

Ecography

**ECOG-00444**

Dapporto, L., Ramazzotti, M., Fattorini, S., Talavera, G., Vila, R. and Dennis, R. L. H. 2013. recluster: an unbiased clustering procedure for beta-diversity turnover. – *Ecography* 36: xxx–xxx.

**Supplementary material**

## Appendix 1

Here we describe how the analyses can be carried out by using the recluster package. The package can be downloaded from the Comprehensive R Archive Network (CRAN) at <http://cran.r-project.org/web/packages/recluster/index.html>.

Manuals for functions are provided in the CRAN web site and a pdf can be downloaded at <http://cran.r-project.org/web/packages/recluster/recluster.pdf>

First of all it is necessary to install the package on R 3.0.0 (or later) and then open it together with the model dataset.

```
library (recluster)
data(datamod)
data(treemod)
```

As a preliminary test we inspect the dissimilarity matrices of turnover for both faunistic and phylogeographic beta-diversity, computed as follows for the model data:

```
# Faunistic dissimilarity

simpdiss <- recluster.dist(datamod)

# Phylogenetic dissimilarity

phylosortdiss <- recluster.dist(datamod, treemod, dist="phylosort")

# The distribution of values can be inspected and plotted

simpdiss
recluster.hist (simpdiss)
phylosortdiss
recluster.hist (phylosortdiss)
```

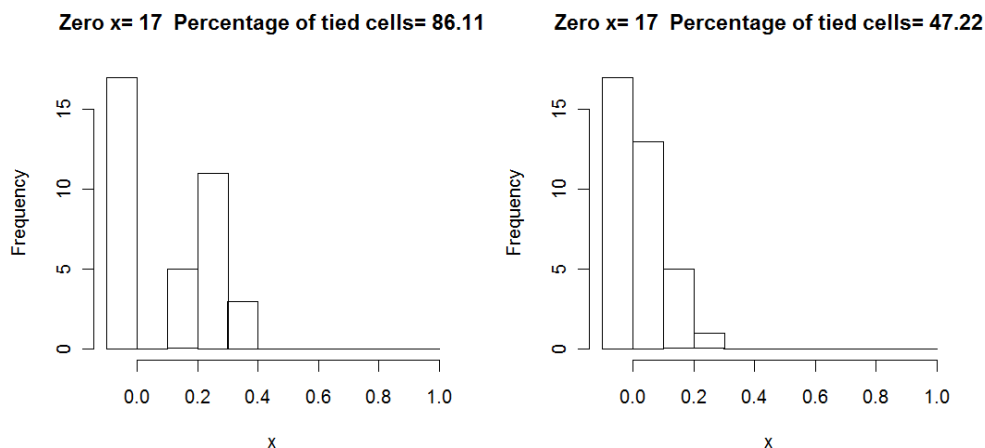


Fig. A1 The number of zero values (indicated by the first bar) is the same for faunistic (left) and phylogenetic beta-diversity (right) while numbers of ties are reduced in phylogenetic beta-diversity since genetic distances contribute to diversify the values. However, distribution of phylogenetic beta-diversity dissimilarities is skewed to lower values.

```
# Compute the tree with consensus rule of 100% for faunistic
beta-diversity
```

```
tree_f <- recluster.cons(datamod, p=1)$cons
plot (tree_f, direction="downwards")
```

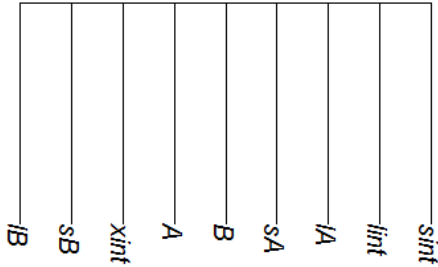


Fig. A2 No structure occurs at 100% consensus because of the high number of zero and tied values resulting in a high frequency of equivalent clustering solutions.

```
# The same result is obtained by phylogenetic beta-diversity
```

```
tree_p <- recluster.cons(datamod, treemod, dist="phylosort", p=1)$cons
```

```
# The tree, identical to the previous one, is not shown.
```

```
# Check row order bias on single nodes by recluster.node.strength
```

```
recluster.node.strength (datamod)
```

```
recluster.node.strength (datamod, treemod, dist="phylosort")
```

