A mitochondrial genome phylogeny of termites (Blattodea: Termitoidae): Robust support for interfamilial relationships and molecular synapomorphies define major clades.

To be submitted as a regular manuscript to *Molecular Phylogenetics & Evolution*

Stephen L. Cameron✉[1], Nathan Lo[2], Thomas Bourguignon[3], Gavin J. Svenson[4] and Theodore A. Evans[5]

[1] Earth, Environment & Biological Sciences School, Science & Engineering Faculty, Queensland University of Technology, GPO Box 2434, Brisbane, QLD 4001, AUSTRALIA.

[2] School of Biological Sciences, University of Sydney, Sydney, NSW, 2006, AUSTRALIA.

[3] Graduate School of Environmental Science, Hokkaido University, Sapporo, Hokkaido, 060-0810, JAPAN.

[4] Department of Invertebrate Zoology, Cleveland Museum of Natural History, 1 Wade Oval Drive, University Circle, Cleveland, OH, 44106, UNITED STATES of AMERICA.

[5] Department of Biological Sciences, National University of Singapore, 117543, SINGAPORE.

✉ Discipline of Biogeosciences, Faculty of Science & Technology, Queensland University of Technology, GPO Box 2434, Brisbane, QLD 4001, AUSTRALIA.

Phone: (+617) 3138 2869; FAX: (+617) 3138 2330; Email: sl.cameron@qut.edu.au

Running Title: Mitochondrial phylogeny of termites

Abstract

Despite their ecological significance as decomposers and their evolutionary significance as the most speciose eusocial insect group outside the Hymenoptera, termite (Blattodea: Termitoidae or Isoptera) evolutionary relationships have yet to be well resolved. Previous morphological and molecular analyses strongly conflict at the family level and are marked by poor support for backbone nodes. A mitochondrial (mt) genome phylogeny of termites was produced to test relationships between the recognised termite families, improve nodal support and test the phylogenetic utility of rare genomic changes found in the termite mt genome. Complete mt genomes were sequenced for 7 of the 9 extant termite families with additional representatives of each of the two most speciose families Rhinotermitidae (3 of 7 subfamilies) and Termitidae (3 of 8 subfamilies). The mt genome of the well supported sister-group of termites, the subsocial cockroach *Cryptocercus*, was also sequenced. A highly supported tree of termite relationships was produced by all analytical methods and data treatment approaches, however the relationship of the termites+*Cryptocercus* clade to other cockroach lineages was highly affected by the strong nucleotide compositional bias found in termites relative to other dictyopterans. The phylogeny supports previously proposed suprafamilial termite lineages, the Euisoptera and Neoisoptera, a later derived Kalotermitidae as sister group of the Neoisoptera and a monophyletic clade of dampwood (Stolotermitidae, Archotermopsidae) and harvester termites (Hodotermitidae). In contrast to previous termite phylogenetic studies, nodal supports were very high for family-level relationships within termites. Two rare genomic changes in the mt genome control region were found to be molecular synapomorphies for major clades. An elongated stem-loop structure defined the clade Polyphagidae + (*Cryptocercus* + termites), and a further series of compensatory base changes in this stem-loop is synapomorphic for the Neoisoptera. The complicated repeat structures first identified in *Reticulitermes*, composed of short (A-type) and long (B-type repeats) defines the clade Heterotermitinae + Termitidae, while the secondary loss of A-type repeats is synapomorphic for the non-macrotermitine Termitidae.

Introduction

Termites are one of the most significant insect groups on the planet due to their status as destructive structural pests of human buildings, their role as peak ecosystem engineers, and as the most speciose group of eusocial insects outside of the Hymenoptera. While only a small percentage of the almost 3000 described termite species consume structural or furniture timbers, this is perhaps the role which most people associate with the group as no other animal group directly attacks human dwellings in this way. Far more significant is their ecological role in recycling ligno-cellulose (Bignell & Eggleton 2000; Yamada et al. 2005) which they carry out via the combination endogenous and symbiotic cellulases and associated enzymes (Watanabe et al., 1998; Lo et al. 2011; Brune & Ohkuma 2011). In addition to being major consumers of wood (Cornwall et al. 2009), including dead timber, dead-limbs of living trees and living trees, termites also include grass harvesters which account for a significant portion of primary production in savanna ecosystems (Eggleton & Tayasu, 2001) and

wholly subterranean soil/humus feeders (Donovan et al., 2001). Different termite families, subfamilies and genera are dietary specialists and dietary type has a major influence on nesting and foraging behaviours (Abe, 1987). Termites are also, along with ants, significant actors in soil structuring allowing improved water infiltration and thus enhanced plant productivity in dry climates, where they replace earthworms in this role (Evans et al. 2011). Finally, termites are the second largest group of social insects, after the social Hymenoptera, and comparisons of how eusociality has evolved in the two groups have greatly advanced our knowledge of sociality in general (Fischman et al. 2011). Putting an evolutionary perspective on the wide variety of diet, ecology and colony systems within termites has, however, long been constrained by confusion about their phylogenetic relationships.

