

# **Biorap Project Papua New Guinea pilot study**

## **Classification procedures and preparation of files for TARGET analysis**

J.L.Stein,  
CRES, ANU  
September, 1998

### **1.Introduction**

These notes describe the derivation of environmental domains, forest type/ domain combinations and species bioclimatic profile group distributions and their preparation for use as biodiversity surrogates for the TARGET program priority area analysis of D.Faith and C.Margules, CSIRO Division of Wildlife and Ecology. The locations to be considered are the Resource Mapping Units (RMU's) from the PNGRIS database. TARGET requires these units be described in terms of their composition of each of the biodiversity surrogate attributes.

### **2. RMU polygons**

PNGRIS files in MAPINFO format , were converted to ARC/INFO (ESRI,1996) for overlay with other data layers. The table allndat.tab with RMU polygons was converted to a coverage first by exporting the table from MAPINFO to Mapinfo Interchange Format (mif), then with the MIFSHAPE command converting to an ARCVIEW shape file. Examination of the dbf file showed some of the attributes associated with each RMU had not converted correctly. This error was corrected by editing the mif file and increasing the field length of some fields before running the MIFSHAPE command. Finally the ARC/INFO commands SHAPEARC, CLEAN and REGIONPOLY were used to convert the shape file to an ARC/INFO coverage and build polygon topology.

To compute areas the RMU coverage was projected from a geographic coordinate system with decimal degree units to UTM projection (Zone 55) with units in metres with the ARC/INFO PROJECT command.

### **3. Environmental domains**

Environmental domains were produced by classifying grid cells according to their similarity across 50 parameters describing key climate, terrain and substrate conditions (table 1), using a non-hierarchical clustering procedure in the PATN software package (Belbin,1987).

#### **3.1 Bioclimatic parameters**

The bioclimatic parameters are biologically meaningful combinations of monthly mean climate values (Hutchinson et.al. 1996). Regular grids of these parameters were calculated by the BIOCLIM program in the ANUCLIM package (Hutchinson et.al. 1997) by combining climate surfaces that are functions of latitude, longitude and elevation, with a 0.01 degree DEM.

This DEM was developed by interpolating topographic data consisting of point and contour elevation data and streamline vectors using the ANUDEM program (Hutchinson,1989). Topographic data was mainly sourced from the Digital Chart of the World 1:1,000,000 scale database with additional spot heights digitised from TPC 1:500,000 mapsheets (Ellis, 1997). Climate surfaces for monthly mean maximum and minimum temperature and precipitation were produced by J.L.Kesteven using programs in the ANUSPLIN package (Hutchinson, 1997) to spatially interpolate monthly mean climate data from the Bureau of Meteorology in Papua New Guinea and Australia with additional rainfall data taken from McAlpine, et.al. (1975) and McAlpine et.al. (1983). The radiation surface used was developed by J.P. Mc.Mahon prior to the commencement of the project.

**Table 1. Attributes describing each grid cell**

Number	Description and units
<i>Attribute Group 1 (Climate)</i>	
1	Annual mean temperature (°C)
2	Temperature seasonality (CofV %)
3	Mean Temperature hottest month (°C)
4	Mean Temperature coldest month (°C)
5	Mean Temperature wettest quarter (°C)
6	Mean Temperature driest quarter (°C)
7	Mean Temperature hottest quarter (°C)
8	Mean Temperature coldest quarter (°C)
9	Annual mean precipitation (square root, mm)
10	Precipitation seasonality (CofV %)
11	Wettest quarter mean precipitation (square root, mm)
12	Driest quarter mean precipitation (square root, mm)
13	Annual mean radiation (MJ/m <sup>2</sup> )
14	Radiation seasonality (CofV %)
15	Wettest quarter mean radiation (MJ/m <sup>2</sup> )
16	Driest quarter mean radiation (MJ/m <sup>2</sup> )
<i>Group 2 (Terrain + landform)</i>	
17	Slope (degree)
18	Curvature (m)
19	Relief (elevation range 4 nearest neighbours) (m)
20	Relief 5k (elevation range 5 km radius) (m)
21	Elevation percentile (elevation rank in 5km radius) (m)
22	Elevation difference (difference from mean elevation in 5km radius) (m)
46	Area of mangroves (in grid cell)
47	Area of beach ridges (in grid cell)
48	Area of swamps (in grid cell)
49	Area of volcanic domes (in grid cell)
<i>Attribute Group 3 (Lithology)</i>	
23	Fine grained sedimentary
24	Coarse grained sedimentary
25	Mixed or undifferentiated sedimentary
26	Mixed sedimentary or limestone
27	Limestone
28	High grade metamorphics
29	Low grade metamorphics
30	Mixed or undifferentiated metamorphics
31	Mixed sedimentary and volcanic
32	Basic to intermediate volcanic
33	Intermediate to acid volcanic
34	Mixed or undifferentiated volcanic
35	Mixed volcanics and limestone
36	Acid to intermediate volcanics
37	Basic igneous
38	Ultrabasic
39	Pleistocene sediments
40	Alluvial deposits
41	Marine sand
42	Estuarine deposit
43	Volcano-alluvial deposits
44	Pyroclastics
45	Overlain with recent ash