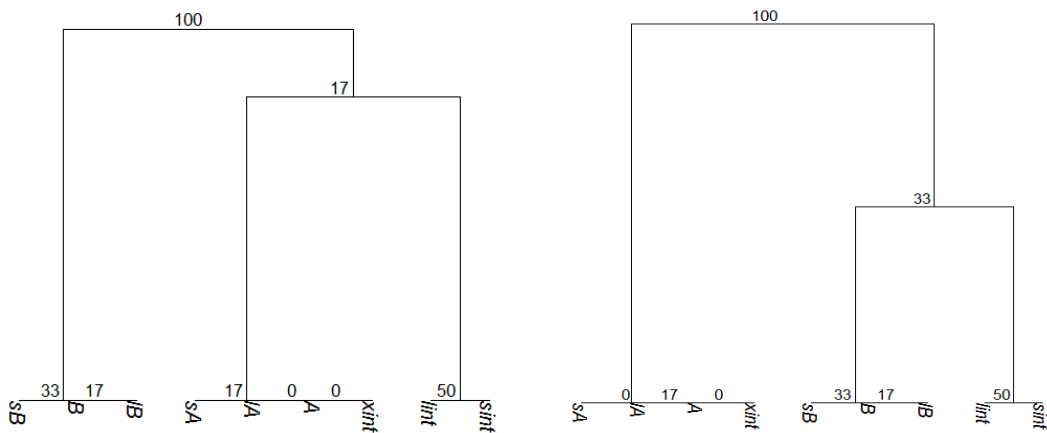


Fig. A3. Plots from `recluster.node.strength` (faunistic beta-diversity, left; phylogenetic beta-diversity, right). The tree obtained with the default UPGMA method using the original order of sites is compared with consensus trees produced after re-ordering of sites for six different consensus rules (from 0.5 to 1 with step 0.1). Values on nodes are the percentages of times each node is repeated among different consensus rules. Only the root is maintained for all consensus solutions and most nodes show very low percentages. This reveals that the row order has a strong influence since only a few nodes are repeated when increasing the consensus rule. Overall low strength indicates a high occurrence of equivalent trees, thus the necessity to use several trees for each consensus and the need to select a low consensus rule if, in an explorative hypothesis, a high number of clusters to be tested should be preferred. In such cases it is suggested that use is made of the lowest default consensus (0.5) with the default 100 re-ordered trees.

```
# Faunistic beta-diversity with 100 resampled trees and p=0.5
```

```
tree_f <- recluster.cons (datamod, p=0.5)$cons  
plot (tree_f, direction="downwards")
```

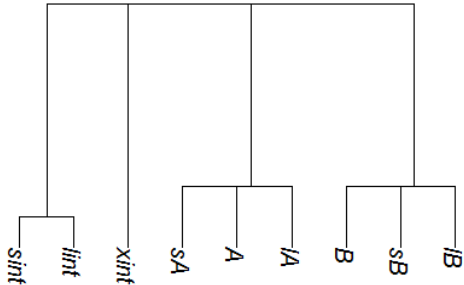


Fig. A4 The 50% consensus tree for faunistic data. With respect to Fig S2 a more realistic picture is obtained and intermediate islands are basal to both A and B groups.

```
# Phylogenetic beta-diversity
```

```
# The phylogenetic signals are weaker than the presence/absence ones  
making the trees more unstable. We thus suggest using 200 trees to  
compute the initial reference tree.
```

```
tree_p <- recluster.cons (datamod, treemod, dist="phylosort", tr=200,  
p=0.5)$cons  
plot (tree_p, direction="downwards")
```

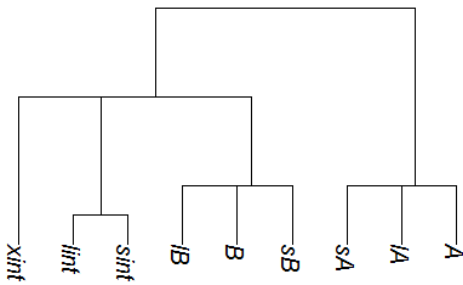


Fig. A5. The consensus tree for phylogenetic data. The topology greatly reflects the relationships for faunistic data except for intermediate islands, which are nearer to B because of the presence of the island endemic that is phylogenetically very similar to a species characteristic of the B fauna (see the main paper).

```
# A successive bootstrap analysis by constructing 1000 consensus trees  
made up of 100 site reordered trees each is suggested, but to obtain a  
faster even less accurate example we set here tr = 20 and boot = 100.
```

```
# Faunistic beta-diversity
```

```
boot_f <- recluster.boot (tree_f, datamod, tr=20, p=0.5, boot=100)  
recluster.plot (tree_f, boot_f)
```

