

ON "THE EVOLUTIONARY GENETICS OF SOCIAL WASPS"  
AND THE PHYLOGENY OF THE VESPINAE  
(HYMENOPTERA, VESPIDAE)

J.M. CARPENTER

*Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138, U.S.A.*

Reçu le 16 juin 1986

Accepté le 12 janvier 1987

SUMMARY

VARVIO-AHO *et al.* (1984) incorrectly analyzed their data on allozymes for eight species of European yellowjackets. Correct analysis produces 50 trees of lower percent standard deviation than their Fitch-Margoliash tree, and three additional Wagner trees of the same length as their network. Construction of consensus trees shows their data to be relatively uninformative on relationships of these vespine species. Their suggestion that *Dolichovespula* is not monophyletic cannot be upheld.

RESUME

Sur la « Génétique évolutive des Guêpes sociales » et la Phylogénie des Vespinae  
(Hymenoptera : Vespidae)

VARVIO-AHO *et al.* (1984) ont fait une analyse incorrecte de leurs données sur des allozymes de huit espèces de guêpes d'Europe du Nord. Une analyse correcte produit cinquante arbres d'une déviation standard d'un pourcent plus basse que leur arbre de Fitch-Margoliash, et trois arbres de Wagner supplémentaires de même longueur que leur réseau. La construction des arbres de consensus indique que leur données ne sont pas très informatives sur les parentés des espèces de vespines. Leur suggestions que *Dolichovespula* ne forme pas un groupe monophylétique n'est pas soutenable.

INTRODUCTION

VARVIO-AHO *et al.* (1984) presented an allozyme data set for eight species of European yellowjackets: *Dolichovespula media*, *norwegica*, *omissa*, *saxonica* and *sylvestris*, *Vespula* (*Vespula*) *austriaca* and *rufa*, and *Vespula* (*Para-*

*vespula) vulgaris*. They used 38 prevalent electromorphs from 13 isozymes in phylogenetic analysis. First they calculated a distance matrix between the species using Nei's genetic distance, and clustered on this matrix using UPGMA, LI's (1981) technique and the Fitch-Margoliash technique. They then recorded the electromorph data into a binary data matrix and used this to produce a Wagner network. From their results they suggested that *Dolichovespula* and perhaps *Vespula s.l.* (i.e., *Paravespula* + *Vespula*) were not monophyletic groups. I will show here that their conclusions rest upon faulty analysis, and that their data are too ambiguous to be phylogenetically useful. Available character data refute their conclusions; these are treated in CARPENTER (1987).

## METHODS

All analyses were performed using the Physys system by J.S. FARRIS and M.F. MICKEVICH, State University of New York at Stony Brook and Maryland Center for Systematic Entomology, as implemented on the VAX 11/780 computer running VMS at Harvard. Nei's distances were as presented in *table IV* of VARVIO-AHO *et al.* For the pairings of *Dolichovespula omissa* with *Vespula rufa* and *austriaca*, which shared no alleles, the infinity value was replaced by the greatest distance found between any other species pair (a distance of 2.54 between *D. saxonica* and *V. rufa*). This was the procedure of VARVIO-AHO *et al.* (p. 379). For the presence-absence coding of electromorphs, the 38 electromorphs listed in their *table III* were treated as binary data, with unscored loci coded as missing. Reanalysis of these data is described below.

## RESULTS

### 1. Distance analysis

The first three techniques applied by VARVIO-AHO *et al.* all produced the same tree (here reproduced as *figure 1*), which is unsurprising, since they are all variations on UPGMA. UPGMA and LI's method do not have an optimality criterion *per se*, but if it is desired to measure how well the phenograms fit the original distance data, the cophenetic correlation coefficient is used. VARVIO-AHO *et al.* provided neither clustering levels nor fit statistics for the phenograms. They only stated (p. 384) that the "root of the UPGMA tree is at the distance  $D = 1.92$ ". This is a misprint; the basal clustering level is actually 1.829. I have recalculated the phenogram using the UPGMA routine of Physys, and indicate the clustering levels in *figure 1*. The cophenetic correlation is .826 as measured by the fit option. The Fitch-Margoliash technique does have an optimality criterion: percent standard deviation (% SD). As it is usually done, one simply calculates a phenogram, then recalculates the branches so as to minimize the % SD statistic for the tree.

