Measuring Forest Disturbance with DNA Barcodes: Molecular Taxonomy of Belizean Rainforest Dung Beetles (Coleoptera: Scarabaeinae)

by

John James Wilson

A thesis submitted in partial fulfilment of the requirements for the degree of Master of Science of the University of London and the Diploma of Imperial College

MSc in Advanced Methods in Taxonomy and Biodiversity

September 2005

Abstract

One of the greatest potential uses of DNA barcoding, sequence databases and economical high-throughput sequencing is the application in biodiversity surveys. Barcoding could provide a new level of efficiency and comparability to ecological surveys involving diverse and difficult-to-identify taxa. This study investigates whether phylogenetic analysis of short mtDNA sequences (16S rRNA) from Belizean rainforest dung beetles produces groupings that can be recognised across survey locations and between surveying time. I examine how these genospecies relate to traditional taxonomic species and the nature of the clustering amongst groups. Analysis of 16S sequences for over 350 individuals from three study sites, one sampled on two occasions, results in a phylogenetic tree that includes a number of distinctive easily recognizable, tip clades (MOTU). These clusters with a few notable exceptions show non-overlapping patterns of sequence divergence; a pattern of low distances within groups and high distances between groups seen in other barcoding studies. Individuals identified to traditional species with morphology (type sequences) fell into distinct clusters making it possible to put Linnaean binomials to 50% of the groups. Mitochondrial DNA clustering was congruent with genotypes produced by analysis of nuclear sequences (28S rRNA) adding supporting evidence to the species hypotheses. 60% of MOTU were collected on multiple sampling expeditions showing the transferability of the DNA barcodes between sites and over time. The present results indicate that an identification system for the Scarabaeinae based on mtDNA fragments will be highly effective and an inventory of MOTU is consistent with higher dung beetle diversity in less disturbed tropical forests. In order for DNA taxonomy to avoid the pitfalls of traditional identification methods, sequences obtained need to be available for comparisons with barcodes collected on future surveys. A web-based DNA barcode database containing the sequence chromatograms and taxonomic information is sorely needed to facilitate this approach.

TABLE OF CONTENTS

Abstract **2**

Introduction **4-5**
1.1. Dung beetles as indicators of forest disturbance 4
1.2. DNA taxonomy 5
1.3. Barcoding biodiversity 5

Materials and methods **6-11**
2.1. Study sites 6
2.2. Field Sampling 7
2.3. DNA extraction and sequencing 8
2.4. Phylogenetic analyses 9
2.5. 16S sequence variation 10

Results **11-17**

Discussion **20-25**
4.1. What makes a good barcode? 20
4.2. What makes a good MOTU? 22
4.3. Dung beetle diversity 23
4.4. Barcoding biodiversity studies: already a reality? 24
4.5. Summary 25

Acknowledgements **26**

References **27-29**

Appendices **30-35**
Appendix A: Taxonomy and historical biogeography of the dung beetles of Belize 30
Appendix B: Laboratory protocols 32
B.1. DNA extraction 32
B.2. PCR amplification 33
B.3. PCR cleanup 34
B.4. Sequencing reaction 34
B.5. Sequencing cleanup 35

1. Introduction

1.1. Dung beetles as indicators of forest disturbance

Perturbations of the forest environment change dung beetle community composition and often decrease diversity (Vulinec 2000). Consequently, dung beetles have been utilized as bioindicators of tropical forest disturbance (e.g. Nummelin & Kaitala 2004; Scheffer 2004; Halffter and Favila 1993). However, their use in this role is limited by difficulties in species identification and consequently lack of accurate inventories. Despite receiving wide attention from systematists (e.g. Linnaeus 1758; Hanski and Cambefort 1991) and ecologists (e.g. Halffter and Matthews 1966; Halffter and Halffter 1989) alike, the 'true' dung beetles (Philips et al. 2004) remain a taxonomic conundrum (Villalba et al. 2002). Due to the large range of structural diversity in the group (Philips et al. 2004), and lack of traditional species descriptions, field study specimens are often assigned to arbitrary numbers (e.g. *Uroxys sp.1*; Andresen 2005) and placement into the contentious higher taxa (Cambefort 1991; Villalba et al. 2002; table A.1) is nearly impossible to a non-specialist. Taxonomic inadequacies precludes correlation of studies carried out by different experts at different locations and times, while the taxonomic impediment means experts are a dwindling resource (Janzen 2004).

1.2. DNA taxonomy

Use of DNA sequences in the identification and delimitation of taxonomic groups has gained widespread acceptance and application in the last decade (see Hebert et al. 2003a; Tautz et al. 2003; Monaghan et al. 2005). DNA barcoding has been shown to give 100% accurate identification of traditional Lepidoptera species (Hebert et al. 2003a) and is already employed for complex groups such as nematodes, mosquitoes and crustaceans (Floyd et al. 2002; Besansky et al 2003; Proudlove and Wood 2003). Barcoding has also proved a success uncovering cryptic species (Hebert et al. 2004a) and in species delineation (Monaghan et al. 2005). Sequence divergence values for conspecific and congeneric pairwise comparisons provides an unambiguous and quantitative criteria for determining the status of an individual (Hebert et al. 2004a). Hebert et al. (2003b) found in pairwise comparisons of 891 traditional beetle species sequenced for the mitochondrial COI gene, 96.2% had sequence divergence (K2P) greater than 2%, with an average of 11.2%. This

compares to intraspecific divergence of generally less than 1% (Avise 2000). DNA barcoding of a mitochondrial gene region can also identify individuals of higher taxonomic groups (Hebert et al. 2003a; Blaxter 2004).

1.3. Barcoding biodiversity

Standardized sampling techniques, high-throughput DNA sequencing of short mitochondrial sequences and phylogenetic analyses to identify specimens could provide a new level of comparability, accuracy and efficiency to many ecological studies. The objective of this study is to explore the potential of DNA barcodes as a tool in surveys of dung beetle diversity and investigate the transferability of DNA barcodes across multiple study sites. DNA sequences, 16S rRNA from the mitochondrial genome and 28S rRNA from the nucleus, were used to determine the number of 'genospecies' or 'MOTU' (Molecular Operational Taxonomic Units; Blaxter 2004) at three study sites in Belize. COI was the marker of choice and is the traditional barcode region. However, in spite of continued efforts, primers failed to amplify the sequence in beetles from two sites so subsequently the sequence was not used in any analyses.

Biogeographical history of the Scarabaeinae means Belize, at the northern edge of the neotropics, hosts a diverse representation of dung beetles including Gondwanaland elements from South America and groups with Afro-Eurasian ancestry from the neartic (see Davies, Scholtz & Philips 2002; appendix A). Individuals of the same taxonomic species are expected to group together into discrete, exclusive genetic clades, so the clustering of the sequences in phylogenetic analyses was used as the primary criterion for determining 'operational taxonomic units'. The three study sites have each been subjected to a different history of natural and anthropogenic disturbance. Chiquibul Forest Reserve was severely damaged by Hurricane Hattie in 1961, Hurricane Iris devastated the Golden Stream Corridor Preserve in 2001 and Bladen Nature Reserve also experienced moderate disturbance during this storm. DNA barcodes for the specimens collected provides a record of diversity at each site and the pattern found was consistent with lower diversity in the more disturbed forests. The Chiquibul Forest Reserve has been sampled on two occasions in 1998 (Inward 2002) and 2004 (Powell 2004) providing a test of the applicability of barcoding biodiversity surveys over time as well as over distance.