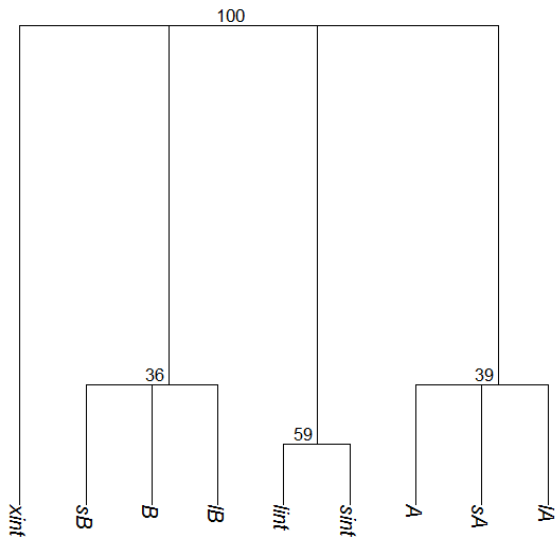


Fig. A6. The faunistic tree with bootstrap values. Supports obtained by using a number of re-sampled species equal to the actual number of species in the sample ( $\times 1$ ) is rather low. This can be a clue for the real weakness of the relationships or for the existence of a small subset of species responsible for turnover with respect to a large number of taxa involved in nestedness phenomena.

```
# Phylogenetic beta-diversity
boot_p <- recluster.boot (tree_p, datamod, treemod, dist="phylosort",
tr=20, p=0.5, boot=100)
recluster.plot (tree_p, boot_p)
```

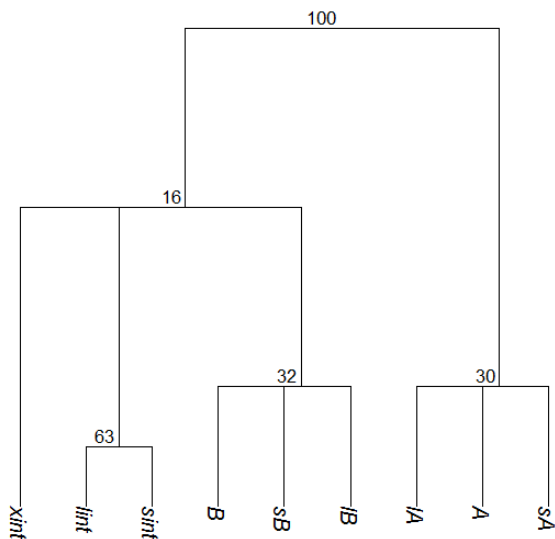


Fig. A7 The phylogenetic diversity tree with bootstrap values

# Perform a series of 10 bootstrap levels with a step of  $2\times$  in order to obtain a series of scales from  $\times 1$  to  $\times 19$ . This needs computation of many matrices and trees and requires extensive computation time, mainly for phylogenetic beta-diversity. To have a faster, even less precise, example we suggest to further reduce tr and boot.

```

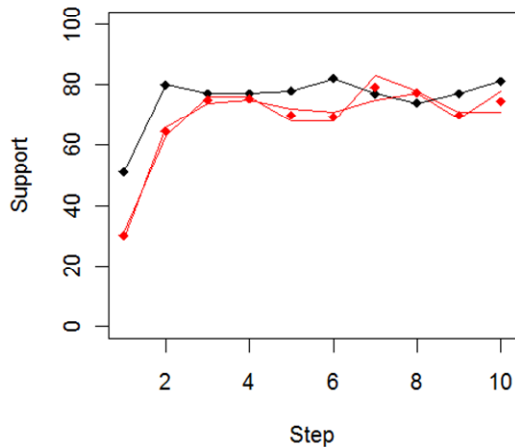
# Faunistic beta-diversity
multiboot_f <- recluster.multi(tree_f, datamod, tr=20, levels=10,
boot=100, step=2)

# Phylogenetic beta-diversity
multiboot_p <- recluster.multi(tree_p, datamod, treemod, dist="phylosort",
tr=20, levels=10, boot=100, step=2)

# Inspect and analyze trends for node support by recluster.identify.nodes.

# Faunistic beta-diversity
id_f <- recluster.identify.nodes(multiboot_f)
id_f

```



```

$nodes
[1] NA NA 2 NA NA 1 2 NA

$scale
[1] 1

#(mean values are omitted)

```

Fig. A8 Trend of support for nodes from step 1 (×1) to step 10 (×19). Identification of highly and poorly supported nodes (black and red, respectively) is done on the basis of Partitioning Around Medoids. Support for each node at different steps shows a substantial and immediate increase of bootstrap values for all nodes and no nodes with constantly weak support. Accordingly, \$scale identifies in the first step (×1) the best discriminating level for one highly and two poorly supported nodes. However, in this case there is no clear evidence for different behaviours of node support at increasing scales and we thus suggest that, in all cases, initial low support is equally due to the fact that only a small subset of species is responsible for turnover (sA, sB, sint and lint groups have only one species involved in turnover). Support tends to stabilize at step 3 (×5) and we suggest using ×1 and ×5 supports in plotting the cluster.

```

recluster.plot(tree_f, multiboot_f, 1, 3)