There has recently been a flowering of interest in termite phylogenetics (see review by Lo & Eggleton, 2011), however a consensus on the number of termite families and their relationships has yet to emerge despite nine studies addressing this issue specifically in a little over a decade. The lack of consensus has involved several issues, yet two among the 'lower termites' (= non-Termitidae) are persistent: (1) the order of family divergences after Mastotermitidae (in particular the Hodotermitidae, Stolotermitidae, Archotermopsidae and Kalotermitidae), and (2) in the degree of paraphyly of the Rhinotermitidae. The earliest phylogenetic studies of termites (e.g. Ahmad, 1950, Emerson & Krishna, 1975) are pre-cladistic treatments which refer to families evolving from within extant genera, do not report trees of family relationships and instead emphasize character-by-character narratives of evolution which do not objectively weigh evidence (e.g. head and wing pilosity link *Mastotermes* with *Hodotermes* but the fused first and second marginal teeth of the left mandible and position of the median vein link *Mastotermes* with *Kalotermes*), and so are risky to interpret posthoc in a tree-like evolutionary pattern. The first explicit termite phylogeny was published by Kambhampati et al. (1996) based on a single gene (*rrnL* = 16S) for 10 species, however there was poor nodal support for interfamilial relationships other than the monophyly of Euisoptera (all termites excluding Mastotermitidae). The year 2000 saw a burst of studies: Kambhampati & Eggleton (2000), 1 gene x 20 species; Thompson et al. (2000), 2 genes x 10 species; and Donovan et al. (2000) based on 196 morphological characters for 49 species. These studies differed in key respects (Figure 1) including the monophyly of families (several are not monophyletic in Donovan et al. 2000 but untestable in the other studies due to lack of taxa) and in interfamilial relationships. A meta-analysis by Eggleton (2001) produced a consensus tree of termite relationships, a perfect Hennigian ladder (Figure 1e). Subsequent major molecular studies greatly increased both the number of genes and the number of species analysed. Inward et al. (2007a), a combined analysis of 3 genes (2000 bp) plus morphology for 250 species, and Legendre et al. (2008), 7 genes (7000 bp) for 40 species, radically increased the available phylogenetic data for termites. However, the two studies differed significantly in four areas (Fig. 1f, g): (1) paraphyly vs monophyly of the stolotermitid-hodotermitid-archotermopsids (henceforth termed the SHA clade), (2) the degree of paraphyly within the Rhinotermitidae, forming either four or two groups, (3) kalotermitids as

either a late or early branching family, and (4) serritermitids as either part of Rhinotermitidae or sister to Rhinotermitidae+Termitidae (Inward et al. 2007a vs Legendre et al. 2008 respectively for each pair). The morphological phylogeny of termites received a thorough review by Engel et al. (2009), 108 morphological characters coded for 76 species including 38 fossil taxa, and resulted in a revised classification scheme including the elevation of 4 new families (3 extant). The most surprising result was the monophyly of Rhinotermitidae, a family which had been consistently shown to be paraphyletic with respect to Termitidae by all previous molecular and the single morphological analyses. Nodal support was relatively weak in Engel et al.'s (2009) tree with few clades supported by unambiguous synapomorphies. This is highlighted by two subsequent studies which both combined the Engel et al. (2009) morphology matrix with different sets of molecular data. Ware et al. (2010) combined 6 genes, largely derived from the Legendre et al. (2008) study, with the morphology data matrix, whereas Lo & Eggleton (2011) combined it with a single gene molecular data set. The topologies of the three analyses each differ (Fig. 1 h – j), largely in two areas long contentious: (1) the resolution of the early diverging families (the Hodotermitidae, Stolotermitidae, Archotermopsidae and Kalotermitidae), and (2) in the degree of paraphyly of the Rhinotermitidae.