2. Materials and methods

2.1. Study Sites

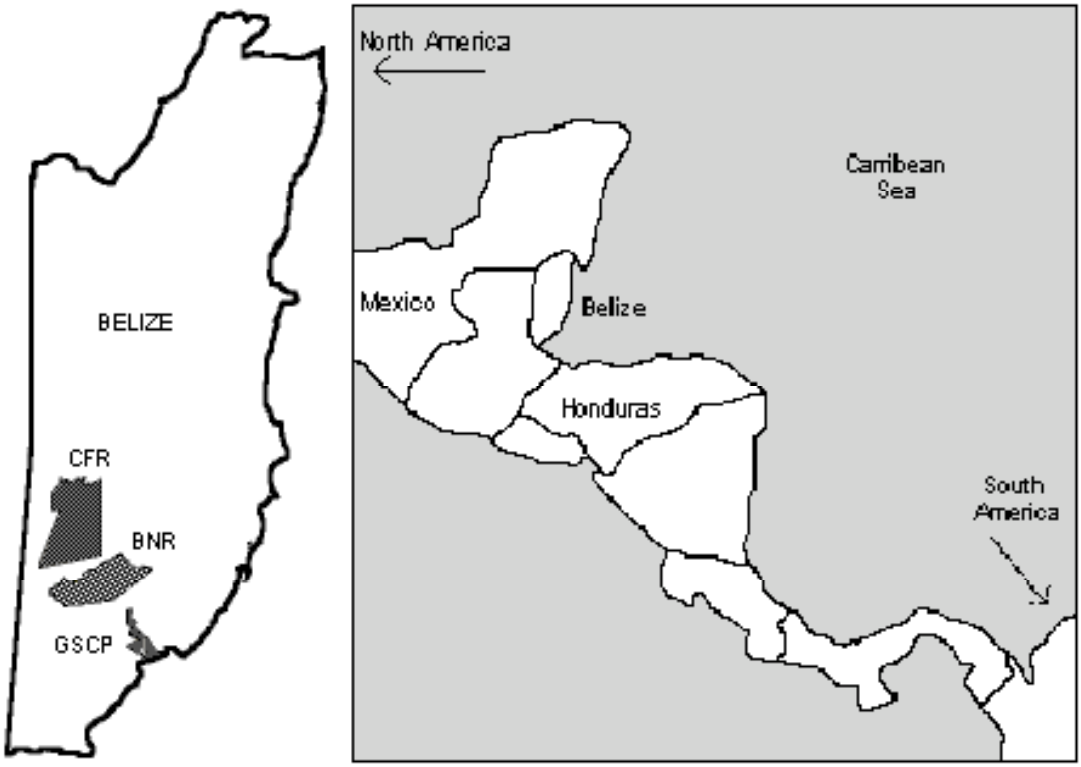


Figure 2.1. The three protected areas used as the study sites and the location of Belize in Central America.

Dung beetles were sampled from three study sites in Belize (figure 2.1):

Chiquibul Forest Reserve

Sampling was conducted around the Las Cuevas Field Research Station. The 77, 000 hectare reserve is situated upon hilly cretaceaous limestone geology and is essentially a secondary forest in structure since it has been subject to frequent disturbances in its history. In 1961 Hurricane Hattie devastated vast areas of the forest, meaning that the canopy height is presently relatively low in most areas, except in the lee of some hills (Inward 2002).

Bladen Nature Reserve

Bladen, an area of 100, 000 hectares, is the largest protected area in Belize, and has experienced only minor disturbance in the past. The area where trapping was conducted, adjacent to the BFREE Field Research Station, was badly hit by Hurricane Iris in October 2001 although not on the scale seen at YCT. This southern half of the forest is a very rugged limestone karst section pitted by conical limestone hills and ledges.

Golden Stream Corridor Preserve

Sampling was conducted around the YCT Field Centre. The preserve, some 20,000 hectares has severe natural disturbance. Hurricane Iris struck in October 2001, resulting in low canopy and very dense understorey at the site currently.

2.2. Field sampling

A standardised trapping protocol (Inward 2002) using baited pitfall traps (figure 2.2), with carrion (raw chicken) and cattle dung bait, was followed at each site. Bait was wrapped in a piece of muslin and suspended on a wire over the mouth of a 250cm^3 vending style cup sunk into the ground. Traps attract insects in the vicinity by odour (Larsen 2005). Since the traps are liable to flooding a plastic plate, held above the trap by stakes, served as a rain roof whilst still allowing access by the insects. Each trap contained approximately 100cm^3 of a solution of salt, detergent and water. The detergent reduces the surface tension of the water, so any insects entering the trap are quickly drowned, whilst the salt acts as an anti-bacterial agent, helping to preserve the specimens in the warm temperatures. Since the dung-baited traps lose their attractiveness quickly, they were re-baited with fresh dung on day two. The carrion bait did not need changing, as it only loses its attractiveness if it becomes desiccated, and actually improves as an attractant as its odour increases.

Flight interception traps were also used to collect specialist feeders, which may not have been attracted to any of these baits. These comprise a baffle of nylon mesh placed across a likely insect flight path, with preservative-filled foil trays beneath to collect insects. A length of tarpaulin is placed above to keep rain out of the trays. Flight intercept traps are particularly effective at collecting beetles, since they tend to fall downwards when they strike the netting in flight. They are most effective when employed amongst dense vegetation, since the black netting seemingly becomes invisible to flying insects in dull light conditions, hence they operate especially well in rainforest. One flight intercept trap was set up and run for approximately two weeks at each site with captured insects collected every 24 hours.

Sampling amounted to approximately 240 trap days at CFR and BNR, and 120 at GSCP (table 2.1; appendix B.2). This was believed to be sufficient to sample a significant proportion of the dung beetle assemblage at these sites (Inward 2002). Individuals from BNR and GSCP were sorted to externally distinct morphotypes in the field. At least five individuals (or all specimens if less than five were collected)

of each morphotype from each locatility were selected for DNA analysis; individuals within morphotypes selected randomly. This should ensure sequences were obtained from all or most species represented in the sample.

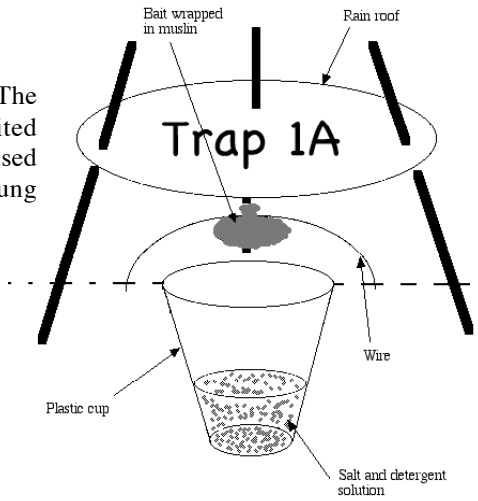


Figure 2.2. The standard baited pitfall trap used to collect dung beetles.

Table 2.1
Location, date of sampling and sampling effort used at each of the study sites

Site	Location	Date	Sampling	Disturbance[d]
Chiquibul Forest Reserve (CFR)	16°44'N; 88°59'W	January 1998[a]	12 transects; 6 of each bait type.	
		June 2004[b]	~9 transects; 5 with dung, 4 with carrion.	
Bladen Nature Reserve (BNR)[c]	16°33'N; 88°42'W	May/June 2005	12 transects; 6 of each bait type.	
Golden Stream Corridor Preserve (GSCP)[c]	16°22'N; 88°47'W	June 2005	6 transects; 3 of each bait type.	

Transects included ten pitfall traps (10m apart), were run for approximately four days, collecting the catch every 24hrs, and were >500m apart. Coordinates were recorded with a handheld GPS receiver.
[a] Sampling by Inward (2002).
[b] Sampling by Powell (2004).
[c] BNR and GSCP were sampled by the author.
[d] Representation of disturbance at the three study sites, based on reported hurricane damage and personal observations.

2.3. DNA extraction and sequencing

For the 2005 samples, collected individuals were preserved in either 100% ethanol or 99% isopropyl alcohol. Total DNA was extracted from heads, legs or whole beetles using the Wizard SV 96 Genomic DNA Purification System (Promega Corporation; appendix B.3). PCR amplification was performed in a volume of 25µL in 1µM Buffer, 3.6µM $MgCl_2$, 0.1µM dNTPs, 0.1µM each primer and Taq DNA polymerase. Sequences of 16S rRNA were amplified as a single fragment of ca. 500

bp, using primers 16Sar (5'-CGCCTGTTTATCAAAAACAT-3') (Palumbi et al. 1991) and 16Sb2 (5'-TTTAATCCAACATCGAGG-3'). A single fragment ca. 650 bp of 28S rRNA was amplified using primers 28S DD (5'-GGGACCCGTCTTGAA ACAC-3') and 28S FF (5'-TTACACACTCCTTAGCGGAT-3') (Inward 2002). Thermal cycling parameters comprised a denaturation phase of 94°C for 30 seconds, extension at 72°C for 60 seconds, and an annealing temperature of 48°C for 30 seconds. The conditions were cycled 40 times.