```

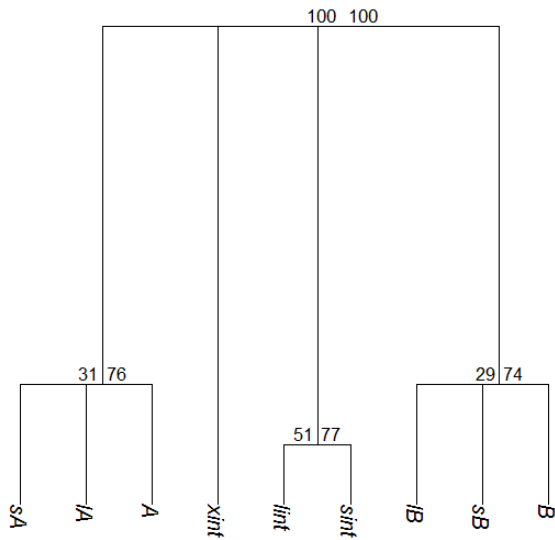
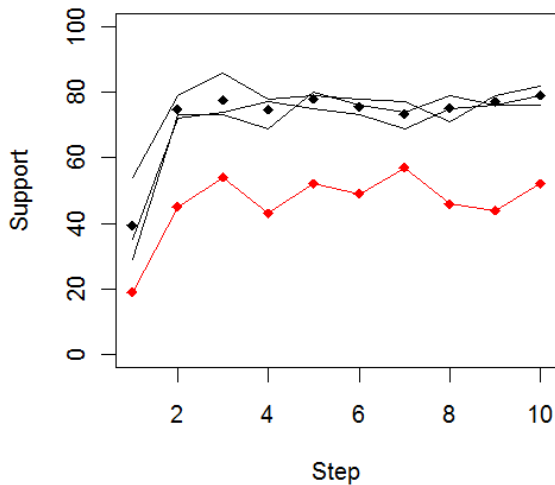


Fig. A9 The final representation for faunistic diversity. The three groups have rather weak support at x1 but stronger support at higher levels. Actually, the few species involved in turnover phenomena are unlikely to be sampled at x1 bootstrap level. Supports for higher levels of bootstrap are also consistent with the rules used to construct the dataset (i.e. formation of A, B and sint-lint groups).

```
id_p <- recluster.identify.nodes (multiboot_p)
id_p
```



```
$nodes
[1] NA 2 NA 1 1 NA 1 NA

$scale
[1] 2
```

Fig. A10 The cluster for phylogenetic beta-diversity is composed of four nodes instead of three whose support in multiscale bootstrap shows two different trends with three nodes rapidly reaching high values of support near to 80%, and one node stabilizing its value under 60%. Accordingly, the function identified differences among these two groups of nodes (red and black, respectively) with the best compromise of separation identified for the second scale ( $\times 3$ ), but any other successive step has a similar trend.

```
recluster.plot(tree_p, multiboot_p, 1, 2, id=id_p$nodes)
```

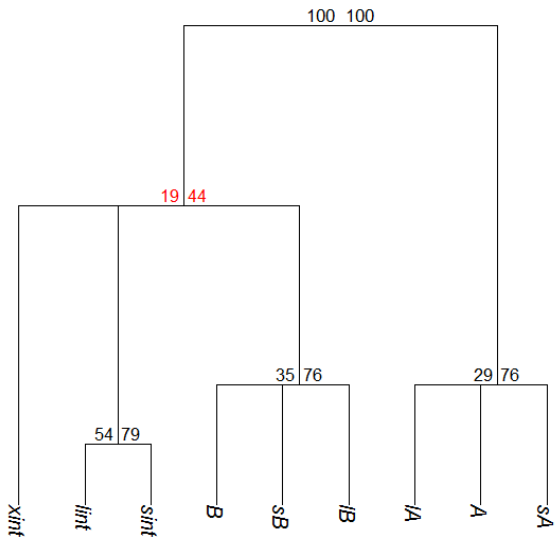


Fig. A11 The phylogenetic diversity tree plotted with ( $\times 1$  and  $\times 3$ ) support for nodes using the same red and black colours used to categorize nodes in the previous graph. The poorly supported node is the one linking the three "int" islands to B. While a link between sint and lint with B islands can be expected in phylogenetic beta-diversity due to the high similarity between the sint and lint endemic with a B species, this cluster also includes the xint island, which has equal dissimilarity to A and B and can be attributed to different groups in different bootstrap runs. This has likely reduced the support for this node.