### 3.2 Terrain parameters

Regular grids of terrain parameters were computed directly from the DEM. Three of the parameters, elevation percentile, elevation difference and relief were computed by the ElevResidGrid program of John Gallant (see <http://cres.anu.edu.au/~johng/elevresid.html>) using a radius of 5 km to define the analysis neighbourhood around each cell.

Other terrain parameters were computed using functions in ARC/INFO GRID. Slope was computed using the SLOPE function with a scale factor of 0.00001 to scale the elevation units in metres to the x,y units in degrees. A second relief parameter was computed using the FOCALRANGE function with a neighbourhood defined to include the cell and its 4 nearest neighbours. Curvature was computed with an ARC/INFO AML program adapted from a FORTRAN routine provided by Mike Hutchinson.

### 3.3 Landform

Seven classes of landform from the 40 classes in the PNGRIS table, allndat.tab were selected as landform features not readily identified from the terrain analysis (table 2). Combining the three classes of swamp reduced these to five classes. The area of each of these five landform classes in each 0.01 degree grid cell was computed by overlaying the polygons with selected landform class on a polygon coverage of grid cells with the ARC/INFO INTERSECT command. The coverage of grid cells was derived with the ARC/INFO GRIDPOLY function from a grid of uniquely numbered cells. Coverages were projected to UTM Zone 55 coordinates for area calculations.

**Table 2. Landform classes selected from PNGRIS**

<b>PNGRIS Landform Class</b>	<b>New Landfor m class</b>	<b>Description</b>
1	1	Mangrove swamps
3	2	Beach ridge complexes and beach plains
22	3	Back swamps
23	3	Blocked or drowned valley swamps or lakes and their associated floodplain
24	3	Undifferentiated swamps
56	4	Volcanic cones and domes
60	5	Lake

### 3.4 Lithology

A digital version of the 1:250,000 Geology of Papua New Guinea was prepared by AGSO and provided in the form of an ARC/INFO polygon coverage. Over 900 lithology types were represented. These were grouped using the lithology classification adopted for PNGRIS by H.A.Nix, assigning a PNGRIS lithology code to each lithology symbol. Removing boundaries between adjacent polygons with the same PNGRIS lithology code value using the ARC/INFO command DISSOLVE reduced the number of polygons. This revised coverage was overlaid on the grid cell polygon coverage to determine the area of a lithology class in each grid cell in a similar procedure to that described for the landform

The 1:250,000 coverage was compared with the lithology boundaries represented in the PNGRIS database. Locational shifts were evident for some features, however, these distortions were not consistent making it difficult to rectify one coverage to the other.

Rectification of the lithology is best done by reference to accurately located topographic features, a task well beyond the scope of this project. As a result there may be occasional anomalies in the combination of landform feature and lithology type for any particular grid cell.

### 3.5 Classification procedure

A FORTRAN program was written to combine the grid cell values of bioclimatic and terrain parameters with the area in each of the landform and lithology classes to produce a datafile in PATN format. Only those grid cells with values for all attributes were considered. This excluded areas west of the PNG border where no landform or lithology information was available, some small islands and occasional grid cells along the coast where one or more data layers were missing due to slight differences between map bases in the delineation of the coastline. The resulting data matrix consisted of 387109 rows (grid cells) and 50 columns (attributes).