VARVIO-AHO *et al.* used the program Phylip (version 2.2) by J. FELSENSTEIN to calculate their Fitch-Margoliash tree. This program presents results as a

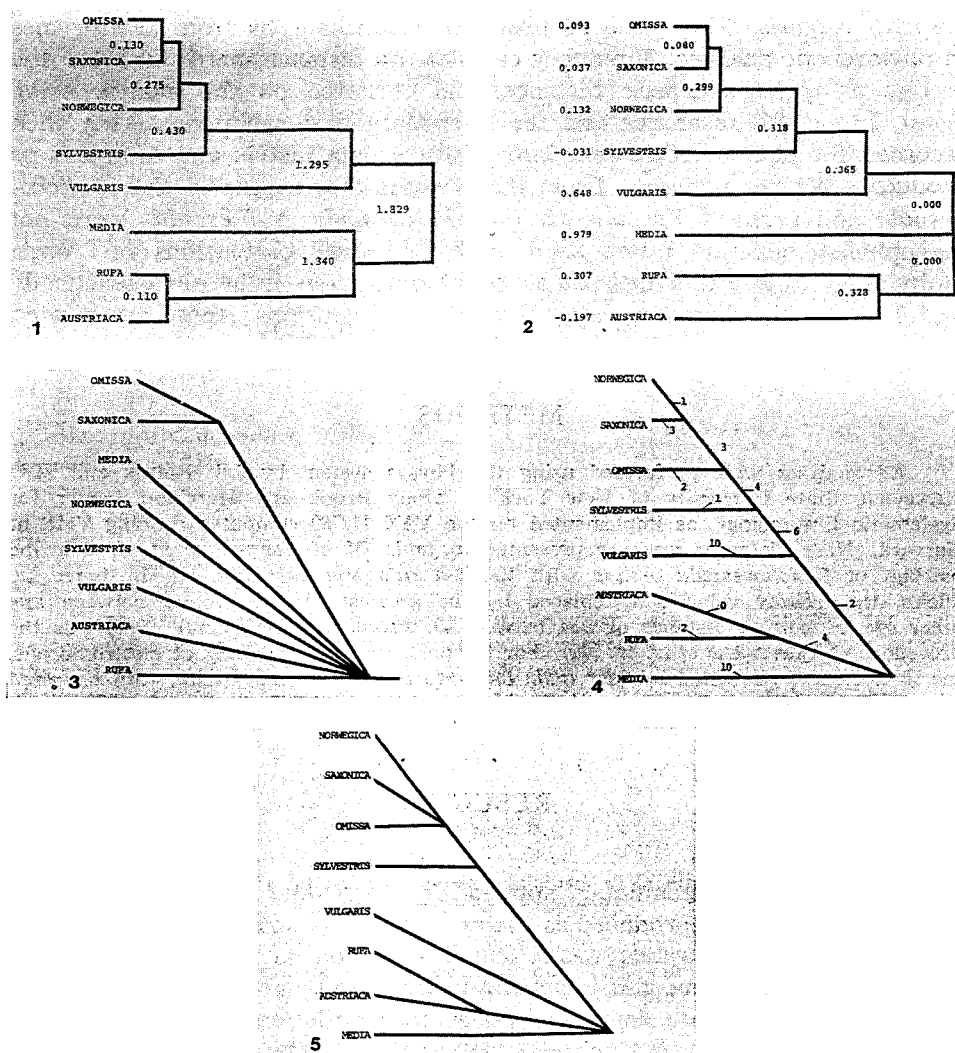


Fig. 1. — UPGMA phenogram for Nei's distance.

Fig. 2. — Best-fitting Fitch-Margoliash tree.

Fig. 3. — Consensus tree of 24 best-fitting Distance Wagner trees.