The same analyses can be conducted on the real butterfly dataset. Together with presence data provided in the package data a phylogenetic tree (treebut) is also available.

```
data(dataisl)
data(treebut)
```

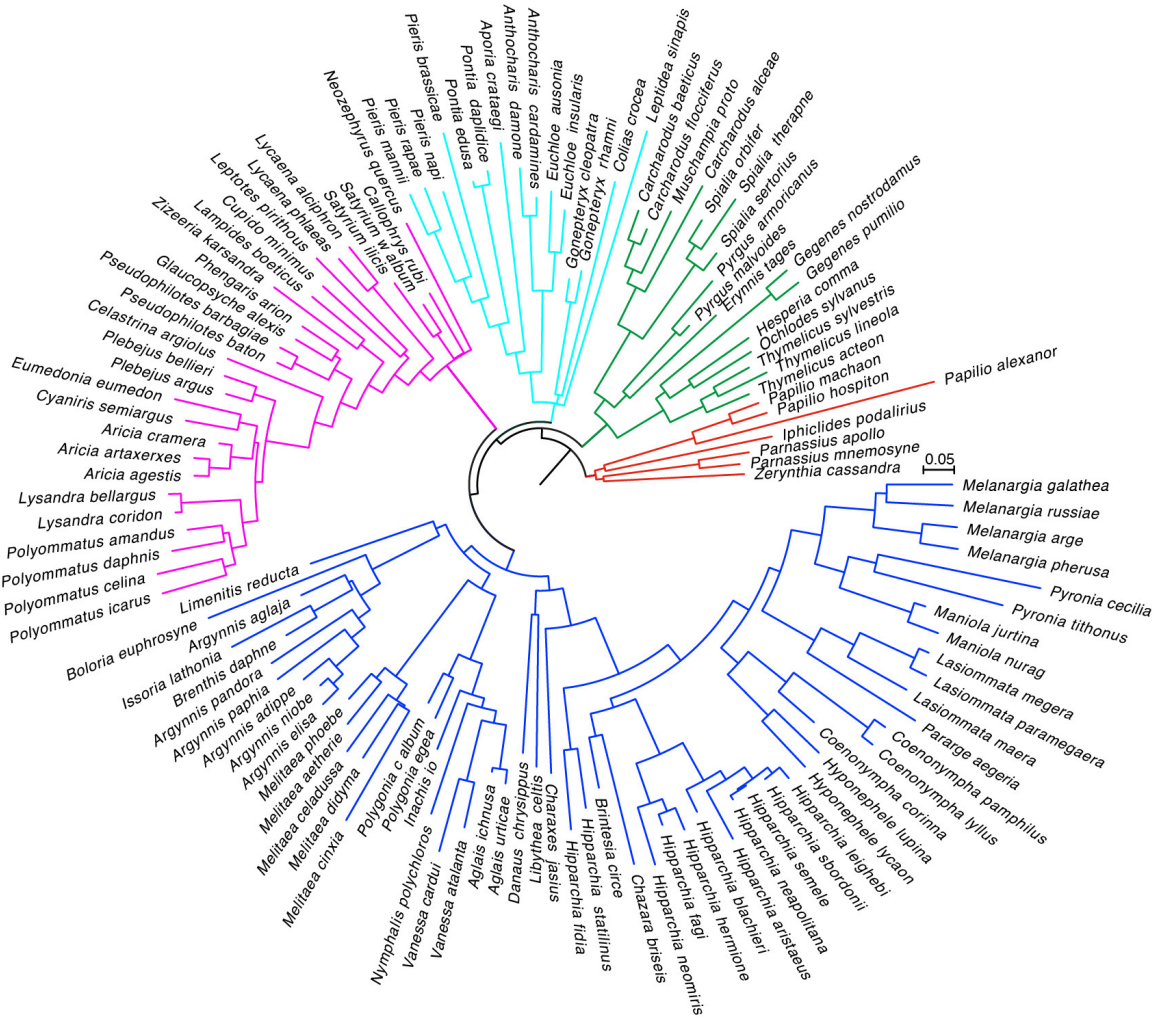


Fig. A12 Maximum Likelihood reconstruction based on COI barcoding sequences (640 bp) for all West-Mediterranean island species available in public databases (GenBank and BOLD). The tree was inferred with RAXML with a gamma model of rate heterogeneity and topological constraints at the levels of family and subfamily following the butterfly phylogeny published in Heikkilä et al. (2012). The different families are represented by different colours (Papilionidae, red; Hesperidae, green; Nymphalidae blue; Lycaenidae, magenta; Pieridae, cyan).

```
# Faunistic dissimilarity
simpdiss_b <- recluster.dist(dataisl)
simpdiss_b
recluster.hist(simpdiss_b)

# Phylogenetic dissimilarity
```

```
phylosortdiss_b <- recluster.dist(dataisl, treebut, dist="phylosort")
phylosortdiss_b
recluster.hist(phylosortdiss_b)
```