The non-hierarchical clustering algorithm ALOB (Belbin,1995) was used to generate a classification of the grid cells. This technique, based on the allocation of objects to a set of seeds, is particularly suited to the analysis of very large data sets. The number of seeds and hence the number of groups generated by ALOB is controlled by the value of the supplied allocation radius. In this case the target of about 600 groups was achieved with a selected allocation radius of 0.053 producing 608 groups.

The Gower metric with built-in range standardization was chosen as the dissimilarity measure, computed over three attribute groups (table 1) such that each attribute group contributed equally to the measure of dissimilarity. ALOB is an iterative procedure, sequentially extracting each object from the group it has been allocated to and allocating it to the group with the nearest centroid. With over 600 groups each iteration took over 1.5 hours computing time on a SUN Ultra computer. After 79 iterations the measure of group heterogeneity, the average deviation, was changing insignificantly so the group definition at this stage was accepted for subsequent analysis.

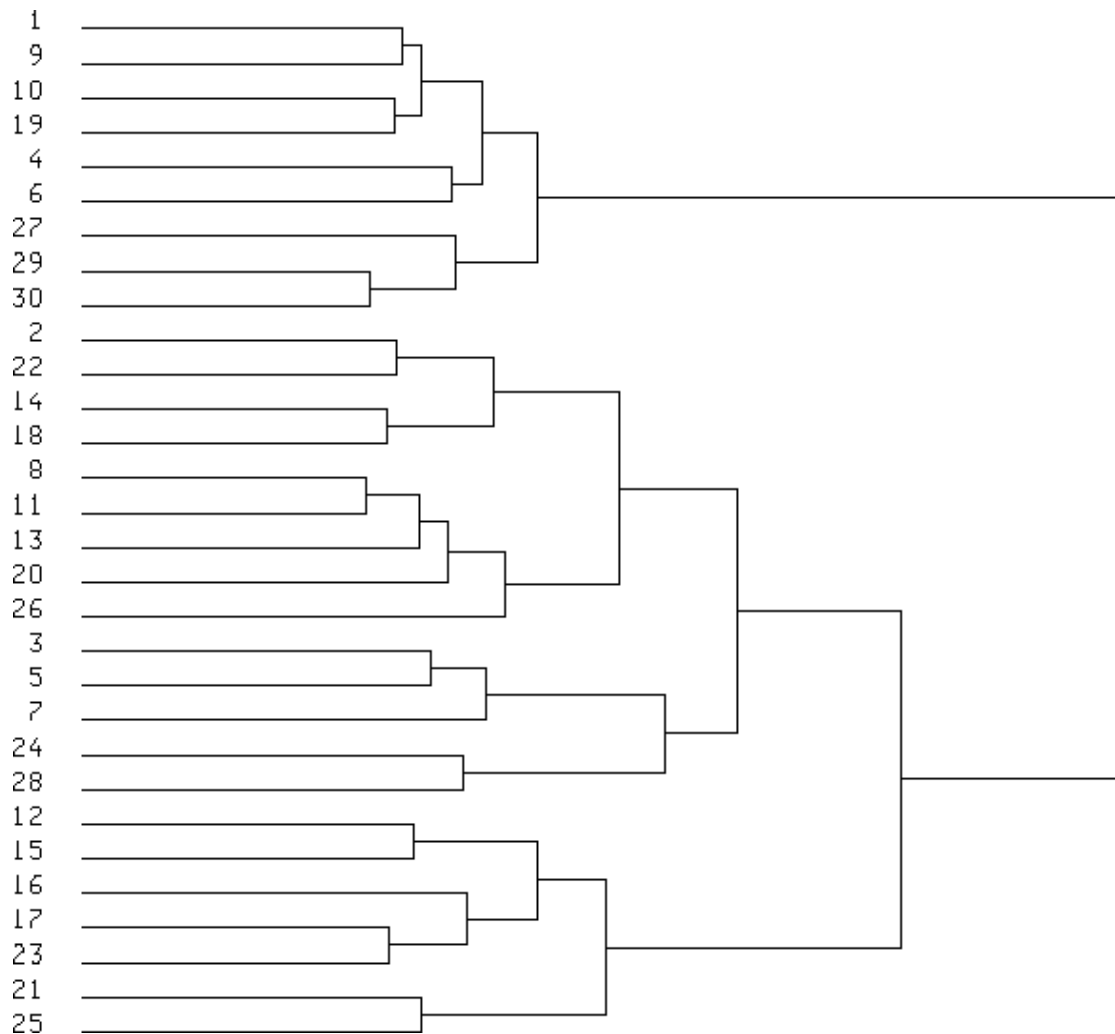
On completion the ALOB program normally computes an inter-group dissimilarity matrix which can be used to examine the higher order group structure. In this case, where ALOB had not completed, this matrix had to be computed independently from the group means, computed with the PATN program, GSTA. A standard hierarchical cluster analysis of the groups was then undertaken in PATN. Firstly, a dissimilarity matrix was computed with the GASO program using the same attribute groupings (table 1) and dissimilarity measure (Gower metric). This matrix was the input to the FUSE program, which provides a number of options for agglomerative hierarchical classification. In this case, the recommended, default options of flexible UPGMA with a beta value of -0.1 were adopted. Finally, a dendrogram (fig.1) was produced from the fusion history. New group definitions for 10 and 30 groups respectively, were also derived from this fusion history using the GDEF program.