Amplification products were purified using a Multiscreen PCRμ96 filter plate (Millipore Corporation; appendix B.5). The plate uses a size exclusion membrane so the product can then be washed of impurities such as unincorporated primers and dNTPs, and resuspended into double-distilled H_2O. Sequencing was performed in both directions using a Big Dye v.1.1 terminator reaction with the same primers used for PCR (appendix B.6). Sequencing reactions were purified by ethanol precipitation (appendix B.7) and analysed on an ABI 377 automated sequencer.

Sequences were trimmed by removing ambiguously resolved parts of the 3' and 5' ends and ambiguities within the sequence were edited using Sequencher 4.5 (Gene Codes Corporation). Sequences of beetles from CFR were obtained from DungBeetleDatabaseAPV8 (BMNH accessions: 668541-668618, 668637-668701). Alignment of 16S and 28S sequences was performed independently with ClustalX (Thompson et al. 1997) using the default parameters (gap cost 15, extension cost 6.66). The alignments were improved by eye in MacClade 4.0 (Maddison & Maddison 2000) using a parsimony tree produced (in PAUP*4b10; Swofford 2002) from the ClustalX output as a guide.

2.4. Phylogenetic analysis

Three types of analyses were conducted; maximum parsimony (MP), Bayesian inference and TCS network analyses. MP and Bayesian trees were rooted with a single outgroup, *Psammodius porcicollis,* from the Aphodiinae (BMNH accession: 679909). Use of a single outgroup taxon was justified, as the objective here is not a detailed exploration of internal phylogenetic relationships within the Scarabaeinae. MP trees were obtained using PAUP*, with gap characters coded as a 'fifth base', and branch lengths optimised under accelerated transformation. Heuristic searches were performed using TBR branch swapping (MulTrees option off) and

1000 random addition replicates saving only a single tree in each case. Because the dataset contained many identical or very similar haplotypes, a large number of trees were found, one of which was selected arbitrarily for further analysis. Robustness of the inferred trees was tested by 'fast' bootstrapping (Felsenstein 1985) with 10,000 replications in PAUP*. The significance of incongruence in phylogenetic signal between genes was tested with the incongruence length difference (ILD) test (Farris et al. 1995) as implemented in PAUP* (PHT) on a combined matrix using 100 homogeneity replicates and default heuristic search parameters.

Bayesian analyses were conducted for the 16S sequences with MrBayes 3.04 (Huelsenbeck and Ronquist 2001), using a GTR+I+Γ model (the optimal model explaining the data as estimated with Modeltest 3.7; Posada and Crandall 1998). I used the default priors (uniform probabilities) starting with random trees, and ran the three heated and one cold Markov chains for 500,000 generations, sampled at intervals of 100 generations. The log-likelihood scores were plotted against generation time, and it was visually determined when the log-likelihood values (and Markov chains) reached a stable equilibrium. To test for the risk of the analysis being trapped in local optima, I repeated the procedure twice, beginning with different starting random trees. If the log-likelihood scores were similar, indicating convergence of the two analyses, the trees (once burn-in samples were discarded) from each analysis were combined in a majority rule consensus, and the percentage of the nodes were taken as a posterior probabilities of the clades under the assumed models, values of 95% or greater considered to be significantly supported (Rannala and Yang 1996).

I used TCS 1.21 (Clement et al. 2000) to generate haplotype cladograms for the 16S sequences with 95% parsimoniously plausible connections between haplotypes, considering gaps as a fifth state (Templeton et al. 1992). Statistical parsimony analysis is useful in this study as it partitions variation into homoplastic and non-homoplastic components (Monaghan et al. 2005). Consequently separate networks can be interpreted as distinct evolutionary lineages.

2.5. Analysis of mtDNA diversity

Variation within and among clusters of 16S sequences was examined using analysis of molecular variance (AMOVA) (Excoffier et al. 1992) of pairwise

differences as implemented in Arlequin 2.000 (Schneider et al. 2000). Clusters were based on the MOTU groupings from the phylogenetic analyses (see below). Using AMOVA the correlation among MOTU distances is used as an F-statistic analog. F_{ST} estimates the proportion of genetic variation within MOTU relative to genetic variation from the whole sample of sequences. MOTU with a single representative were excluded from the analysis because within-group variation could not be measured. Additionally uncorrected *p*-distances within and between clusters were calculated with PAUP*. Because of the huge number of possible pairwise comparisons (i.e. 351x351) I limited calculations to values of *p*-distances within MOTU and between adjacent MOTU pairs on the 28S MP tree (figure 3.1) or nearest 'new' MOTU on the 16S MP tree (figure 3.2). These values are included in table 3.1.

3. Results

Less than 50% of specimens selected for sequencing from BNR and GSCP were successfully incorporated into the 28S analysis. 16S sequencing performed better with 80% of all BNR and GSCP specimens selected having sequences successfully incorporated.

Four hundred and eleven variable positions (93 parsimony informative) in the aligned 28S matrix (667 characters) produced 36 genotypes. One of these (the morphologically identified specimen of *Canthon cyanellus*) was believed to be an erroneous construction due to spurious optimisation of missing characters in this unusually short sequence. The sequence was inspected visually and found to be identical to the Motu24 genotype over its entire (albeit short) length and was not considered as a separate MOTU in the following analyses. Genotypes differed by a minimum of one nucleotide.

MP trees for the mtDNA sequences (478 characters, 216 parsimony informative positions) had a very similar topology to the Bayesian inference trees and uncovered 36 clusters of similar 16S sequences, plus 4 isolated sequences without close relatives. As with the MP tree set recovered, MP and Bayesian trees had minor differences in the interrelationship of clusters and in the resolution of haplotypes within such clades but neither of these differences influence the general interpretation of the results. The replicate Bayesian analyses were visually compared and revealed a similar topological structure, suggesting convergence of the two analyses to the same general solution. Although I discuss phylogenetic results in the context of the MP tree presented in figure 3.2, the discussion applies equally well to Bayesian trees (a Bayesian consensus tree is presented in figure 3.3).

Clustering of 16S sequences showed complete congruence with the 28S genotypes. All individuals with the same 28S genotype formed monophyletic clusters of 16S haplotypes, consistent with a slower rate of evolution in nuclear genes. In the combined analysis of both sequences, incongruence length difference was not significant (P=0.1). Additional clusters not represented by 28S genotypes were found in the 16S analysis for beetles with a history of poor 28S amplification (the Phanaeini tribe; AP Vogler, personal communication). This made a total of 40 genospecies overall (35 28S genotypes and 5 additional MOTU comprising Phanaeini and *Deltochilum* clusters). For purposes of this study I designate a MOTU

as a cluster of individuals with similar 16S sequences and identical 28S genotypes or if no 28S sequence was available for the specimen, the individual's 16S sequence grouped within these clusters. For groups where 28S sequencing failed completely, monophyletic groups of individuals with 16S sequences differing by 0.02 (uncorrected *p*-distance) or less were designated MOTU.

With three exceptions, individuals of morphologically defined species (type sequences) grouped into separate well-defined MOTU clusters. Two specimens identified as *Uroxys sp1* fell into two separate clusters; *Canthon sp1* and *sp2* formed a single grouping. A specimen of *Onthophagus nitidior* was included in the *Onthophagus longimanus* MOTU, and did not cluster with the other specimen identified as belonging to this species. Two of these cases involve undescribed morphospecies and the third the extremely difficult to identify Onthophagus tribe.

Most clusters were well supported as monophyletic with high bootstrap values and show the characteristic of low sequence divergence within, but high sequence divergence between clusters, seen in other barcoding studies (Monaghan et al. 2005). Maximum divergence (uncorrected *p*-distances) of sequences within tip clades ranged from zero to 0.087 (although this top value could represent multiple genospecies) and was typically <0.01 while between cluster distances were typically >0.03 (table 3.2), exceptions are discussed below. Uncorrected *p*-distances measured within and between the tip clade clusters were mostly non-overlapping (table 3.2). Based on pairwise differences in AMOVA, within group variation accounted for only 2% of the total variation in the data set (F_{ST}=0.9800).