Fig. A13 Faunistic beta-diversity, left; phylogenetic beta-diversity, right. Despite showing the same number of zero values phylogenetic beta-diversity encompassed less tied values and, as for the model data, the dissimilarity distribution is highly skewed to low values.

```
# Node strength for faunistic data
recluster.node.strength(dataisl)
```

Fig. A14 Recluster.node.strength confirmed that the tree for faunistic data is highly dependent on row order.

```
# Node strength for phylogenetic data
# tr has been reduced to 50 to shorten the phylogenetic analysis
recluster.node.strength(dataisl, treebut, tr=50, dist="phylosort")
```

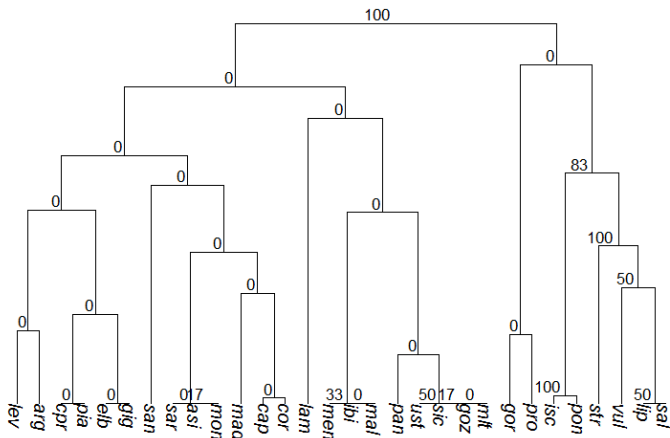


Fig. A15 A high dependency on row order occurs also for phylogenetic beta-diversity.

```
# Faunistic data
# We selected default tr=100 and p=0.5 to create the first consensus tree.

tree_bf <- recluster.cons (dataisl, p=0.5)$cons
plot(tree_bf, direction="downwards")
```

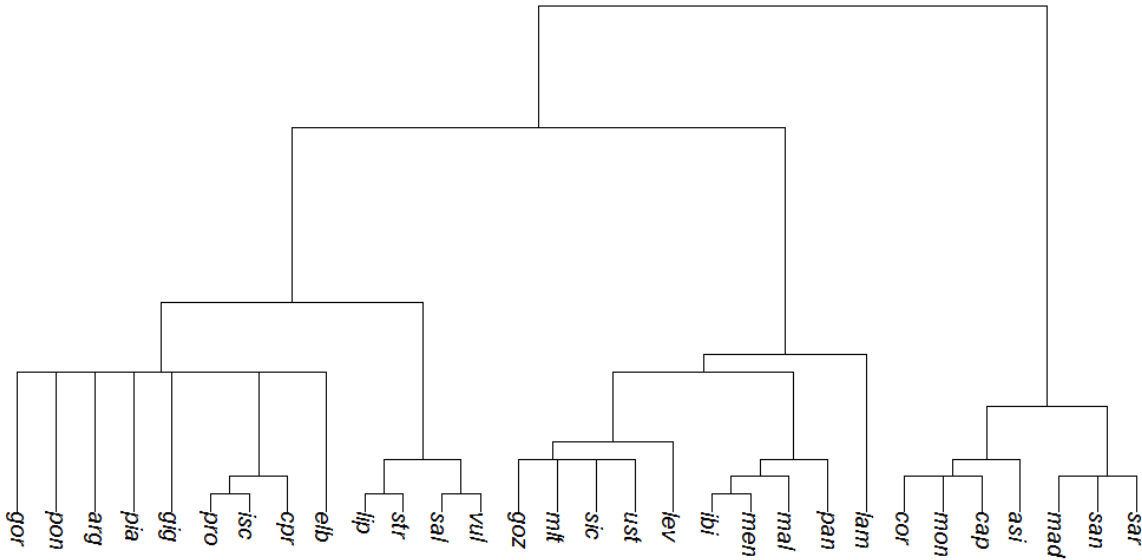


Fig. A16. The consensus tree for faunistic data.

```
tree_bp<- recluster.cons (dataisl, treebut, dist="phylosort", p=0.5)$cons
plot(tree_bp, direction="downwards")
```

