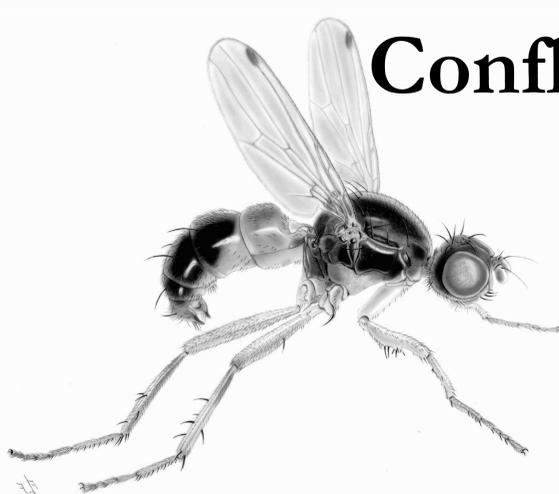


# Conflict and Congruence Between DNA And Morphology In Insect Systematics

Lim Shimin Gwynne Supervisor: A/P Meier, Rudolf

Department of Biological Sciences, National University of Singapore S2 #02-01, Science Drive 4, Singapore 117543



## Introduction

Conflict and congruence between molecular and morphological sources of information are important topics in contemporary systematics. At the taxonomic level, widespread sepsid species are particularly useful for studying conflict because of the large spatial and temporal isolation of populations. At the phylogenetic level, congruence and conflict between DNA and morphology in phylogeny studies can be quantified using Partitioned Bremer Support (PBS).

## Widespread Sepsid Species Populations: Native Or Introduced?

Sepsid species are closely associated with humans, being saprophagous on cattle dung, which may lead to secondary introductions due to human transcontinental traffic. This can be assessed by looking at pairwise divergences of multiple samples from different populations.

Out of the 253 specimens extracted, 135 yielded usable *COI* sequences. These were added to a preexisting 56-sequence dataset, totaling 191 sequences from 61 morphological species. Large genetic distances consistent with sepsid populations being native were found for *Sepsis neocynipsea* (2.44%, Turkey/US), *S. thoracica* (4.73%, Africa/China), *S. punctum* (5.87% Germany/China), *S. dissimilis* (3.29% Comoros/Thailand), *S. flavimana* (6.13% Germany/US), *Australosepsis niveipennis* (3.12% Malaysia/Comoros) and *Themira biloba* (5.94% Denmark/US). It appears that native populations have adapted and exploited cow dung as a food source as domesticated cattle spread into their species range. Two notable exceptions include *T. putris* and *T. minor*, where pairwise divergences between Nearctic and Palearctic populations were very low and did not show regional clustering.

## DNA Taxonomy

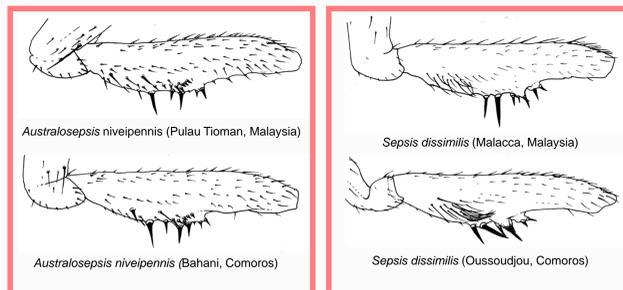
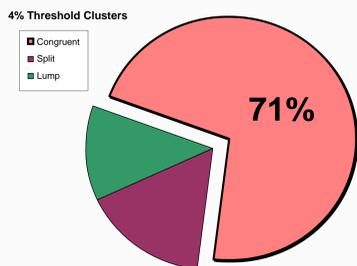
Some researchers have proposed that species boundaries can be determined by fixed percentage pairwise distances. Here, we test this by using DNA divergence thresholds to assign species names to morphologically pre-sorted species.

Table 1: Range of threshold differences and identification success

	1% Threshold	2% Threshold	3% Threshold	4% Threshold	5% Threshold
Proportion Congruent	59%	60%	66%	71%	67%
Proportion Split	30%	24%	19%	16%	15%
Proportion Lumped	11%	16%	15%	13%	19%

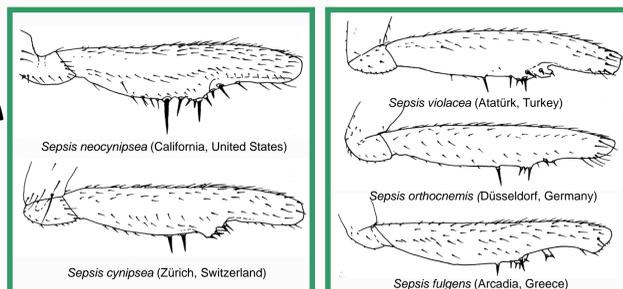
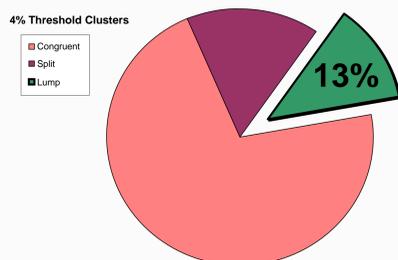
### Scenario 1: Congruence

71% of morphological species were recognised by a 4% threshold. These species can be effectively delimited by either method.