Some of the ambiguous mtDNA clusters matched incongruence between the number of networks and haplotypes produced in the TCS analysis, and tip clusters on the MP and Bayesian inference trees. Statistical parsimony analysis of the 84 16S haplotypes within the limits of parsimony (*sensu* Templeton et al. 1992), i.e. the number of steps by which two haplotypes have a 95% statistical probability of being linked without homoplasy, was calculated to be 30 steps or less. The analysis resulted in 39 independent networks within which connections between each of the haplotypes fulfilled this criterion. These networks correlated with clusters identified with MP and Bayesian analyses with a typical ratio of a single network per MOTU. There were a few exceptions. Haplotypes of the Phanaeini tribe (Motu36, Motu37 and Motu38) formed a large single network including the haplotypes from Motu4. Motu16 and the Motu17 singleton which grouped together closely on all trees

formed a single network. Two *Canthon* groupings Motu21 and Motu22 (see later) formed a single network. The unusually divergent haplotypes in Motu6 formed four separate networks and Motu11 formed two networks.

Table 3.2
Dung beetles included in the present study

Motu	Morphological identification[a]	Locality present[b]				No. of individuals 28S	No. of individuals 16S	Haplotypes in TCS network (number of networks)	Maximum pairwise distance (16S sequences)[c]	Minimum pairwise distance (16S sequences)[d]
		CFR98	CFR04	BNR	GSCP					
1	*Eurysternus sp1*	-	X	X	X	15	14	3 (1)	0.004	0.038
2	*Eurysternus magnus*	-	X	-	-	6	7	3 (1)	0.004	
3	*Eurysternus angustulus*	-	X	-	-	8	13	1 (1)	0.000	0.037
4	*Eurysternus carabaeus*	X	X	X	X	11	16	10 (0.25)	0.022	
5		-	X	X	X	5	10	3 (1)	0.004	0.131
6		X	X	X	-	4	5	5 (4)	0.087	
7	*Ateachus chrysopyge*	X	X	-	-	16	20	1 (1)	0.000	0.057
8		-	-	X	X	3	1	1 (1)	-	
9	*Dichotomious amplicollis*	-	X	-	-	5	5	3 (1)	0.006	0.066
10	*Dichotomious satanus*	-	X	-	-	1	4	2 (1)	0.002	
11	*Uroxys sp1*	X	X	-	-	8	8	2 (2)	0.019	0.091
12	*Uroxys sp1*	X	X	-	-	5	4	3 (1)	0.038	
13		X	-	-	-	1	1	1 (1)	-	0.157
14		X	X	-	-	2	2	1 (1)	0.000	
15		-	-	X	-	4	4	1 (1)	0.000	0.099
16	*Copris laeviceps*	X	X	X	X	17	23	6 (0.5)	0.000	
17		-	-	X	-	1	1	1 (0.5)	-	0.111
18		X	X	-	-	3	3	1 (1)	0.000	
19		-	-	X	-	5	10	3 (1)	0.004	0.083
20		-	X	-	X	0	7	3 (1)	0.033	
21	*Canthon subhyalinus*	X	X	-	-	11	12	5 (0.5)	0.004	0.016
22	*Canthon sp1+2*	-	X	-	-	11	12	2 (0.5)	0.004	
23	*Canthon lamprimus*	-	X	X	-	4	5	2 (1)	0.000	0.032
24	*Canthon cyanellus*	-	X	X	X	18	21	3 (1)	0.000	
25	*Onthophagus sp1*	-	X	-	-	7	8	2 (1)	0.006	0.042
26	*Onthophagus anthracinus*	-	X	X	-	7	8	2 (1)	0.002	
27	*Onthophagus rhinolophus*	X	X	-	-	3	4	2 (1)	0.008	0.046
28		-	-	X	X	3	2	2 (1)	0.002	
29		X	-	-	-	1	1	1 (1)	-	0.123
30		-	-	X	-	1	2	1 (1)	-	
31		X	-	-	-	3	3	2 (1)	0.011	0.089
32	*Onthophagus crinitis panamensis*	X	X	X	X	23	25	4 (1)	0.000	
33	*Onthophagus longimanus*	X	X	-	-	13	15	3 (1)	0.010	0.049
34	*Onthophagus nitidior*	X	X	-	-	3	4	1 (1)	-	
35		-	-	X	-	4	2	1 (1)	-	0.142
36	*Coprophanaeus telamon*	-	X	X	X	0	44	10 (0.25)	0.018	
37	*Phanaeus sallei*	X	X	X	-	0	5	2 (0.25)	0.000	0.030
38	*Phanaeus endymion*	-	X	X	X	0	6	1 (0.25)	0.012	
39	*Deltochilum gibbosum*	-	X	-	-	0	3	2 (1)	0.002	0.092
40	*Deltochilum pseudoparile*	-	X	X	X	0	11	5 (1)	0.016	
40	**24**	**17**	**30**	**20**	**12**	**232**	**351**	**84 (39)**	**0.009 (Ave.)**	**0.038 (Ave.)**

[a] Type sequence specimens identified by F. Krell, Scarab Research Group, Entomology Department, NHM.
[b] See table 2.1 for location codes. X indicates presence of MOTU at the locality.
[c] Based on pairwise base distances (uncorrected *p*) as calculated by PAUP*.
[d] Pairwise base distances between MOTU pairs (uncorrected *p*) as calculated by PAUP*.

4. Discussion

4.1. What makes a good DNA barcode?

Despite appearing to be a 'true' barcode (i.e. invariant within species and variable between species), and even if this is a real and widespread phenomenon, several disadvantages means the 28S sequence would not make a good barcode region in reality. The high failure rate of PCR amplification and abundant contamination with the 28S primers could negatively influence its use as a DNA barcode target sequence. The relatively high concentration of mtDNA means amplification of these genes is more likely to be successful. The 28S genotype groupings recovered cannot be considered species by a genealogical species concept (Hudson and Coyne 2002) because some groups were not reciprocally monophyletic (e.g. Motu3 was a paraphyletic grouping). Although a 28S genotype would hint at reciprocal monophyly in mtDNA another gene sequence would be needed to confirm this. One base pair difference would result in the establishment of a new genospecies. Unfortunately DNA sequencing is not always this reliable, and in the current widely used databases (i.e. GenBank) no sequence chromatograms are provided. Therefore it is difficult to assess the certainty of any 'questionable' base call. This was a problem I encountered with sequences from the Scarabaeinae database (AP Vogler, unpublished), and is being addressed by the online COI bank (Hebert et al 2005), which archives electropherograms with submitted sequences. Results presented here question whether 28S is actually a 'true' barcode. Four putative genospecies (i.e. within Motu6) possessed the same 28S sequence despite grouping into separate 16S clusters and TCS networks with *p*-distances typical of independent MOTU (although three are singletons). However, when these haplotypes were treated as separate clusters in an AMOVA analysis no major increase in the F_{ST} value was observed (i.e. when the three sequences representing possible new MOTU within Motu 6 were removed F_{ST}=0.9848). This compares with a major decrease in F_{ST} when two definitive MOTU are treated in the analysis as a single cluster (e.g. with motu7 and motu24 combined F_{ST}=0.9382) also noted by Monaghan et al. (2005). The pattern seen in Motu6 is perhaps not surprising as nuclear sequences generally evolve more slowly than mtDNA, so this could represent speciation in progress (MT Monaghan, personal communication)! Unusually high within genospecies distances was also seen in Motu20. This could in

fact represent two MOTU (it is not strictly a MOTU under my definition as only one sequence of 28S was analysed) with an obvious separation into the two clusters with high bootstrap support seen on the MP tree, although TCS analysis produced a single network for the current grouping. Further investigation and sequencing of more individuals is needed to resolve these groupings. The minimum distance between Motu21 and Motu22 was unusually low (0.016), but still well above the distances seen within this MOTU pair. These two groupings are obviously very closely related taxa judging by the adjacent positions on the 16S tree and one base pair separates the 28S genotypes. They could represent recently speciated sibling species, and presents an opposite case to the two above with nuclear segregation, yet short mitochondrial distances. However, establishment of 28S as a DNA barcode was not the aim of this study, and the sequence performed well as a measure of congruence between mtDNA and nuclear markers, supporting the establishment of MOTU based on monophyletic clusters of 16S sequences. Although individuals were grouped into clusters in the exact same way whether based on 28S or 16S sequences (16S may subdivide groups further) paraphyletic groupings in 28S means the exclusivity criterion of species delimitation (Baum and Shaw 1995; Sites & Marshall 2003) is not fulfilled on all occasions.