In order to display the environmental domains the group definition file, in the form of a group affinity vector, was converted to a grid in ARC/INFO ASCIIGRID format using a FORTRAN program which first merged the grid cell centroid x,y location with group number. A colour file was produced by assigning each group an RGB colour definition based on its position in the 3 dimensional space of an ordination generated by the PATN program, SSH, using all default options. This technique, developed by Belbin, Marshall and Faith (1983), allows groups to be displayed in colours that reflect inter-group dissimilarities.

### 3.6 Overlay with RMUs

To compute the area of each domain in an RMU the grid of domains was converted to a polygon coverage with the ARC/INFO GRID function GRIDPOLY, projected to UTM Zone 55 co-ordinates and overlaid with the RMU coverage with the ARC command INTERSECT. The ARC FREQUENCY command was used to produce a summary table of RMU, domain and area. This table was converted to the .loc format required by TARGET with a FORTRAN program, ARC2TARG. This program used a lookup table, rmuseqno.tab, to assign a sequence number to each RMU consistent with those in the map base polygon file, rmuseqno.pol. Another summary table was produced with the total area of each domain.

**Figure 1. Dendrogram showing group structure at the 30 group level**



#### **4. Forest types**

RMUs were also described by their component forest types. These were derived from the FIMS dataset, converted from MAPINFO format and projected to UTM Zone 55 co-ordinates by Andre Zerger at CRES. Each FMU polygon in the dataset was associated with one of 642 forest types. C.R.Margules, CSIRO Division of Wildlife and Ecology, provided a lookup table to reduce these to 209 classes by combining similar types. Non-forest or disturbed types were excluded from further analysis. This coverage was overlaid with the polygon coverage of environmental domains at the 10-group level, created from the domain grid with the GRIDPOLY function and projected to UTM co-ordinates. The area of each of the resulting 634 unique combinations of forest type and 10 group domain were then summarised by RMU. Combinations where either forest type or domain had a value of zero, for example, where the underlying mapbase coastlines did not match, were excluded. Finally, a TARGET .loc file for the combination attributes was produced with the ARC2TARG program.

#### **5. Bioclimatic profile group distributions**

The third set of attributes compiled for TARGET analysis are the bioclimatic distributions of 10 groups of species classified according to the similarity of their bioclimatic profiles. The data available for this analysis was collated from herbarium specimen records provided by CPBR, Canberra, Emoia specimen locations prepared by Bishop Museum, Hawaii and

records for 3 species of Birds of Paradise provided by DEC, PNG. Most of this data was imported into the Biorap Database.