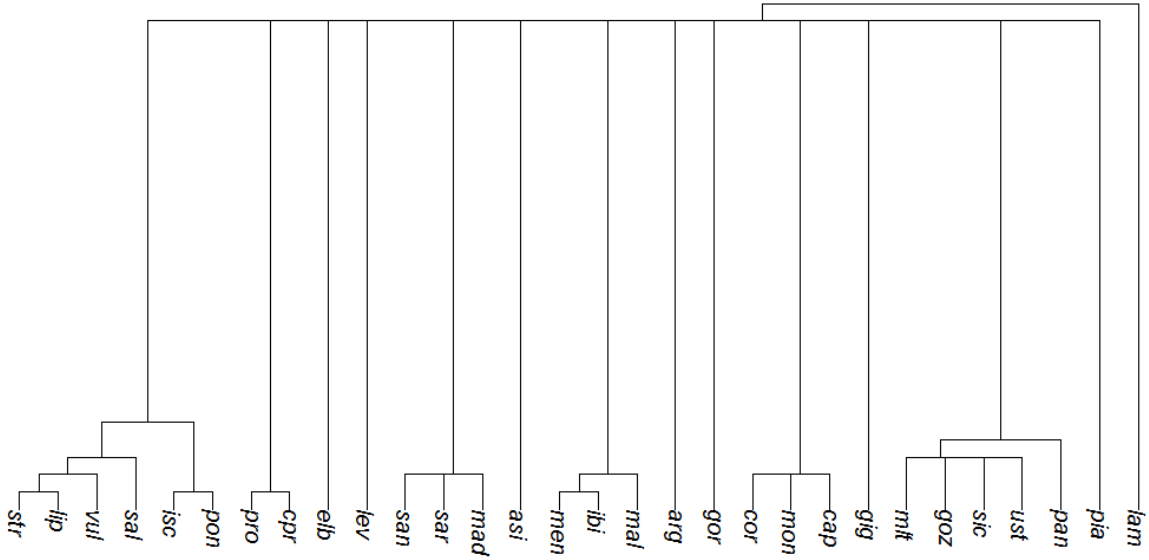
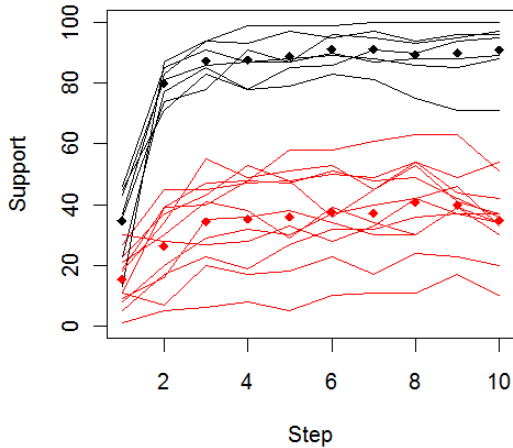


Fig. A17 The consensus tree for phylogenetic data.

```
# Multiscale bootstrap for faunistic beta-diversity
multiboot_bf <- recluster.multi (tree_bf, dataisl, tr=20, boot=100,
levels=10, step=2)
id_bf<-recluster.identify.nodes(multiboot_bf)
id_bf
```



```
$nodes
[1] NA 2 2 2 NA NA NA NA NA 2 2 1 2 2 2 2 1 NA NA 1 1 1
1 2 2 NA 1 NA

$scale
[1] 2
```

Fig. A18. Recluster.identify.nodes recognized two well identifiable groups of nodes with a best compromise at level 2 (x 3).

```
# Plot the tree with bootstrap supports
```

```
recluster.plot (tree_bf, multiboot_bf, 1, 2, id=id_bf$nodes)
```

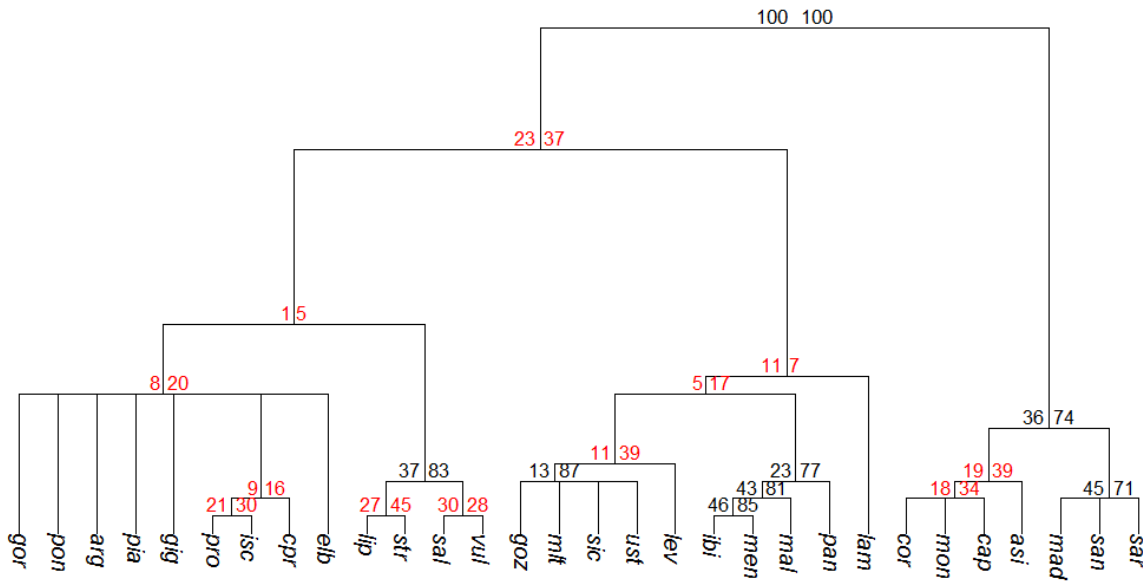
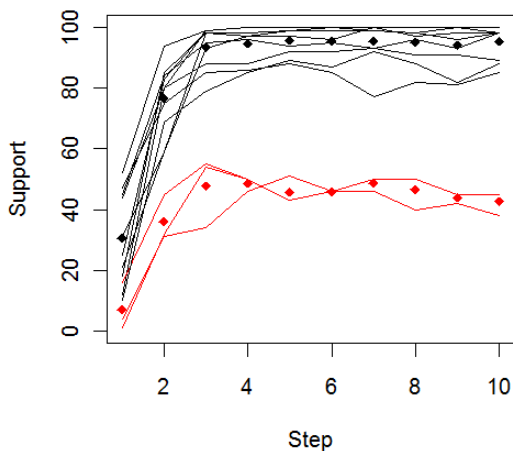