Taken together, the nine comprehensive phylogenetic studies of termites published to date display some consistent relationships, the Mastotermitidae are the sister of all other families, a clade termed the Euisoptera, Rhinotermitidae +Termitidae are the most derived group (= the Neoisoptera) and of the termitid subfamilies, the fungus farming Macrotermitinae are generally the sister of the remaining subfamilies (exceptions being subfamilies which have by some authors been considered part of Macrotermitinae). The remaining interfamilial relationships are unresolved, as is the question of the monophyly and relationships between subfamilies within more speciose groups such as the rhinotermitids and termitids. A complicating factor in weighing these diverse phylogenetic hypotheses is that nodal support values for the majority of trees are low and difficult to compare directly due to differing methods of tree reconstruction. For the majority of studies, most nodes are not statistically significant (see Fig. 1). The exceptions are Thompson et al. (2000) and Legendre et al. (2008). The bootstrap support values reported by Legendre et al. (2008) should however, be treated with caution as they were generated from implied alignments produced by the program POY (Wheeler et al. 2006) and so are constrained which artificially inflates bootstrap support values (Yoshizawa, 2010). Poor nodal supports and conflicting topologies produced by studies that vary by small degrees e.g. different combinations of exemplar genera for the same families with many in common across multiple studies, suggests that current molecular datasets based on a limited number of genes needs to be greatly augmented if a robust phylogeny of the termites is to emerge. Clearly, a new approach is needed to resolve these persistent issues in termite phylogeny.

Analyses based on whole mitochondrial (mt) genomes are a practical and efficient approach to applying phylogenomic scale data to difficult phylogenetic questions, such as the evolution of termites. Coding for 37

genes – 13 protein-coding (PCGs), 2 ribosomal RNAs (rRNAs) and 22 transfer RNAs – the typical metazoan mt genome is a circular molecule approximately 15,000 bp in size and readily amplified by PCR using a combination of universally conserved and purpose designed primers. Mt genomes have been repeatedly used to resolve deep-intraordinal relationships within insects including within Phasmatodea (Komoto et al. 2011), Orthoptera (Fenn et al. 2008), Hemiptera (Hua et al., 2008; 2009), Hymenoptera (Cameron et al. 2008; Dowton et al. 2009a), Coleoptera (Song et al. 2009; Timmermans et al., 2010), Diptera (Cameron et al. 2007a) and Lepidoptera (Kim et al. 2011). Mt genome phylogenies have been shown to be robust over broad time scales and to provide sufficient signal for resolution even of short internodes (Cameron et al. 2007a). Strategies for dealing with among site rate heterogeneity and non-stationarity have been developed (Cameron et al. 2009; Sheffield et al. 2009; Song et al., 2010), extending the range of evolutionary conditions over which this data type can be applied.

Accordingly, we set out to develop a more robust molecular dataset for inferring the phylogenetic relationships within termites using whole mt genomes. Our aim was to sequence mt genomes for representatives of all extant taxa which have been proposed at the family rank following Engel et al. (2009) as this classification recognizes the largest number of families. Given that there is strong support from previous analyses for the monophyly of almost all of the groups accorded family status by Engel et al. (2009) (except Rhinotermitidae and Archotermposidae), an exemplar approach to phylogenetic reconstruction is appropriate (*sensu* Yeates 1995). All 13 of the genera included in the present study have been included in previous, more comprehensive phylogenies of termites, e.g. 11 of 13 are included in Legendre et al. (2008) and Engel et al. (2009), and 12 of 13 are included in Inward et al. (2007). The family assignment of each genus used here match those used in previous analyses with the exception of representatives of the families Stolotermitidae and Archotermopsidae which was recently proposed by Engel et al. (2009) and were classified as the Termopsidae in earlier studies. This overlap of representative genera between the present and previous analyses allows for direct comparison of family-level relationships between different studies. DNA or tissue was obtained for representatives of all families except the Stylotermitidae, and whole genomes sequenced for all of those except the Serritermitidae from which we were unable to amplify long PCRs. Whole mt genomes were sequenced for 13 species consisting of 12 termite species plus their well supported sister-group, the subsocial roach *Cryptocercus* (a group recently termed the Xylophagodea, Engel, 2011). Phylogenetic relationships were thus inferred between 7 of 9 extant termite families, 3 of 7 rhinotermitid subfamilies and 3 of 8 termitid subfamilies and potential molecular synapomorphies identified for several major termite clades.