From the total of more than 10,000 records, specimen locations for 87 species (table 3) were extracted. These species selected were those from the list of preferred genera with reasonable numbers of records with full geocoding (ie.complete latitude, longitude and elevation) or species of particular interest, such as the orchids. Species, such as the Bird of Paradise, *Astrapia stephaniae*, with latitude and longitude recorded only to the nearest degree, were not used. The list of species also included 26 species analysed as part of the Biorap training program in October 1997 by Gerard Natera, David Mowbray, Jane Mogina and Anthea Dee. Data for other species were prepared for bioclimatic analysis by Gerard Natera, David Mowbray and Ursula Grott.

Preparation included extracting the records from the Biorap database with the "Export to Bioclim" option and identifying geocoding errors. Mapping specimen locations with either IDRISI or ARC/INFO identified obvious location errors, for example, sites in the ocean. Checking outliers on the bioclimatic profiles generated by the BIOCLIM program (Hutchinson et.al., 1997) identified other errors. Wherever possible errors were corrected from 1:100,000 scale topographic maps. However, many anomalous records were simply removed, for example, where either the locality description for the specimen was not sufficiently detailed to allow the collection site to be accurately located or the described locality was not found on the maps. A list of erroneous sites or those requiring further queries has been compiled for CPBR and Bishop Museum. Changes were only made to the text files; no corrections have yet been made to records in the Biorap database

A bioclim profile was thus produced for each species. These were combined into one data matrix where rows are the species and columns are the 11 percentile values for each of the 16 bioclim parameters from table 1 (a total of 176 attributes). Species were then classified using a standard PATN hierarchical agglomerative clustering strategy with the Gower dissimilarity measure and the flexible UPGMA fusion option with the default beta value of -0.1. Attributes were not grouped for the computation of dissimilarity thus the 8 temperature attributes had a greater weight than either the 4 precipitation or 4 radiation attributes. Examination of the dendrogram (fig.2) suggested about 10 groups as an appropriate level to break the hierarchy. Group statistics generated for each of the groups confirmed 10 groups as a reasonable division of the species. At this level there was still biologically meaningful discrimination between the two closest groups.

To generate a predicted bioclimatic distribution for each species group a bioclim profile was produced from the combined specimen records of the members of the group. The predicted distribution of the group was determined by matching the values of the bioclimatic parameters estimated for each grid point on the 0.01 degree DEM to the bioclimatic profile values with the BIOMAP program. A species group was predicted to occur at a grid point if the values of the 16 bioclimatic parameters in table 1 were all within the range of the corresponding parameter in the profile. Unlike the environmental domains, more than 1 species group may be predicted to occur at a grid point.

The grids of predicted distributions were then converted to a polygon coverage with the GRIDPOLY command, projected to UTM Zone 55 co-ordinates and overlaid on the RMU coverage with the INTERSECT command. From the resulting combined coverage a summary table was generated with the FREQUENCY command giving the area of one or more of the 10 species groups predicted to occur within each RMU. This table was converted to the TARGET .loc format with the ARC2TARG program.

Table 3. Species selected for bioclimatic analysis.