Past work has provided conflicting perspectives on the likely efficacy of mtDNA markers in delineating species boundaries (see Lipscomb et al. 2003, Moritz and Cicero 2004). However, as a barcoding region 16S fills the criteria of low sequence divergence within, but high sequence divergence between 'species' and with a few exceptions, these divergence values are completely non-overlapping. Genealogical concordance (Avise and Ball 1990) is prevalent in the data i.e. with minor exceptions, specimens identified using morphological criteria fall into separate, well-defined molecular clusters. The few morphospecies that did not fit into a single MOTU are mostly undescribed species with the arbitrary numbering system discussed above. This demonstrates how correlating barcodes with traditional morphological species can only be as good as the original morphological identification. In the majority of cases using sequences of 'known' taxa as comparators we can assign genospecies to described dung beetle taxa. For the remaining clusters I hypothesise these represent either an undescribed species or species for which no type sequence was included. A parallel morphological survey will be needed to determine in detail how the diversity measured by molecular

methods relates with that found by traditional classification and 'species' under any concept (see Mayden 1997).

Although the alignment was reasonably straightforward in the two genes, there is the possibility that ambiguous homology in hyper-variable regions could dilute the effectiveness of the barcodes. One approach could be that followed by Floyd et al. (2002); aggressive alignment to remove potentially noisy data (i.e. all missing and indel positions). When this was attempted (results not presented) many taxa with missing terminal data had to be removed to avoid a very short barcode and results were still almost identical to those obtained with missing data and gap positions included. The aggressive alignment approach is not very practical for large-scale biodiversity studies where sequences obtained are likely to vary in length and quality in all possible ways. To produce the most informative study overall and include as much data as possible, judgement calls in the alignment and inclusion of missing data are unavoidable. This does add a level of subjectivity to the barcodes, and this study in particular, which could potentially be avoided with a protein-coding sequence. Any sequences submitted to barcoding databases as references (type sequences in particular) should be good quality (i.e. clean, single-peaked chromatograms) and sufficient length so that short fragments can be unambiguously aligned with them.

4.2. What makes a good MOTU?

Species delineation in DNA taxonomy is still problematic. The various methods attempted in this study, namely statistical parsimony networks, phenetic distance thresholds, thresholds based on calculations adapted from population genetics analysis, and genealogical concordance, could not be applied universally to the data and provide ‘satisfactory’ results. Despite clear and distinctive molecular clusters on the phylogenetic trees, separating these clusters in an unambiguous and quantitative way remains problematic, and ultimately the clusters have an intuitive element, similar to the traditional species boundaries, that DNA taxonomy was trying to counter. The current consensus on methods to separate species seems to be *p*-distance thresholds. But these are not universal (e.g. a sequence divergence >0.03 and you have a new species) and are currently applied on a case-by-case basis (see Hebert et al. 2004a; Hebert et al. 2004b; Hogg and Hebert 2004). Previous studies have mostly involved a priori identified taxonomic species, so to a certain extent the

limits of species are already determined. Barcoding is in its infancy and as application extends, more scientists will become faced with these difficulties. A solution may come to hand, or sequence divergence measures could prove to be adequate.

4.3. Dung beetle diversity

Any comparison of dung beetle diversity at the sites in this study, is heuristic as the standardized trapping protocol was not followed strictly on all occasions (e.g. in 2004 only nine transects were used and no flight intercept traps, and at GSCP only half the trapping regime i.e. six transects could be employed due to time constraints; see table 2.1), but it does provides a context in which to examine the barcoding capability for the group. It is unpractical to sequence every individual collected, and the selection method used; selecting individuals for sequencing based on distinct morphology was in line with practice followed in other barcoding studies (Hogg and Hebert 2004; Monaghan et al. 2005). However, this approach could benefit from a statistical measure of uncertainty in obtaining sequences from all MOTU in the sample, and sampling all the MOTU at the site.

The differences in sequences (and MOTU) found on the two sampling trips to CFR was pronounced and this is the subject of another MSc thesis (Powell 2004). Sampling was conducted at different times of the year, and dung beetles assemblages are known to be seasonal, with abundances varying between the rainy season, dry season, and transition periods (see Inward 2002). Nonetheless, 13 of the MOTU in this study were found at CFR both years. Comparing different sites sampled at the same time of year (i.e. CFR 2004, BNR, GSCP) CFR has a lot of species represented by very few individuals and not seen at the other sites. Sampling on this occasion included canopy traps and pitfall traps in grass areas around the field station, which could artificially inflate diversity at this site. CFR does appear to be much more diverse (29 MOTU in 2004). This was expected as Chiquibul Forest escaped the impact of Hurricane Iris in 2001, which damaged areas of the other two reserves. The effects of disturbance were plainly seen in the sample from GSCP where two species (subsequently identified by their barcodes as *Coprophanaeus telamon* and *Canthon cyanellus*) dominated during the trapping period, each comprising 25% of the total catch. Dominance of a few dung beetle species is often associated with disturbed habitats (Andresen 2005). Even with the shorter trapping period, it is obvious this

site has very low diversity, only 12 of the MOTU in this study, were found at GSCP and very few beetles were collected in total (178 compared to >300 at BNR with the same number of traps).

4.4. Barcoding biodiversity surveys: already a reality?

This is a preliminary study and although some specimens were problematic and could not be placed definitively into 28S genotypes or 16S tip clusters, mainly due to the short sequences of these individuals and spurious optimisation of missing data, 99% of specimens were placed conclusively within genospecies. These ambiguous taxa however, are not insignificant as fragmented pieces, sequencing errors and crude phylogenetic reconstruction methods are likely to be features of barcoding used in the field, by para-taxonomists and amateurs. Therefore these problems need to be addressed if barcoding is to fulfil its potential.

The fact that identical sequences were sampled from sites 50km apart and at the same site six years later shows mtDNA barcodes can travel and are transferable between studies. 60% of genospecies were found on more than one sampling expedition, 25% were found at all three sites and 10% were found every time sampling was conducted. This partly reflects the unstable and changeable nature of dung beetles communities (Inward 2002), so the fact we found stable molecular sequences within this group is very promising. DNA barcodes provide a link between individuals collected in different localities or in different studies in a way that arbitrary morphospecies designations (e.g. *Uroxys sp1*) do not, and immediately provides the data needed for future DNA barcoding identification.

If dung beetles are to continue being used as indicators of forest disturbance, 'operational taxonomic units' (Blaxter 2004), need to be identified easily. DNA barcodes differ from the standard traits used for species discrimination in the following important way: they can be obtained in a mechanised manner (Hebert et al. 2004a). Hence they can be used without much background knowledge, are easily databased and easily compared. A database of sequences would create an identification system for Scarabaeinae dung beetles that enables identification of specimens by anyone with access to a basic sequencing laboratory (Hogg & Hebert 2004). This approach allows MOTU to be designated objectively and repeatedly by anyone using the sequence database. As a result, mtDNA data can form the basis of testable taxonomic hypotheses that could be examined with additional types of data

in the future (such as Motu6 in this study). It provides rapid division into probable groups of reproductively isolated individuals and provides a perspective on their evolutionary past (Avise & Walker 1999). Distinctive morphological characters correlated with MOTU and referenced on the database, could aid preliminary identification of specimens and groups in the field, until the barcoder's enthusiastic vision of hand-held DNA sequencers or chip-based DNA arrays (Stoeckle et al. 2004; Janzen 2004) becomes reality.

