 12: 99–124.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. – Evolution 39: 783–791.

Felsenstein, J. 1988. Phylogenies from molecular sequences: inference and reliability. – Annu. Rev. Genet. 22: 521–565.

Hudson, R. 1990. Gene genealogies and the coalescent process. – Oxf. Surv. Evol. Biol. 7.

Maddison, D. R. and Maddison, W. P. 2000. MacClade: analysis of phylogeny and character evolution, version 4.0. – Sinauer Associates Sunderland.

Remsen, J. and O'Grady, P. M. 2002. Phylogeny of Drosophilidae (Diptera), with comments on combined analysis and character support. – Mol. Phylogenet. Evol. 24: 249–264.

Ruiz, A., Wasserman, M. and Heed, W. B. 1990. Evolution of the mojavensis cluster of cactophilic *Drosophila* with descriptions of two new species. – J. Heredity 81: 30–42.

Simmons, M. P. and Ochoterena, H. 2000. Gaps as characters in sequence-based phylogenetic analysis. – Syst. – Biol. 49: 369–381.

Swofford, D. L. 2002. PAUP*: phylogenetic analysis using parsimony (and other methods), version 4.0 beta. – Smithsonian Institution.

Thompson, J. D., Gibson, T. J., Plewniak, F. et al. 1997. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. – Nucleic Acids Res. 24: 4876–4882.

Vilela, C. R. 1983. A revision of the *Drosophila repleta* species group (Diptera: Drosophilidae). – Rev. Bras. Entomol. 27: 1–114.

Wasserman, M. 1992. Cytological evolution of the Drosophila repleta species group. – In: Krimbas, C. B. and Powell, J. R. (eds), *Drosophila* inversion polymorphism. CRC Press, pp. 455–552.

Wasserman, M. and Koepfer, H. R. 1977. Character displacement for sexual isolation between *Drosophila mojavensis* and *Drosophila arizonensis*. – Evolution 31: 812–823.