Species	Abbreviation	Number of records (PNG)
<b>Plants</b>		
Amyema artensis	amyemart	40
Castanopsis acuminatissima	castaacu	111
Casuarina oligodon	casuaoli	15
Dimorphanthera amplifolia	dimoramp	76
Dimorphanthera fissiflora	dimorfis	17
Dimorphanthera keysseri	dimorkey	62
Ficus adenosperma	ficusade	99
Ficus ampelas	ficusamp	21
Ficus arfakensis	ficusarf	38
Ficus armiti	ficusarm	23
Ficus baeuerlenii	ficusbae	19
Ficus benjamina	ficusben	20
Ficus bernaysii	ficusber	17
Ficus botryocarpa	ficusbot	38
Ficus calodictya	ficuscal	7
Ficus casearioides	ficuscas	39
Ficus complexa	ficuscom	20
Ficus congesta	ficuscon	16
Ficus copiosa	ficuscop	39
Ficus dammaropsis	ficusdam	18
Ficus drupacea	ficusdru	25
Ficus erythrosperma	ficusery	79
Ficus glaberrima	ficusgla	22
Ficus glandulifera	ficusgan	19
Ficus gul	ficusgul	49
Ficus itoana	ficusito	22
Ficus macrorhyncha	ficusmac	20
Ficus melinocarpa	ficusmel	24
Ficus microcarpa	ficumicr	13
Ficus microdictya	ficusmic	19
Ficus mollior	ficusmol	52
Ficus nodosa	ficusnod	11
Ficus odoardi	ficusodo	14
Ficus pantoniana	ficuspan	27
Ficus pungens	ficuspun	64
Ficus septica	ficussep	60
Ficus subcuneata	ficussbc	21
Ficus subtrinervia	ficussbt	21
Ficus subulata	ficussub	154
Ficus tinctoria	ficustin	61
Ficus trachypison	ficustra	124
Ficus trichocerasa	ficustri	107
Ficus variegata	ficusvar	48
Ficus virens	ficusvir	58
Gymnostoma papuanum	gymnopap	41
Lithocarpus celebicus	lithocel	46
Lithocarpus lauterbachii	litholau	16
Lithocarpus megacarpus	lithomeg	7
Lithocarpus rufovillosus	lithoruf	70
Mediocalcar agathodaemonis	medioaga	17
Mediocalcar bifolium	mediobif	21
Melicope aequata	melicaeq	50
Melicope bonwickii	melicbon	60
Melicope conjugata	meliccon	48
Melicope durifolia	melicdur	88
Melicope mucronata	melicmuc	106
Melicope pachypoda	melicpac	64
Melicope robbinsii	melicrob	54

Species	Abbreviation	Number of records (PNG)
Melicope trachycarpa	melictra	26
Melicope vitiflora	melicvit	34
Melicope xanthoxyloides	melicxan	145
Nothofagus grandis	nothogra	35
Nothofagus perryi	nothoper	43
Nothofagus pullei	nothopul	54
Thelymitra papuana	thelypap	18
Vaccinium amblyandrum	vacciamb	98
Vaccinium finisterrae	vaccifin	15
Vaccinium shoddei	vaccisho	15
<b>Scinidae</b>		
Emoia atrocostata ssp.atrocostata	emoiaatr	21
Emoia bismarckensis	emoiabis	8
Emoia caeruleocauda	emoiacae	87
Emoia cyanogaster	emoiacya	10
Emoia jakati	emoiajak	101
Emoia kordoana	emoiakor	37
Emoia longicauda	emoialon	50
Emoia loveridgei	emoialov	26
Emoia nigra	emoianig	8
Emoia obscura	emoiaobs	31
Emoia pallidiceps ssp. mehelyi	emoiapam	23
Emoia pallidiceps ssp.pallidiceps	emoiapap	66
Emoia physcina	emoiaphy	15
Emoia physicae ssp.physicae	emoiaphp	39
Emoia physicae ssp.purari	emoiaphu	10
Emoia popei	emoiapop	19
Emoia veracunda	emoiaver	22
<b>Birds of Paradise</b>		
Paradisaea raggiana	paradrag	34
Pteridophora alberti	pterialb	11