Materials & Methods

Mitochondrial genome sequencing

Taxonomic classification (following Engel et al. 2009) and collection details for each study species are given in Table 1. All specimens were preserved in 100% ethanol and stored at – 20°C until DNA extraction. Australian samples were identified by SLC & TAE using Hill (1942) and the termite subcollection of the Australian National Insect Collection, Canberra (ANIC – CSIRO); non-Australian samples were identified by the collaborators who provided the samples. Whole genomic DNA was extracted from the head and thorax of workers with the DNeasy Tissue kit (QIAGEN); abdomen and gut were excluded to avoid contamination from food sources and symbiotic microbes. Exploratory long PCRs were performed using primers conserved either across all insects or designed from consensus dictyopteran sequences (Table 2); the remaining portion of the mt genome was amplified with primers specifically designed for each sample (Supp. Table S1). As PCR success varied between samples, the exact amplification strategy followed also varied between species; for the full amplification strategy and primer sequences see Suppl. Table S1. Within each long PCR product the full, double stranded sequence was determined by primer walking (primers available from SLC upon request). Long PCRs were performed using Elongase (Invitrogen) with the following cycling conditions: 92°C for 2 min; 40 cycles of 92°C for 30 sec, 50°C for 30 sec, 68°C for 12 min; and a final run out step of 68° for 20 min. PCR amplifications were cleaned using ExoSAP-IT (GE Healthcare) prior to sequencing. Sequencing was performed using ABI BigDye ver3 dye terminator sequencing technology and run on an ABI 3770 or ABI 3740 capillary sequencer. Sequencing PCR conditions were 28 cycles of 94°C/10 sec, 50°C/5 sec, 60°C/4 min. Raw sequence files were edited and assembled into contigs in Sequencher ver. 4 or 5 (GeneCodes Corporation). Transfer RNA inference was conducted using tRNAscan-SE (Lowe & Eddy 1997) using invertebrate mitochondrial predictors and a cove score cut off of 1. *trnS*$^{(AGN)}$ was the only tRNA which was routinely not found by tRNAScan-SE; it was identified by eye, through reference to secondary structure models for this gene from other dicytopteran insects. Reading frames between tRNAs were found in Sequencher and identified using translated BLAST searches (blastx) (Altschul *et al.* 1997) as implemented by the NCBI website (http://www.ncbi.nlm.nih.gov/). Annotations of the ribosomal RNA genes were done by eye with reference to previously published insect mt rRNA gene secondary structures (c.f. Cameron & Whiting 2008). Structural features of the mt genome such as stem-loops and repeat regions were identified as part of the annotation process.

Phylogenetic Inference

Alignments were made of the mt genomes of the 13 species newly sequenced for this study plus those of five additional dictyopteran species taken from GenBank, consisting of an additional termite, *Reticulitermes santonensis*, three roaches, *Periplaneta fuliginosa*, *Blatella germanica* and *Eupolyphaga sinensis* and a mantis, *Tamolanica tamolana*. Mt genome of three additional polyneopteran orders, Mantophasmatodea (*Sclerophasma paresiense*), Phasmatodea (*Megacrania alpheus*) and Orthoptera (*Locusta migratoria*), were

used as outgroups. Alignments were made of each gene separately using Muscle (Edgar *et al.* 2004, implemented in MEGA5: Tamura *et al.* 2011). PCGs were aligned as DNA codons in MEGA5, whereas RNA genes were directly aligned as DNA. Individual gene alignments were concatenated in MacClade 4.06 (Maddison & Maddison, 2005). Nucleotide composition statistics, A+T% and nucleotide skew, were calculated in MEGA5. Nucleotide skew measures the relative proportions of A's to T's, (A-T)/(A+T), and C's to G's (C-G)/(C+G). Positive skew indicates an excess of A's or C's whereas negative skew indicates an excess of T's and C's. The statistic scales from 0 (equal proportions of each nucleotide) to 1 (complete absence of one nucleotide) (Perna & Kocher, 1995).

