PSWC   P. S. Ward Collection, University of California at Davis, CA, USA

## CO1 methods

Specimens were preserved in 95% ethanol in Madagascar and upon return to California were loaded into ScrewTop TrakMates® boxes (Matrix Technologies) and shipped to the University of Guelph. There, DNA was extracted from tissues rich in mitochondria (e.g. legs), employing primers with high universality, and amplifying a PCR product approximately 600 bp in length. Total genomic DNA extracts were prepared from small pieces ($\leq$1 mm) of tissue using the NucleoSpin® 96 Tissue kit (Macherey-Nagel Duren, Germany), following the manufacturer's protocols. Extracts were resuspended in 30 µl of dH$_2$O, and a 650base-pair (bp) region near the 5′ terminus of the CO1 gene was amplified following standard protocol [9]. Briefly, full length sequences were amplified using primers (LF1-ATTCAACCAATCATAAAGATATTGG and LR1-TGATTTTTTGGACATCCAGAAGTTTA [10]). In cases where a 650 bp product was not successfully generated, internal primer pairs (LF1–ANTMR1-(see Table 1)) and (MLF1 – GCTTTCCC-ACGAATAAATAATA [11] – LR) were employed to generate shorter overlapping sequences that allowed the creation of a composite sequence (contig). PCR reactions were carried out in 96 well plates in 12.5 µl reaction volumes containing: 2.5 mM MgCl$_2$, 5 pmol of each primer, 20 µM dNTPs, 10 mM Tris HCl (pH 8.3), 50 mM KCl, 10–20 ng (1–2 µl) of genomic DNA, and 1 unit of TaqDNA polymerase (Platinum® Taq DNA Polymerase - Invitrogen) using a thermocycling profile of one cycle of 2 min at 94°C, five cycles of 40 sec at 94°C, 40 sec at 45°C, and 1 min at 72°C, followed by 36 cycles of 40 sec at 94°C, 40 sec at 51°C, and 1 min at 72°C, with a final step of 5 min at 72°C. Products were visualized on a 2% agarose E-Gel® 96-well system (Invitrogen) and samples containing clean single bands were bidirectionally sequenced using BigDye v3.1 on an ABI 3730xl DNA Analyzer (Applied Biosystems). Contigs were made using Sequencher v4.0.5 (Gene Codes) and were subsequently aligned by eye in Bioedit [12]. Sequence divergences were calculated using the K2P distance model [13] and a NJ tree of distances [14] was created to provide a graphic representation of the patterning among-species divergences using MEGA3 [15], and

BOLD [16]. Sequence neutrality [Tajima's D - 17] and rates of substitution were calculated with DnaSP v.3 [18]. Sequences and other specimen information are available in the project file "Revision of Malagasy *Anochetus* and *Odontomachus*" in the Published Projects section of the Barcode of Life website (www.barcodinglife.org) with complete collection information for each specimen deposited at www.antweb.org. All sequences from the barcode region have been deposited in Genebank (CO1: EF610629: EF611041, EF999925-EF999945).

A composite representation of variation within the CO1 DNA barcode for each of the eight species revised here is presented in Figures 15 and 16. We used the online program Fingerprint [19 - http://evol.mcmaster.ca/fingerprint] to illustrate the heterogeneity at a specific site within the barcode region as a percentage of the vertical line drawn to represent each base pair.

Diagnostic base pairs (or combination of base pairs) for each species within the Malagasy region are presented. This more cladistic interpretation of the DNA barcode data is very sensitive to the number of specimens analyzed – and the fewer specimens incorporated, the greater the likelihood that a rare haplotype is not reflected in the data. We present this method of analysis not to that our coverage of each species is sufficient to reflect the variation within a species, but rather to demonstrate that such an analysis is possible within a group of taxa or region, when there is good representation of the variability within a species. The nucleotide position given refers to the barcode region, and can be compared to their full mitochondrial position by adding 48 (as aligned to the *Bos taurus* complete mitochondrial genome sequence Genbank ref AY676873). The standard IUPAC ambiguity codes are used to denote intra-specific variation.

Complementary genetic analyses. In addition to the CO1 barcode, for some specimens we amplified portions of the rRNA gene regions: 18S, 28S (D2) and ITS1. Within the variable D2 region of 28S, the forward primer corresponds to positions 3549–3568 in *Drosophila melanogaster* reference sequence (Genbank M21017). Within the 18S sequence, the forward primer corresponds to positions 375–406 in *Drosophila melanogaster* reference sequence while the ITS1 sequence was generated using primers where the forward primer corresponds to positions 1822–

**Table 1.** Primers used to generate sequences and molecular tests.

| Primer Name | Primer sequence (5′-3′) | Amplicon region | Primer source | Used for sequencing (Y/N) |
|---|---|---|---|---|
| LepF1 | ATTCAACCAATCATAAAGATATTGG | CO1 | [56] | Y |
| LepR1 | TAAACTTCTGGATGTCCAAAAAATCA | CO1 | [56] | Y |
| MLepF1 | GCTTTCCCACGAATAAATAATA | CO1 | [57] | Y |
| MLepR1 | CCTGTTCCAGCTCCATTTT | CO1 | [58] | Y |
| C_ANTMR1D-RonIIdeg_R | GGRGGRTARAYAGTTCATCCWGTWCC | CO1 | [Modified from 59] | N |
| C_ANTMR1D-AMR1deg_R | CAWCCWGTWCCKRMNCCWKCAT | CO1 | [Modified from 60] | N |
| CAS18Fs1 | TACACACCGCCCGTCGCTACTA | ITS1 | [61] | Y |
| CAS5p8s1Bd | ATGTGCGTTCRAAATGTCGATGTTCA | ITS1 | [Modified from 61] | Y |
| D2B | GTCGGGTTGCTTGAGAGTGC | 28S | [62] | Y |
| D3Ar | TCCGTGTTTCAAGACGGGTC | 28S | [62] | Y |
| 18H3 | AGGGTCGATTCCGGAGAGGGAGCCTGAGAA | 18S | [63] | Y |
| 18SWR | CTTGGCAAATGCTTTCGC | 18S | [63] | Y |
| *wsp* 81F | TGGTCCAATAAGTGATGAAGAAAC | *Wolbachia* surface protein | [20] | Y |
| *wsp* 691R | AAAAATTAAACGCTACTCCA | *Wolbachia* surface protein | [20] | Y |