Fig. 4. — Wagner network for presence-absence coding of electromorphs.

Fig. 5. — Consensus tree of Wagner trees.

Fig. 1. — Phénogramme UPGMA à partir de distances de Nei.

Fig. 2. — Arbre meilleur-ajustant de Fitch-Margoliash.

Fig. 3. — Arbre de consensus de vingt-quatre arbres meilleur-ajustant de Wagner.

Fig. 4. — Réseau de Wagner à partir du code de présence-absence d'électromorphes.

Fig. 5. — Arbre de consensus des arbres de Wagner.

network, and indicates both branch lengths and "average percent standard deviation", but VARVIO-AHO *et al.* again provided neither clustering levels nor the fit statistic, % SD value. I have calculated the best-fitting Fitch-Margoliash tree with the same topology as theirs (*fig. 2*) using the Psdopt routine; its % SD is 20.493. Note that there are two negative branch lengths. VARVIO-AHO *et al.* did not comment on the uninterpretability of negative branch lengths and the problems these present for analysis of distance data (FARRIS, 1972, 1981, 1985, 1986). As is well known (FARRIS, 1981), negative branch lengths are often required to optimize % SD for nonmetric distances such as Nei's. Phylip 2.2's version of the Fitch-Margoliash procedure defaults to permit only nonnegative branch lengths, and P. PAMILO (*in litt.*) stated that VARVIO-AHO *et al.* did not consider negative branch lengths. The % SD value they actually obtained, 38.3 (P. PAMILO, *in litt.*), was therefore not close to optimal either for these data or this topology.

Better % SD values are achieved with the Distance Wagner technique than the Fitch-Margoliash procedure, provided that the branch lengths in the Distance Wagner analysis are selected so as to optimize % SD (FARRIS, 1981). *Inter alia*, this amounts to permitting negative branch lengths. Better fit still may be achieved if multiple trees are sought (FARRIS, 1985). The Pwagner routine of Physys was used to calculate Distance Wagner trees with low % SD, with the default number of 50 trees calculated. These range in % SD from 17.726 to 25.666. All 50 trees fit better than the Fitch-Margoliash tree of VARVIO-AHO *et al.*, and all have negative branch lengths. Of the 50 trees, 24 have lower % SD than the best-fitting tree with the same topology as that of VARVIO-AHO, ranging from 17.726 to 20.282 (the cophenetic correlation on these trees ranges from .917 to .888, better than either the UPGMA phenogram or the % SD optimized version of this tree). The best-fitting tree differs from the second best-fitting tree by a value of .012. There is something like a gap (1.098) between the third and fourth best-fitting trees, but the only pronounced gap, 2.749, occurs between the twenty-fourth and twenty-fifth best-fitting trees.

PRAGER and WILSON (1978) advocated presenting information from multiple trees as a simplified (less resolved) tree, and POST and UZZELL (1981) restricted attention to multiple trees separated from others by a pronounced gap in goodness-of-fit. FARRIS *et al.* (1982) and FARRIS (1985) used consensus trees to indicate the information on grouping common to multiple trees below a pronounced gap. The Adams consensus tree for the 24 best-fitting trees is shown as *figure 3*. It has just one informative group (this is also the strict consensus tree). The Adams consensus and the strict consensus for all 50 trees are also the same. They are completely unresolved. The distance data are so ambiguous as to be virtually worthless, a result shared with many other distance data sets (FARRIS, 1985 and pers. comm.).