4. 5. Summary

Short sequences of mitochondrial and nuclear DNA facilitates placement of dung beetle specimens into distinctive genetic clusters, which I designated as MOTU. With minor exceptions these groupings correlated well with the few specimens of morphologically identified species included in this study and had non-overlapping levels of sequence divergence i.e. low divergences within the group (typically <0.01 for intra-cluster divergences) and large divergences between MOTU (typically >0.03). Individuals were separated into 40 MOTU found over three forests in Belize. 60% were found on more than one sampling expedition, 25% were found at all three sites and 10% found every time sampling was conducted. Some of the groupings were ambiguous and further investigation is needed (including sequencing of more specimens) to determine the most practical and definitive way of characterising these groupings. Further work would include an examination of how MOTU based on 16S and 28S sequences compare with COI-5', the traditional barcoding region used in most studies, and with morphological species boundaries. The results presented demonstrate that an identification system based on DNA barcoding is feasible for the Scarabaeinae, and would provide an efficient way of measuring changes in the diversity between locations and over time in dung beetle assemblages.

6. Acknowledgements

Thanks are owed to the following. Daegan Inward who designed the trapping protocol and provided invaluable advice on fieldwork, and also for the use of his DNA collection from Belize. Liz Powell who also collected and sequenced beetles from the Chiquibul Forest Reserve. Jacob and Kelly Marlin (BFREE), Nick Wicks and everyone at YCT for assistance in Belize. Hector Mai, The Forestry Department and Government of Belize for allowing research and exportation of dung beetles from their country. Sam Bridgewater my NHM contact in Belize. Ruth Wild, who has been very supportive through the practical work and analysis, and provided excellent training in the laboratory protocols. All the helpful people in the molecular systematics laboratory and my good friends on the MSc course. Michael Monaghan for direction and assistance with the data managing side of molecular phylogenetic analysis. Alfried Vogler for support and understanding through what has been a difficult couple of months!

7. References

Andresen E (2005) Effects of season and vegetation type on community organization of dung beetles in a tropical dry forest. *Biotropica* **37**: 291-300.

Avise JC (2000) *Phylogeography. The History and Formation of Species*. Harvard University Press, Cambridge, MA, USA.

Avise JC & Ball MR (1990) Principles of genealogical concordance in species concepts and biological taxonomy. *Oxford Surveys in Evolutionary Biology* **7**: 45-67.

Avise JC & Walker D (1999) Species realities and numbers in sexual vertebrates: perspectives from an asexually transmitted genome. *Proceedings of the National Academy of Science* **96**: 992-995.

Baum DA & Shaw KL (1995) Genealogical perspectives on the species problem. In '*Experimental and Molecular Approaches to Plant Biosystematics*' (Eds. PC Hoch & AG Stephenson) pp 289-303. Missouri Botanical Garden.

Besansky NJ, Severson DW & Ferdig MT (2003) DNA barcoding of parasites and invertebrate disease vectors: what you don't know can hurt you. *Trends in Parasitology* **19**: 545-546.

Blaxter ML (2004) The promise of DNA taxonomy. *Philosophical Transactions of the Royal Society of London Series B* **359**: 669-679.

Blaxter ML & Floyd R (2003) Molecular taxonomics for biodiversity studies: already a reality. *Trends in Ecology and Evolution* **18**: 268-269.

Cambefort Y (1991) Biogeography and evolution. In '*Dung Beetle Ecology*' (Eds. I Hanski & Y Cambefort) pp 22-35. Princeton University Press, Princeton, NJ, USA.

Clement M, Posada D & Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9**: 1657-1660.

Davies ALV, Scholtz CH & Philips TK (2002) Historical biogeography of Scarabaeinae dung beetles. *Journal of Biogeography* **29**: 1217-1256.

Excoffier L, Smouse P & Quattro J (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**: 479-491.

Farris JS, Kallersjo M, Kluge AG & Bult C (1995) Testing significance of incongruence. *Cladistics* **10**: 315-319.

Felstenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**: 783-791.

Floyd R, Abebe E, Papert A & Blaxter ML (2002) Molecular barcodes for soil nematode identification. *Molecular Ecology* **11**: 839-850.

Halffter G & Edmunds D (1982) The nesting behaviour of dung beetles. An ecological and evolutive approach. Institute de Ecologia, Mexico DF, Mexico.

Halffter G & Matthews E (1966) The natural history of dung beetles of the subfamily Scarabaeinae (Coleoptera, Scarabaeidae). *Folia Entomol Mex* **12-14**: 1-312 [Reprinted by Medical Books (2nd Ed.) Palermo, Italia].

Halffter G & Favila ME (1993) The Scarabaeinae (Insecta: Coleoptera) an animal group for analysing, inventorying and monitoring biodiversity in tropical rainforests and modified landscapes. *Biology International* **27**: 15-21.

Hanski I & Cambefort Y (1991) *Dung beetle ecology*. Princeton University Press, Princeton, NJ.

Hebert PDN, Cywinska A, Ball SL & deWaard JR (2003a) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London Series B* **270**: 313-322.

Hebert PDN, Penton E, Burns JM, Janzen DH & Hallwachs W (2004a) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Science* **101**: 14812-14817.

Hebert PDN, Ratnasingham S & deWaard JR (2003b) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London Series B* **270**: S596-S599.

Hebert PDN, Ratnasingham S and Door R (2005) Barcode of life http://www.barcodinglife.org (Accessed 01/09/2005).

Hebert PDN, Stoeckle MY, Zemlak TS & Francis CM (2004b) Identification of birds through DNA barcodes. *PLOS Biology* **2**: 1657-1663.

Hogg ID & Hebert PDN (2004) Biological identification of springtails (Hexapoda: Collembola) from the Canadian Artic, using mitochondrial DNA barcodes. *Canadian Journal of Zoology* **82**: 749-754.

Huelsenbeck JP & Ronquist F (2001) MrBayes: Bayesian inference of phylogeny. *Bioinformatics* **17**: 754-755.

Hudson RR & Coyne JA (2002) Mathematical consequences of the genealogical species concept. *Evolution* **56**: 1557-1565.

Inward DJG (2002) *The evolution of dung beetle assemblages*. PhD thesis, The Natural History Museum/ Imperial College London

Janzen DH (2004) Now is the time. *Philosophical Transactions of the Royal Society of London Series B* **359**: 731-732.

Larsen TH (2005) Trap spacing and transect design for dung beetle biodiversity studies. *Biotropica* **37**: 322-325.

Lipscomb D, Platnick N & Wheeler Q (2003) The intellectual content of taxonomy: a comment on DNA taxonomy. *Trends in Ecology and Evolution* **18**: 65-66.

Maddison WP & Maddison DR (2000) MacClade: Analysis of phylogeny and character evolution, version 4.0, Computer program distributed by Sinauer, Sunderland, MA, USA.

Mayden RL (1997) A hierarchy of species concepts: the denouement in the saga of the species problem. In '*Species "The Units of Biodiversity"'* (Eds. MF Claridge, HA Dawah & MR Wilson) pp 381-424. Chapman and Hall Ltd., London.

Monaghan TM, Balke M, Gregory TR & Vogler AP (2005) DNA-based species delineation in tropical beetles using mitochondrial and nuclear markers. *Philosophical Transactions of the London Society Series B, in press.*

Moritz C & Cicero C (2004) DNA barcoding: promise and pitfalls. *PLOS BIOLOGY* **2**: 1529-1531.

Nummelin M & Kaitala S (2004) Do species dominance indices indicate rain forest disturbance by logging? *Biotropica* **36**: 628-632.

Palumbi SP, Martin AP, Ramano S, McMillan WO, Stice L & Grabowski G (1991) *The simple fool's guide to PCR version 2*. University of Hawaii, Zoology Department, Honululu.

Philips TK, Pretorius E & Scholtz CH (2004) A phylogenetic analysis of dung beetles (Scarabaeinae: Scarabaeidae): unrolling an evolutionary history. *Invertebrate Systematics* **18**: 53-88.

Posada D & Crandall KA (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**: 817-818.

Powell E (2004) *A critique of DNA taxonomy for the use of studying long term local diversity of a dung beetle community in a tropical American forest*. MSc thesis, Imperial College London.

Proudlove G & Wood PJ (2003) The blind leading the blind: cryptic subterranean species and DNA taxonomy. *Trends in Ecology and Evolution* **18**: 272-273

Rannala B & Yang Z (1996) Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *Journal of Molecular Evolution* **43**: 304-311.