## 6. Combining attributes

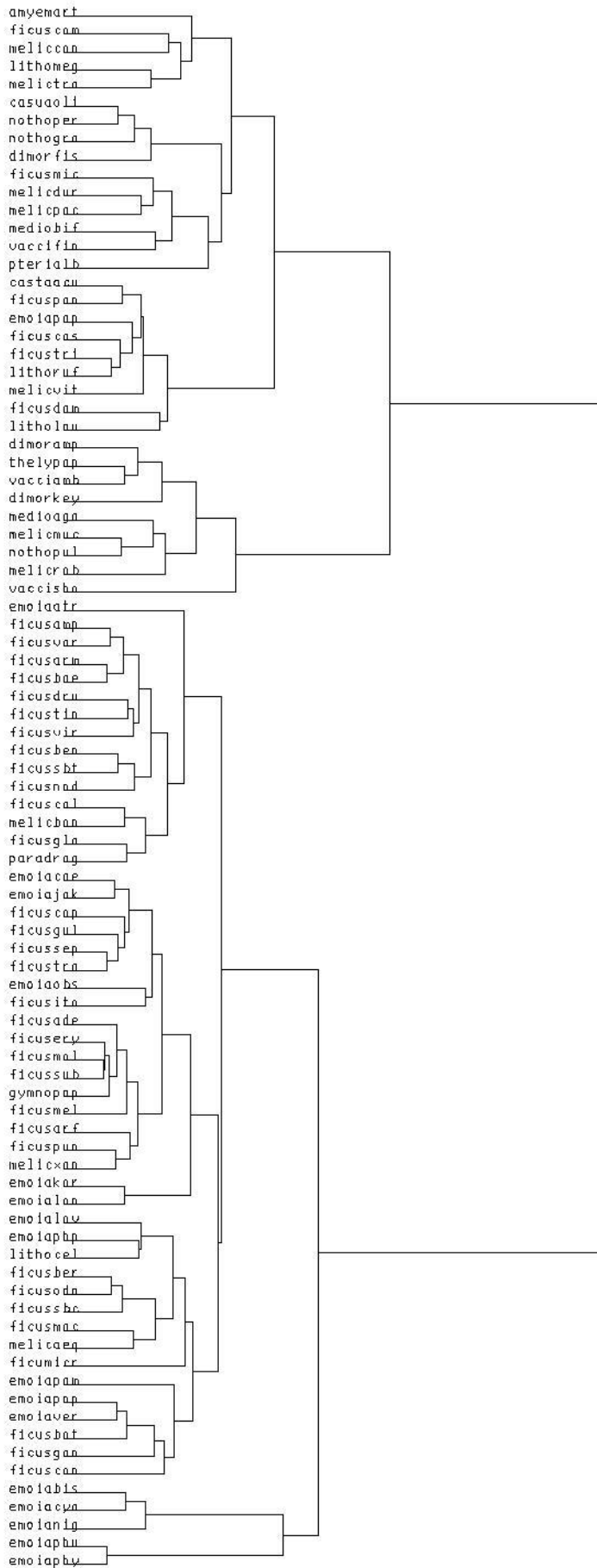
The final step in preparing the files for TARGET analysis involved combining the .loc files created independently for the three sets of attributes. This was done by incrementing the attribute numbers for the second and third sets of attributes by the maximum attribute value of the first and second sets respectively (table 4).

**Table 4. Attribute renumbering for combined .loc file, rmuall3.loc**

Original attribute range	Combined attribute range	Description
1 - 608	1 - 608	Environmental domains
1 - 634	609 - 1242	Forest type/ domain combination
1 - 10	1243 - 1342	Species group predicted distributions



Figure 2. Dendrogram from the classification of species based on their bioclim profile



## **7.References**

- Belbin,L. (1987) *PATN Manuals* CSIRO Wildlife and Ecology, Lyneham, ACT
- Belbin,L., Marshall,C., and D.P.Faith (1983) *Representing relationships by automatic assignment of colour*. Australian Computer Journal 15:160-163
- Ellis, F. (1997) *Making the New Guinea Digital Elevation Model (DEM)*, Centre for Resource and Environmental Studies, Australian National University, Canberra
- ESRI (1996) *ARC/INFO Version 7.0.4*, Environmental Systems Research Institute, Inc, Redlands, California, USA
- Hutchinson, M.F. (1989) *A new method for gridding elevation data with automatic removal of pits*. J.Hydrol. 106:211-232
- Hutchinson, M.F. (1997) *ANUSPLIN Version 3.2 User Guide*, Centre for Resource and Environmental Studies, Australian National University, Canberra
- Hutchinson, M.F., Belbin,L., Nicholls, A.O., Nix, H.A., McMahon, J.P. and K.D.Ord (1996) *BioRap. Volume Two. Spatial Modelling Tools*, The Australian BioRap Consortium, Canberra
- Hutchinson, M.F., Nix, H.A., Houlder, D.J. and J.P.McMahon (1997) *ANUCLIM User Guide*, Centre for Resource and Environmental Studies, Australian National University, Canberra
- McAlpine, J.R., Keig, G. and K.Short (1975) *Climatic Tables for Papua New Guinea*, Division of Land Use Research Technical Paper No. 37, CSIRO, Australia
- McAlpine, J.R., Keig, G. and R.Falls (1983) *Climate of Papua New Guinea*, CSIRO, ANU Press, Canberra