Material & Methods

1. Sources of datasets

We acquired the datasets using morphological characters that include colour/pattern characters in the analyses from various sources of journal and book articles. In total 58 morphology-based phylogenetic datasets were obtained from either the journal websites or the authors and were assembled to investigate the "behaviour" of colour pattern characters in phylogenetic reconstruction (Table 1). Among these datasets, 46 data matrices were already included in Areekul & Quicke's study.

2. Partitioning of datasets

2.1 Strategy I

In the present study, the datasets were first partitioned into five categories based on phylogenetic predictions of character distributions involving different circumstances of aposematism and mimicry: A. the taxa of the clade being analysed do not exhibit any aposematic or mimetic colouration, but their colouration can possibly contribute to crypsis or camouflage (Fig. 1A); B. the taxa of the clade being analysed exhibit aposematic but not mimetic colouration, and the distribution of aposematic characters can be either random or concentrated (Fig. 1B); C. multiple mimetic forms occur in related or unrelated taxa in the clade and each of the mimicry forms is participated by the species not included in the phylogeny being analysed, and distribution of the mimetic characters tend to be random or concentrated (Fig. 1C); D. multiple unrelated taxa

included in the phylogeny being analysed participate in the same mimicry complex, and distribution of the mimetic characters tend to be random (Fig. 1D); and E. one mimicry complex is participated by related taxa forming a monophyletic group, so the mimetic characters support the monophyly of the taxa involved in the mimicry complex (Fig. 1E). Each dataset was further divided into two subsets, “colour/pattern” and “non-colour/pattern”, based on the attribute of each character included in the data matrix. The colour/pattern subset includes, for instance, the characters which are described with colours, and the characters involving stripes, spots, lines or patches. The datasets used in the present study cover from 3% to 72% of colour/pattern characters (see Table 1)

2.1 Strategy II

To compare the effect of different dataset partitioning strategies on detecting phylogenetic influence of colour/pattern characters using the protocol shown in Fig. 2, we also attempted employing the partitioning method of Areekul & Quicke (2006). First, the 3 categories (C, D and E) were combined into one, termed category C+D+E, which involves all different circumstances of mimicry. However, the category B (aposematism) was retained to avoid confusion in character interpretation involving aposematism and mimicry. Under this partitioning strategy, 3 dataset categories were obtained. The same analysis were then proceeded as described above.

3. Tree-searching criteria & character fit

Each of the datasets was reanalysed with all characters set as the same situations in their respective original studies (i.e. weighting scheme and character ordering). All the phylogenetic analyses were implemented with the complete datasets (simultaneously analysed colour/pattern and non-colour/pattern characters) and performed by PAUP*, version 4.0b10 (Swofford, 2000). In order to obtain trees fairly similar to the most parsimonious trees (MPTs) in a reasonable time, even though it would not generally find most parsimonious trees with extensive datasets, we followed Areekul & Quicke's (2006) strategy to set heuristic search (HS) with 1000 random addition and then ran tree bisection reconnection (TBR) branch swapping for each dataset in equal weighted to get MPTs, and we applied the same strategy on subsequently weighted analyses. We then separately mapped the colour/pattern and non-colour/pattern character partitions onto the MPT(s) from the analysis of the complete original dataset, and calculated the ensemble retention indices (RIs) for each partition of each equal-weighted (EW) trees.

4. Assessments of effect of colour/pattern characters

4.1 Our idea of reassessment

Comparing the up-weighting protocols proposed by Østergaard *et al.* (2003) and Areekul & Quicke (2006), respectively, we found some differences in the "trees input AST method" (see the next paragraph for further explanation) between them. Østergaard *et al.* (2003) computed a strict consensus tree of for each equal or up-weighted analysis, and then preceded agreement subtree of "consensus tree from equal weighted data

matrix" and each "consensus tree from up-weighted data matrix" (see Østergaard *et al.*, 2003: 309, Fig.4). On the other hand, Areekul & Quicke (2006) did not compute any consensus tree of result tree populations for each weighting analysis, but they kept tree populations and then randomly picked pairs of trees from equal and unequal weighted analyses for agreement subtree method. They obtained the numbers of taxa remained in ASTs. The former protocol shows a problem that the resolution of strict consensus tree may decline while there is aarge number of original trees.