Scheffer PY (2004) Dung beetle (Coleoptera: Scarabaeidae) diversity and community structure across three disturbance regimes in eastern Amazonia. *Journal of Tropical Ecology* **21**: 9-19.

Schneider SD, Roessli D & Excoffier L (2000) ARLEQUIN version 2.000: A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Sites JW & Marshall JC (2003) Deliminating species: a renaissance issue in systematic biology. *Trends in Ecology and Evolution* **18**: 462-470.

Stoeckle M (2003) Taxonomy, DNA and the barcode of life. *Bioscience* **53**: 796-797.

Stoeckle M, Waggoner P & Ausubel J (2004) *Barcoding life: Ten reasons*. https://phe.rockefeller.edu/barcode/docs/TenReasonsBarcoding.pdf (Accessed 12/8/05).

Swofford DL (2002) PAUP* Phylogentic Analysis Using Parsimony (*and other methods). Version 4.0b10. Sinauer Associates, Sunderland, MA, USA.

Tautz D, Arctander P, Alessandro M, Thomas RH & Vogler AP (2003) A plea for DNA taxonomy. *Trends in Ecology and Evolution* **18**: 70-74.

Templeton AR, Crandall KA & Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* **132**: 619-633.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F & Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **24**: 4876-4882.

Villalba S, Lobo JM, Martin-Piera F & Zardoya R (2002) Phylogenetic relationships of Iberian dung beetles (Coleoptera: Scarabaeinae): insights on the evolution of nesting behaviour. *Journal of Molecular Evolution* **55**: 116-126.

Vulinec K (2000) Dung beetles (Coleoptera: Scarabaeidae), monkeys and conservation in Amazonia. *The Florida Entomologist* **83**: 229-241.

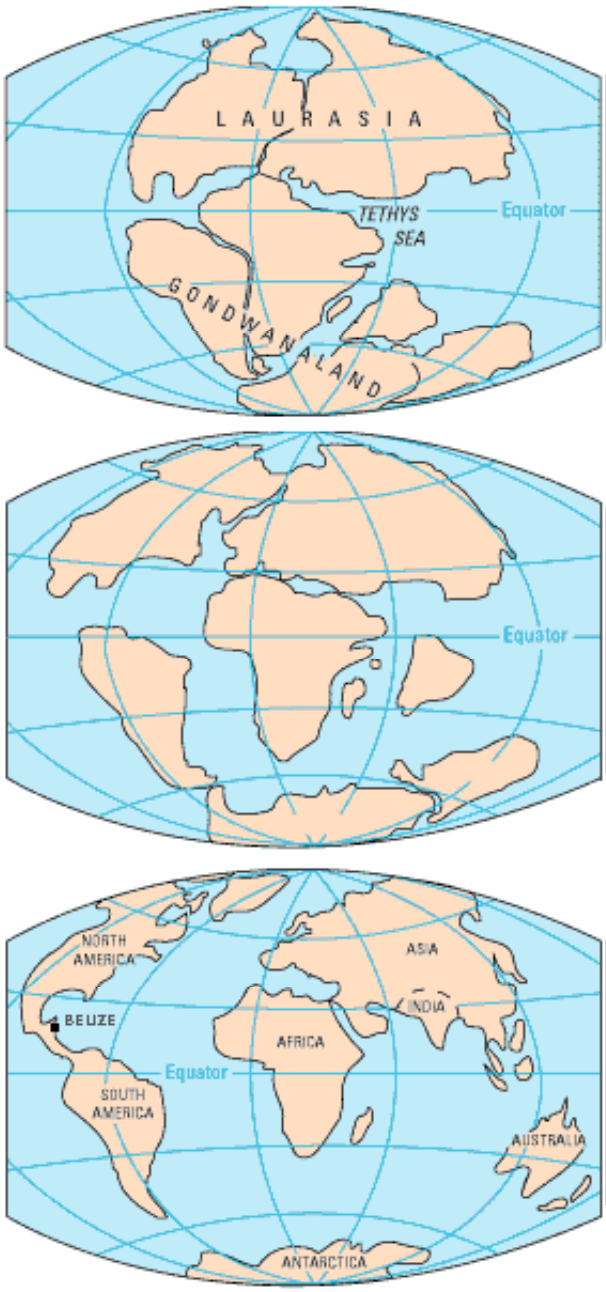
Appendix A: Taxonomy and historical biogeography of the dung beetles of Belize

Table A.1
World Scarabaeinae tribes, their relative ages, and number of known genera and traditional species in different geographical regions

	World		Neotropics		Belize		Age
Tribe	Genera	Species	Genera	Species	Genera	Species	
Canthonini	93	771	27	297	3	10	Gondwanan
Coprini	10	374	1	17	1	1	Modern
Dichotomiini	30	722	19	563	6	11	Gondwanan
Eucraniini	4	19	4	19	0	0	Intermediate
Eurysternini	1	26	1	22	1	6	Intermediate
Gymnopleurini	4	104	0	0	0	0	Intermediate
Ontiticellini	14	165	0	0	0	0	Modern
Onthophagini	35	2213	1	95	1	8	Modern
Onitini	18	195	0	0	0	0	Intermediate
Phanaeini	12	148	12	146	2	3	Intermediate
Scarabaeini	11	145	0	0	0	0	Intermediate
Sisyphini	3	58	1	2	0	0	Modern
Total	235	4940	66	1161	14	39	

Data from Villalba et al. 2002 (World), Davies et al. 2002 (Neotropics), Inward 2002 (Belize) and Cambefort 1991 (Ages).

Three principal long-term dispersal/vicariance events are hypothesized to account for the global distribution patterns of the Scarabaeinae (Davies, Scholtz & Philips 2002; summarized in figure A.1). Of the tribes present in the Neotropics, two (Canthonini, Dichotomiini) have a worldwide distribution with principle generic richness in the large southerly regions (Afrotropical, New Guinea, Madagascar, Australia and Neotropical). This is considered to represent a Gondwanaland pattern (Davies, Scholtz & Philips 2002), which may stem partly from their long periods of isolation from the late Cretaceous until the Pliocene (Neotropical), or the present (Madagascar, Australia), which barred all but over water immigrants. Three tribes (Phanaeini, Eucraniini, Eurysternini) show endemic distributions restricted, to the Americas. These Gondwanaland and indigenous groups are especially characteristic of Neotropical forest habitats but probably more recent taxa have also radiated widely into more open habitats and northward into the Neartic region. Two other tribes are found in the Central American fauna, Copris and Onthophagus, of which only the latter extends into South America. Both of these genera are widely dispersed around the world, but centred on the Afro-Eurasian region, and it has been suggested that these taxa colonised North America from Eurasia via the Bering bridge (Cambefort 1991). The clear African origin of Onthophagus makes this seem quite likely. These northern elements with Afro-Eurasian ancestry were probably also derived from the Nearctic during the Great American Interchange and have spread southward, although taxon richness decreases steeply from Central to South America. This was therefore a relatively recent event, and taxa (and assemblages) may still become established, particularly due to the intermingling of independently derived lineages (Inward 2002).



TRIASSIC 200 Mya
Mesozoic taxon radiation within Gondwanaland followed by continental fragmentation, vicariance and taxon divergence to tribal and generic levels.

CRETACEOUS 65 Mya
Some modern Afro-Eurasian elements reach Australia and North America followed by speciation in North America possibly caused by vicariance on two principle occasions, one in the Miocene and the other in the Pleistocene.

PRESENT DAY
Following the Great American Interchange derived Gondwanaland tribes moved northwards into North America coupled with southward movement of Neartic elements with derived Afro-Eurasian ancestry.

Figure A.1. Biogeographical hypotheses for distribution patterns of the Scarabaeinae (Davies, Scholtz & Philips 2002; Maps modified from https://pubs.usgs .gov/publications/text/historical.html [Accessed 10/08/2005]).

Appendix B. Laboratory protocols

B.1. DNA extraction

1. The beetle or beetle part (small beetles were cut in half, separating head and thorax, larger beetles the leg was used and cut into pieces) were placed in a well in a 96 well (deep) plate. 275µl of digestion solution (table B.1) was added to the well.