The unsuitability of Nei's distances for analysis by branch length fitting

is well known (FARRIS, 1981, 1985, 1986), but deserves further comment. This distance is nonmetric, which obviates any interpretation of clocklike divergence (FARRIS, 1981), and for these data, the departure from metricity is marked. When the Mtest routine of Physys was used to test for violations of the triangle inequality, 18 were found, out of 56 possible triplets. Non-metric distances cannot be interpreted as amount of evolutionary change when analyzed by branch length fitting (FARRIS, 1981), nor even as the statistical expectation of amount of change (FARRIS, 1985, 1986), and so it is unknown what these distances represent. Restriction of analysis to only trees with nonnegative branch lengths is therefore completely unjustified, and this is particularly true of trees that fit an ultrametric to the distances, as in UPGMA. This was done by VARVIO-AHO *et al.* because they assumed a molecular clock for Nei's distances (p. 376), but since nonmetric distances cannot be truly clocklike (FARRIS, 1981, 1985, 1986), it sacrifices fit for nothing. If distances are to be used in phylogenetic analysis, as when no other data are available, they should be properly analyzed. That is, phenetic techniques should not be used, because that amounts to forcing a clocklike interpretation even when the data do not support it. But use of distances when other data are available, as is true of Nei's distance where the original electromorphs may be analyzed as characters, is a mistake. It discards information present in the original character data. Nei's distance should be abandoned.

## 2. Wagner analysis

VARVIO-AHO *et al.* also used Phylip to calculate their Wagner network. It differed from the phenograms only in the relative relationships of *Dolichovespula norwegica*, *omissa* and *saxonica*. Phylip is relatively deficient for Wagner analysis (LUCKOW and PIMENTEL, 1985). Aside from poorer performance of the Phylip Wagner procedure than that of other widely available Wagner programs (LUCKOW and PIMENTEL, 1985), the output of version 2.2 consists only of a network, total length, and the number of changes in each character. Branch lengths are not provided; the user must estimate these by hand. VARVIO-AHO *et al.* (p. 381) discussed estimation of the branch lengths, but did not indicate the results of their calculation nor give the total length of the network. The assignment of character states and estimation of branch lengths are ambiguous for these data. I have calculated a set of values for their topology with the Diagnose routine, and present the result here as *figure 4*, using media as reference point. As measured by LFIT, the length is 50, with consistency of .76.

The XWAGNER routine of Physys was used to perform an exact solution via a branch and bound algorithm for all most parsimonious trees. There are four distinct trees for these data, all of length 50, consistency .76. VARVIO-AHO *et al.* thus found only one of the Wagner trees for their data. *Figure 5* is the consensus tree (both ADAMS and NELSON) for all the possible

Wagner trees. The data are ambiguous: only three informative groups are present.

## DISCUSSION

VARVIO-AHO *et al.* stated (p. 383) "the various phylogeny-generating techniques gave congruent results" and "there are some evident patterns which we can discuss reliably". As shown above, neither claim is correct. Instead, for the distances no clear patterns of any kind are shown: these data are so ambiguous as to be virtually uninformative. The presence-absence data are also ambiguous, and to the extent that they show any pattern, the results of the various techniques are not congruent with each other.

VARVIO-AHO *et al.* stated (p. 383) "The observed genetic heterogeneity within *Dolichovespula*, caused by *D. media*, indicates that this group may not be monophyletic but consists of those yellowjackets which have retained preclude monophyly. But to the extent that this suggestion is consistent with analysis of their presence-absence data falling to show *Dolichovespula* as a some primitive characters". Of course plesiomorphy in any given feature in itself does not preclude monophyly. But to the extent that this suggestion is consistent with analysis of their presence-absence data falling to show *Dolichovespula* as a group (*cf. fig. 4*), it is not supported by the other possible trees for these data. The basal multifurcation in the consensus tree (*fig. 5*) is not inconsistent with monophyly of *Dolichovespula*. It is simply uninformative about this. Further, morphological and behavioral data are available which strongly uphold the monophyly of *Dolichovespula*. CARPENTER (1987), in the first comprehensive cladistic analysis of the vespine subgenera, cites no fewer than 12 autapomorphies of this genus.

VARVIO-AHO *et al.* claimed (p. 384) "One important conclusion from our genetic studies is that the two social parasites examined originate from the same evolutionary lineages as their hosts. This renders baseless those taxonomic subdivisions which separate the parasitic species into (sub) genera different from their hosts". Although *austriaca* and *rufa* clustered together, and *omissa* clustered near (not with) its host *sylvestris*, the taxonomic claim is illogical. These parasites were separated categorically from their hosts because of their differences from the hosts, which would still exist regardless of the lineages the parasites originated from. These differences are autapomorphies (CARPENTER, 1987) associated with parasitism, and the separation is therefore certainly fallacious, but on the grounds that it renders the categories of the hosts paraphyletic.