Therefore, we followed the protocol of Areekul & Quicke (2006) (Fig. 2) to detect the effects of colour/pattern characters. We employed weighting method to observe the influence of each partition on topologies, and compared the "degree of consensus" between topologies from serial weighting data matrix and equally weighted ones by agreement subtree(s) (ASTs). The maximum agreement subtree approach is a method designed for reconciling different evolutionary hypotheses (trees) for the same set of taxa. An agreement subtree maintains a subset of the taxa for the equivalent restricted subtree (under a suitable definition) in all given trees (Amir & Keselman, 1997). The more taxa remain in ASTs, the more assentient between the weighed partition and the original equal weighted data matrix.

The following strategy is designed to detect whether one partition carries more outstanding signals than the other, and contributes more influence in tree reconstruction. If one partition carries stronger apparent signals

than the other or contains less noise, data constituting this partition may fit better onto the MPT from simultaneous analysis of both partitions and be dominant in the combined analysis. Both the number of characters in one partition and the strength of the intrinsic phylogenetic signal per character can affect the dominancy of one data partition over the other. A series of differential re-weighting schemes was employed to test the potential dominancy of one partition over the other one, and then compare the eventual trees with those obtained from equally weighted analysis by the AST method suggested by Areekul & Quicke (2006). MPTs from analysis of up-weighted dominant partitions will be more similar to those from equally weighted analysis, and the strength of dominance could therefore be assessed by examining how the numbers of taxa remain in the agreement subtree(s) (ASTs) was affected by differential weighting of colour/pattern and non-colour/pattern characters.

The proportion of taxa remained in ASTs of each reweighed scheme was then charted to a broken-line graph (Fig. 4), and the symmetry of these lines could show the degree of conflict or consensus between colour/pattern and non-colour/pattern characters partitions.

4.2 Measuring relative signal strength of partitions

Heuristic search was run as described above with both data partitions simultaneously, differential weighted scheme was applied by assigning weight to colour/pattern characters of 1.25, 1.5, 2, 5, 10 and 1000 x relative to non-colour/pattern ones and vice versa (see Fig. 3), in total, 12

re-weighted analyses were performed for each data matrix.

MPT(s) were generated from each differentially (asymmetrically) re-weighted scheme analysis, and then form a group of trees population. We employed the AST function implemented in PAUP*(version 4.0b10) (Swofford, 2002) for estimating tree similarities between equally weighted (EW) and asymmetrically weighted (AW) trees. If MPTs calculated from each analysis resulted in more then ten trees, or there were more than ten possible pair combinations, ten randomly selected pairs of individual trees from EW and AW populations were compared by AST method (Østergaard *et al.*, 2003), and then calculated the mean of the range of AST values. Eventually, we obtained twelve mean AST values of AW comparison for each data matrix (see Fig. 2).

The degree of asymmetry of the mean percentage of taxa remained in ASTs under differential re-weighed schemes showed by broken-line graph (Fig. 4) was measured to assess the level of real conflict between the two apparent phylogenetic signal partitions originating from each data partition or due to different amounts of noise (Dolphin *et al*., 2000; Areekul & Quicke, 2006). If the two partitions of one data matrix carry equally strong apparent phylogenetic signals, the plot of the mean percentage of taxa remained in ASTs (from differential re-weighted and equally weighted MPTs) versus the up-weighting factor will form a symmetric reverse V shape line (Areekul & Quicke, 2006). The more conflict the two partitions are, the more eminent the reverse V shape line is. If this plot shows an asymmetric line as a glacis, which indicates the

two partitions carry different strength of phylogenetic signal and the proportion of taxa remained in ASTs of the weaker partition will decline more rapid with the degree of up-weighting (Areekul & Quicke, 2006).

Asymmetry in plots of the proportion of taxa remained in ASTs in each category was measured employing the Wilcoxon matched pairs test to compare the mean ratios of taxa remained in ASTs from the non-colour/pattern up-weighted analyses with those from equivalently colour/pattern up-weighted analyses. The ratios of taxa remained in ASTs within different weights were compared between the five categories by Kruskal-Wallis test, and Dunn's Multiple Comparisons Test was further performed as post-test to find the specific pairs which caused the significant variation while the p-value was smaller than 0.05.