Table B.1
Composition of the digestion solution

	Volume per sample	Master mix for 96 samples
Nuclei Lysis Solution	200μl	22.0ml
proteinase K, 20mg/ml	20μl	2.2ml
0.5M EDTA (pH 8.0)	50μl	5.5ml
Total Volume	**270μl**	**29.7ml**

2. The plate was covered to stop evaporation and incubated overnight (16–18 hours) in a 55°C water bath. It is not necessary to shake the plate during the incubation.
3. Following overnight incubation at 55°C, 250μl of the Wizard SV lysis buffer was dispensed into each well of the deep well plate containing lysate. The contents of each well were mixed by pipetting several times. The lysates were processed as soon as possible after the Wizard SV lysis buffer has been added (lysate must still be warm).
4. The vacuum manifold was assembled according to figure B.1.A and the binding plate was placed in the vacuum manifold base. The binding plate was orientated in the vacuum manifold with the numerical column headers toward the vacuum port. The vacuum line was attached to the vacuum port on the manifold base.
5. The tissue lysates were transferred to the wells of the Binding Plate. Vacuum was applied until all of the lysate has passed through the Binding Plate.
6. 1ml of Wizard SV wash solution (verify that ethanol has been added to the wash solution) was added to each well of the Binding Plate.
7. Vacuum was applied until the wash solution passed through the binding plate.
8. Steps 6 and 7 were repeated two more times for a total of 3 washes with the wash solution.
9. After the wells had emptied, vacuum applied until the seal broke.
10. The binding plate was removed from the manifold base and blotted by gently tapping onto a clean paper towel to remove residual ethanol.
11. The plate was placed in an incubator at 65°C for 15 minutes.
12. The 96-well deep well plate was placed in the manifold bed and the vacuum manifold collar positioned on top. The plate was orientated with the numerical column headers toward the vacuum port.
13. The binding plate, manifold collar 96-well deep well plate and manifold bed were set up as in Figure B.3.C.
14. 250μl of room temperature nuclease-free water was added to each well of the binding plate.
15. Vacuum was applied until the nuclease-free water passed through the binding plate.
16. Steps 14 and 15 were repeated for a second elution.
15. The apparatus was dismantled and the genomic DNA plates covered and stored at –20°C.

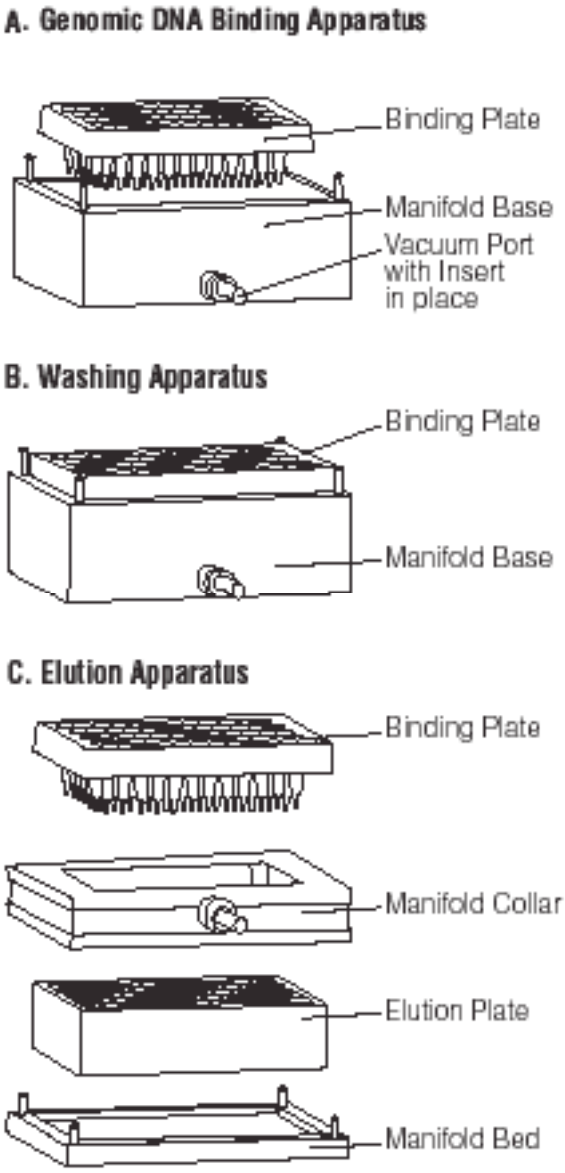


Figure B.1. The Vac-Man 96 Vacuum Manifold. Panels A, B and C show the manifold and plate combinations necessary to accomplish genomic DNA binding, washing and elution, respectively, for manual genomic DNA purification.

B.2. PCR amplification

Table B.2
Composition of the PCR reaction

	Volume per sample	Master mix for 96 samples
ddH_2O	17.65µl	1765 µl
Buffer (10xNH4)	2.5µl	250 µl
MgCl2 (50µM)	1.8µl	180µl
Primer 1 (10µM)	0.25µl	25µl
Primer 2 (10µM)	0.25µl	25µl
dNTPs (10µM)	0.50µl	50µl
Taq polymerase	0.055µl	5.5µl
Total Volume	**23µl**	**2300.5µl**

I used 1-3µl of template DNA from elution 2, and 23.5µl of 'master mix' (table B.2) in the PCR for each of the two genes.

Thermocycler settings:
1. Incubate at 94°C for 2 minutes
2. Incubate at 94°C for 30 seconds
3. Incubate at 48°C for 30 seconds
4. Incubate at 72°C for 1 minute
5. Cycle to step 2 for 39 more times
6. Incubate at 72°C for 7 minutes
7. Incubate at 4°C forever

3µl of the PCR products (loaded using 3µl of dye) were separated out electrophoretically on a 150ml 1M TAE agarose (3g) gel containing 3µl ethidium bromide, at 150V for 20 minutes, and then detected under UV light.

B.3. PCR cleanup

1. 50µl of binding buffer was added to the PCR product (making a total of approximately 75µl in each well).
2. The PCR product/binding buffer solution was then transferred onto the Multiscreen PCRµ96 filter plate (Millipore Corporation, Cat # MAFBNOBIO).
3. The vacuum system was set up as in Figure B.3.C and vacuum was applied.
4. 100µl of 80% ethanol wash was added to each well on the filter plate and drawn through with the vacuum.
5. Step 4 was repeated.
6. The filter plate was dried for 30 minutes at 37°C.
7. 50µl of ddH_2O was used to resuspend the DNA, this was drawn through the filter by centrifuge (4000rpm for 20 seconds).
8. The filter plate was washed with ddH_2O for reuse.

Following the cleanup products were checked again by electrophoresis as above.

B.4. DNA sequencing

Table B.3
Composition of the sequencing PCR reaction

	Volume per sample	Master mix for 96 samples
ddH_2O	6.5µl	650µl
2.5x Buffer	2.4µl	240µl
Primer (10µM)	0.32µl	32µl
Big Dye	0.6µl	60µl
Total Volume	**9.8µl**	**982µl**

I used 3µl of template DNA from the PCR cleanup and 7µl of 'master mix' (table B.3) for sequencing PCR for each of the three genes.
Thermocycler settings:
1. Incubate at 94°C for 2 minutes
2. Incubate at 96°C for 15 seconds
3. Incubate at 50°C for 15 seconds
4. Incubate at 60°C for 4 minutes
5. Cycle to step 2 for 35 more times
6. Incubate at 4°C forever

B.5. Sequencing cleanup

Table B.4
Composition of plate sequencing cleanup solution

	Volume per sample	Master mix for 96 samples
100% ethanol	50µl	5000µl
Na Acetate 3M	2µl	200µl
Total Volume	**52µl**	**5200µl**

1. 52µl of solution (table B.4) was added to each well.
2. This was incubated at room temperature for 15 minutes.
3. The plate was then centrifuged at 4000rpm for 30 minutes.
4. The liquid was poured off the plate.
5. 100µl of 70% ethanol was added to each well.
6. The plate was centrifuged at 4000rpm for 5 minutes.
7. The ethanol was poured off and the plate was spin dried onto paper to remove any remaining liquid.
8. The plate was dried at room temperature for 30 minutes and was now ready for sequencing on a gel.