Fig. A19 Consensus tree with  $\times 1$  and  $\times 3$  support values. Terminal nodes (island groups) received better support with respect to higher (regional) patterns.

```
# Phylogenetic beta-diversity
```

```
multiboot_bp <- recluster.multi (tree_bp, dataisl, treebut,
dist="phylosort", tr=20, boot=100, levels=10, step=2)
id_bp <- recluster.identify.nodes(multiboot_bp)
id_bp
```



```
$nodes
 [1] NA 1 NA NA NA NA NA 1 1 2 2 1 1 NA 1 NA NA NA 1 1 NA NA 1
NA NA 2 1 NA NA

$scale
 [1] 3
```

Fig. A20. Recluster.identify.nodes suggests the existence of two groups of nodes with a greatest separation at scale 3 ( $\times 5$ ).

```
recluster.plot (tree_bp, multiboot_bp, 1, 3, id=id_bp$nodes)
```

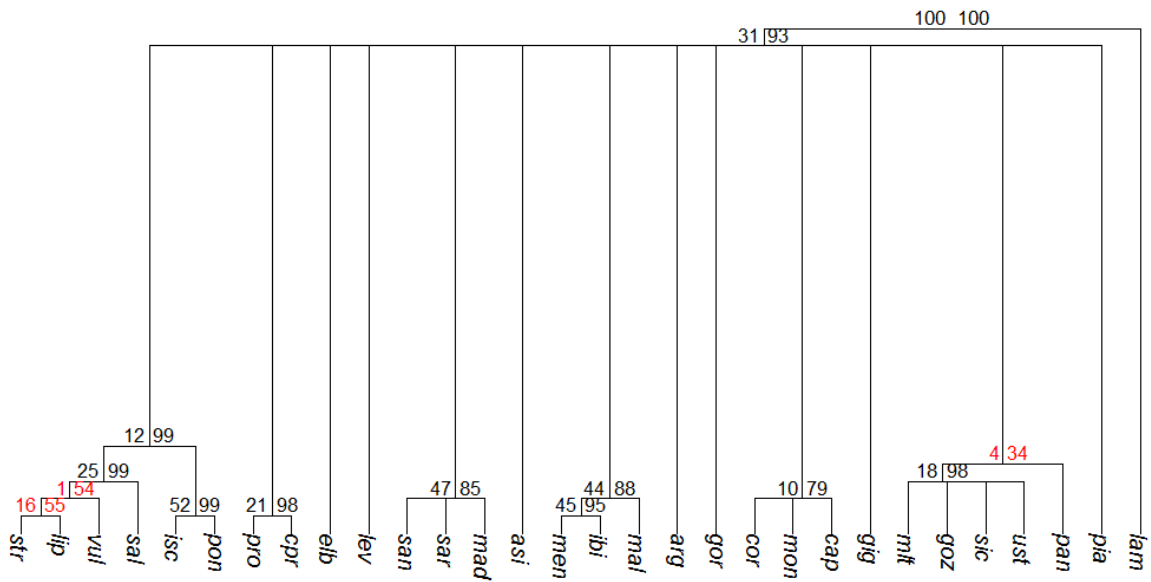


Fig. A21 A plot of the phylogenetic diversity tree with supports obtained in  $\times 1$  and  $\times 5$  levels.

#### References

Heikkila, M., et al. 2012. Cretaceous origin and repeated tertiary diversification of the redefined butterflies. - *P Roy Soc B-Biol Sci* 279: 1093-1099.