Proponents of the molecular clock, such as VARVIO-AHO *et al.*, like to claim that their approach is superior because (p. 376) "If the characters used to construct the phylogeny are subject to selection pressures, the conclusions will be biased because of either convergent or divergent evolution".

This can only mean that homoplasy will mislead investigators using the phylogenetic criterion of parsimony, but as FARRIS (1983) points out, abundance of homoplasy in itself is no grounds for adopting some other approach. The cited claim also implies that selection in fact will cause results to be misleading, which can scarcely be true generally. Advocacy of the molecular clock comprises some further, unstated claims, namely that evolution at the biochemical level has in fact proceeded only homogeneously, which is false as shown for example by the data discussed in this paper (see FARRIS, 1981, 1985 for citations of additional data sets showing heterogeneity), and that heterogeneity could never be brought about in the absence of selection, which also cannot be true. In this case, adoption of the molecular clock led VARVIO-AHO *et al.* into specious results. It is to the credit of phylogenetic methods that the weaknesses of a priori assumptions may be clearly seen - they may be contradicted by the data which had seemed to corroborate them, when these data are analyzed without the unrealistic assumptions.

ACKNOWLEDGMENTS. — I thank P. PAMILO, J.S. FARRIS, and R. HONEYCUTT for commenting upon the manuscript. Use of the computer was supported by a Milton Fund Grant and NSF Grant BSR-8508055 to the author. Jenny C. CORNELL helped prepare the figures, and Laurie BURNHAM and Kathy HOY shared their knowledge of French.

#### References

- CARPENTER J.M., 1987. — Phylogenetic relationships and classification of the Vespinae (Hymenoptera : Vespidae). *Syst. Ent.* (In press).
- FARRIS J.S., 1972. — Estimating phylogenetic trees from distance matrices. *Am. Nat.*, 106, 645-668.
- FARRIS J.S., 1981. — Distance data in phylogenetic analysis. In *Advances in Cladistics*. Proceedings of the first meeting of the Willi Hennig Society. Ed. V.A. Funk, D.R. Brooks, pp. 3-23. New York Bot. Gard., Bronx.
- FARRIS J.S., 1983. — The logical basis of phylogenetic analysis. In *Advances in Cladistics 2*. Proceedings of the second meeting of the Willi Hennig Society. Ed. N.I. Platnick, V.A. Funk, pp. 7-36. Columbia Univ. Press, New York.
- FARRIS J.S., 1985. — Distance data revisited. *Cladistics*, 1, 67-85.
- FARRIS J.S., 1986. — Distances and statistics. *Cladistics*, 2, 144-157.
- FARRIS J.S., KLUGE A.G., MICKEVICH M.F., 1982. — Immunological distance and the phylogenetic relationships of the *Rana boylei* species group. *Syst. Zool.*, 31, 479-491.
- LI W.O., 1981. — Simple method for constructing phylogenetic trees from distance matrices. *Proc. Nat. Acad. Sci.*, 78, 1085-1089.
- LUCKOW M., PIMENTEL R.A., 1985. — An empirical comparison of numerical Wagner computer programs. *Cladistics*, 1, 47-66.
- PRAGER E.M., WILSON A.C., 1978. — Construction of phylogenetic trees for proteins and nucleic acids: empirical evaluation of alternative matrix methods. *J. Molec. Evol.*, 11, 129-142.
- POST T.J., UZZELL T., 1981. — The relationships of *Rana sylvatica* and the monophyly of the *Rana boylei* group. *Syst. Zool.* 30, 170-180.
- VARVIO-AHO S.L., PAMILO P., PEKKARINEN A., 1984. — Evolutionary genetics of social wasps (Hymenoptera, Vespidae, *Vespula*). *Insectes Soc.* 31, 375-386.