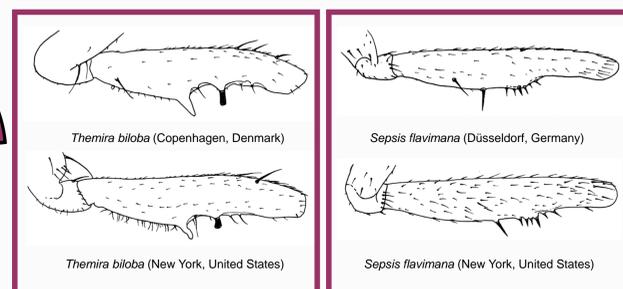
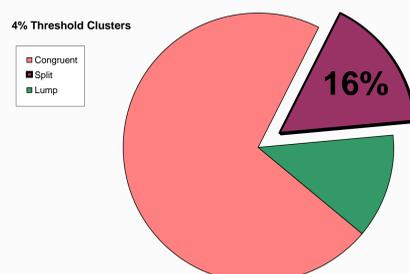
### Scenario 2: Incongruence – Lumped Species

If identifications were incongruent, tests of reproductive compatibility were conducted when possible in order to externally validate current species limits. Using the same 4% criteria, 13% of morphologically distinct sister species were merged.



### Scenario 3: Incongruence – Split Species

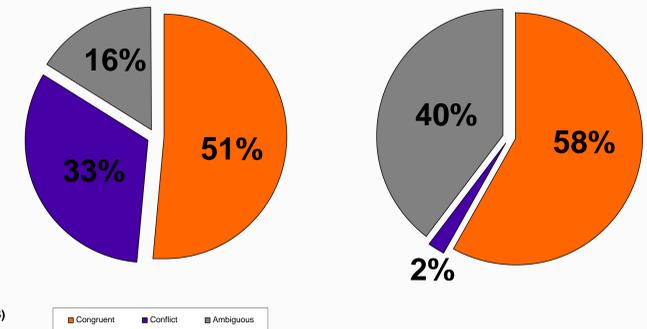
16% of morphologically indistinguishable species were split into multiple species.



## DNA Barcoding

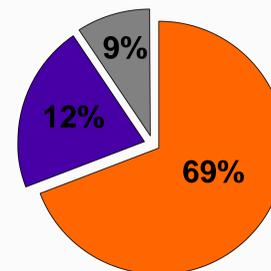
The recent surge of interest in DNA barcoding has highly polarised systematic research, with many researchers offering various critiques of the method. DNA barcoding aims to identify unknown species using a barcode *COI* sequence. We test this method of identification by using a neighbour-joining (NJ) tree and pairwise distances.

NJ Tree-Based Species Delimitation (Hebert *et al.*, 2003) NJ Tree-Based Species Delimitation (Meier *et al.*, 2006)



1. NJ Tree-Based: This follows the methods of Hebert *et al.* (2003: Proc. R. Soc. Lond. B.), which uses a monophyletic assemblage containing the query sequence as a criterion of identification success and Meier *et al.* (2006: Syst. Biol.), which requires the query sequence be at least one node into a conspecific clade, or in a conspecific polytomy.

Best Close Match Based Species Delimitation (Meier *et al.*, 2006)



2. Pairwise Distance-Based: Best Close Match (Meier *et al.*, 2006: Syst. Biol.) uses a threshold level of intraspecific genetic divergence to assemble all generated *COI* sequences into clusters, where species representation can be assessed. For this study, the threshold was 3%

**BOLD-IDS**: Sequences were also input into the full database of BOLD-IDS, an identification database for the Barcode of Life System (Ratnasingham and Hebert, in press: Mol. Eco. Notes) in order to identify species based on sequence similarity measurement. Only 22% of sequences were successfully identified with no species-level match for the rest, most likely due to inadequate sampling within Sepsidae in the database.

## Conflict In Phylogeny Reconstruction

Differences in molecular and morphological information may result in conflicting signal in phylogenetic analyses. This can be quantified by assessing relative node support using Partitioned Bremer Support (PBS), which apportions relative support of all defined partitions (in this case morphological and molecular information) on the most parsimonious tree or consensus of all most parsimonious trees.

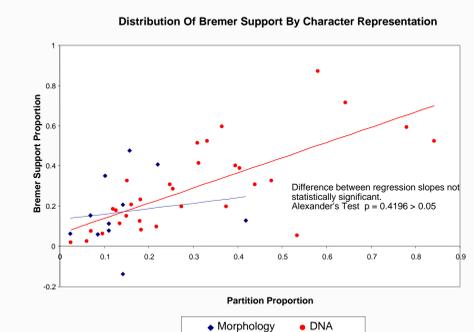
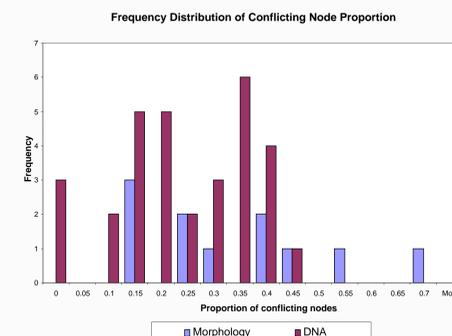
11 datasets representing studies on Coleoptera, Diptera, Lepidoptera, Hymenoptera, and Hemiptera were obtained. PBS was performed and partitions assessed for differences in node conflict as well as absolute and relative support between morphological and DNA information.

### 1. Node Conflict

No significant difference in proportion of node disagreement was found. (morphology average: 31%, DNA average: 22%. 2-tailed homocedastic t-test  $p > 0.05$ )

### 2. Strength of support

Due to large (85%) size, DNA partitions provided three times more support. However no significant difference was found for per-character relative support. (morphology average: 0.8, DNA average: 0.73%. 2-tailed heterocedastic t-test  $p > 0.05$ )



## Summary

We found widespread conflict between morphology and DNA in widespread species taxonomic identification.. However, there was no significant evidence for difference between both categories of information at the phylogenetic level.