Two analytical approaches were used to infer phylogenetic trees, likelihood and Bayesian inference, to determine the affect of analytical method on topology and nodal support. Conflict between codon- and gene-based signals was assessed by partitioning either by codon or gene for each inference method. The effect of base compositional bias was examined using partitioning by codon, LogDet transforms (implemented in PAUP 4.0b10, Swofford, 2002), PHASE (Gowri-Shankar & Rattray, 2007) and outright removal of third codon positions as the most compositionally biased partition (see below). Partitioning by gene resulted in 16 total partitions, 13 PCGs, 2 rRNAs and a combined partition for the 22 tRNAs as each individual tRNA is too short (circa 65 bp) for accurate parameter determination if analysed separately. Partitioned by codon results in 5 (without third codons) or 6 (with third codons) total partitions, 1 for each included codon position, 1 each for the 2 rRNAs and 1 combined for the 22 tRNAs. Analyses were performed with the RaxML Black-Box webserver (http://phylobench.vital-it.ch/raxml-bb/index.php; Stamatakis et al., 2008) for likelihood and MrBayes ver 3.1.2 (Huelsenbeck & Ronquist 2001) for Bayesian analysis. All Bayesian analyses were run with unlinked partitions, appropriate models of molecular evolution selected for each partition and each dataset analysed using 2 independent runs, each of 4 chains (3 hot and 1 cold chain), for 3 million generations with sampling every 1000 generations; convergence was achieved by all analyses within 3 million generations as determined using Tracer ver. 1.4 (Rambaut & Drummond 2007). Completed Bayesian analyses were examined for asymptotic behavior of each parameter and of total tree likelihood; trees collected prior to this asymptotic point were treated as burn-in and discarded (generally the first 30-60,000 generations). Partition models were chosen using AIC as implemented in ModelTest (Posada & Crandall 1998). Bayesian run files are available for each analysis from SLC upon request.

Results

Genome Sequences

Complete mt genomes were sequenced for 12 termite and 1 cockroach species and have been submitted to GenBank (see Table 1 for accession numbers). Termite mt genomes sequenced in this study range in size from

15,483 (*Zootermopsis*) to 16,542 bp (*Drepanotermes*) i.e. similar in size to other insect mt genomes. There may be a slight tendency toward an increase in mt genome size within termites as all but 2 termite species, *Mastotermes* and *Zootermopsis*, have larger mt genomes than any of the roaches from which they are descended, and which range in size from 14,996 (*Periplaneta*) to 15,553 bp (*Eupolyphaga*). As in other mt genomes, the majority of genome size variability is due to variation in the size of the A+T rich region (= putative control region) and much of that variability is ascribable to large repeat units (covered further below). All species have the insect ancestral mt genome arrangement. Within Dictyoptera, there is limited length variability in the PCGs and a high degree of conservation of start- and stop-codons across homologous genes (Supplementary Table S2). The one exception is *nad1* in *Cryptocercus* where a single base deletion has caused a frame shift, removed the in-frame stop codon and resulted in a gene 20bp longer than that found in other dictyopterans. Similar frame shifts at this location have been found in beetles (Sheffield et al., 2008) associated with a conserved regulatory element, the binding site of the transcription termination peptide mtTERM (Cameron & Whiting, 2008), and it is possible that the additional 20 bp are not translated. The majority of PCGs utilize canonical start (M or I) and stop (TAA, TAG, TA, or T) codons. One exception is use of V as a start codon in *nad5* (*Porotermes*, *Zootermopsis*, *Microhodotermes*, *Heterotermes*, *Coptotermes*) and *atp8* (*Mastotermes*). In all six instances, the V is coded for by the triplet GTG which is one base removed from the M coding triplets (ATN) and may be "corrected" by post-transcriptional modification. As with the use of non-canonical start codons for *cox1*, the use of V is actually widespread within insects (e.g. Cameron et al. 2011) however it has yet to be examined by transcript mapping in the way *cox1* has been (e.g. Margam et al. 2011).

As previously noted for the mt genomes of *Reticulitermes* (Cameron & Whiting, 2007), the mt genomes of termites have a high degree of nucleotide compositional bias. Compared to other dictyopterans, there is a strongly reduced A+T% (an average 7% decrease), an increase in C-skew and a strong increase in T-skew in termites (Table 3; Supplementary Table S3). The reduced A+T bias in termites is strongest in the third codon position, which are approx. 14% lower than non-termite dictyopterans, reflecting a stronger background mutational pressure towards C's and G's at silent sites in termites than in non-termite dictyopterans. While the A+T% of *Cryptocercus* is not significantly reduced relative to other non-termite dictyopterans, the skew statistics are biased with C-skews intermediate between those of termites and non-termites and A-skews stronger than *Mastotermes*.

Mitochondrial Genome Phylogeny of Termites

Inference method and partitioning strategies had no effect on topology and limited effect on nodal support; exclusion of third codon positions however had a major effect on the topology in the non-termite dictyopterans (Figure 2; Supp. Fig. S1). Dictyoptera was monophyletic in all analyses. When third codon positions are

included Dictyoptera divides into two clades: termites plus *Cryptocercus* and the remaining roaches with Mantodea (*Tamolanica*) derived from within Blattodea as the sister-group of *Blatella* (Supp. Fig. S1). When third codon positions are excluded Mantodea represented by *Tamolanica* is the sister of the remaining dictyopterans and the roaches are strongly paraphyletic with *Eupolyphaga* sister to *Cryptocercus*+termites (Fig. 2). The difference between inclusion and exclusion of third codon positions is likely due to the high degree of nucleotide compositional heterogeneity between termites and roaches which is not modeled adequately by RaxMl or MrBayes for this dataset. Computational methods of correcting for base compositional bias (LogDet transforms, PHASE) resulted in the same tree topologies as those inferred by other methods when third codon positions were included or excluded, suggesting that they fail to correct for this bias. For this reason and due to a synapomorphic rare genomic change (see below) we consider the topology excluding third codon positions to be more accurate.

Cryptocercus+termites and the same set of relationships within the termites were found in all analyses including those including versus excluding third codon positions. As expected the Mastotermitidae are the sister of the remaining termites (= Euisoptera). There is strong support for the monophyly of a clade composed of the stolotermitids, hodotermitids and archotermopsids (=SHA clade), with Hodotermitidae plus Archotermopsidae. Kalotermitids are the sister of the Neoisoptera which is also strongly supported. Within the Neoisoptera, Rhinotermitidae is consistently paraphyletic with respect to Termitidae, with Rhinotermitinae (represented by *Schedorhinotermes*) sister to the remaining rhinotermitids plus termitids. A strongly supported sister grouping between *Heterotermes* and *Coptotermes* renders Heterotermitinae paraphyletic. Resolution within the Termitidae consistently supported macrotermitines as the sister of the remaining termitids and while there is support for a monophyletic Termitinae (i.e. *Drepanotermes +Macrognathotermes*) to the exclusion of *Nasutitermes* (Nasutitermitinae) from both inference methods, it is not significant in the maximum likelihood analyses.

Nodal support was stronger for the Bayesian than the maximum likelihood analyses and for codon-based partitions than for gene-based partitions (Fig. 2). Support for most of the relationships between the roach genera was not significant in all analyses except the Bayesian analysis of the codon-partitioned dataset. Nodal support within the termites was very strong across all analyses except for the sister paring of *Drepanotermes*+ *Macrognathotermes* which was significantly supported only in the Bayesian analyses.

Rare Genomic Changes as Potential Synapomorphies

A series of rare genomic changes (RGCs), complex molecular features shared between species, in the non-coding A+T rich (=putative control) region were identified and mapped onto the consensus phylogenetic tree to

determine if they represented molecular synapomorphies (Figure 2). RGC 1 is a major hairpin loop found in the A+T rich region which is greatly elongated in all termites, *Cryptocercus* and *Eupolyphaga*; it is probably the origin of replication for the mt genome (c.f. Saito et al. 2005). The stem ranges in size from 16 paired bases in *Zootermopsis* to 38 paired bases in *Mastotermes* and the loop from 11 (*Schedorhinotermes*) to 14 bp (*Cryptocercus*) in size. The stem bases are highly conserved and readily alignable (Supp. Fig S2). Furthermore, a series of base substitutions in the more conserved distal part of the stem are consistent with the monophyly of the Neoisoptera (RGC 1a: Figure 2). This hairpin loop is located an average of 305 bp from the 5' end of *rrnS*; the range is 225 (*Cryptocercus*) to 406 bp (*Drepanotermes*); *Reticulitermes* is an outlier at 672 bp from *rrnS* due to an additional repeat unit in this region (see below). Hairpin loops of this sort are a common feature of metazoan mt genomes however the single conserved stem-loop is much smaller in the remaining dictyopterans, 7-10 stem bases, 10-17 loop bases, and located much closer to *rrnS*, 31-98 (avg 67) bp from the 5' end. It is thus not likely to be homologous to the major hairpin loop.

RGC 2 is the presence of the complicated double repeat units first found in *Reticulitermes*, consisting of short (type-A, 186 bp) repeats adjacent to the *rnS* end of the AT-rich region and long (type-B, 552 bp) repeats, which contain the long hairpin loop structure discussed above, adjacent to the *trnI* end. This same complicated repeat structure is also found in *Coptotermes*, *Heterotermes* and *Macrotermes*. In each of these three genera the repeat units consist of one full A unit, one partial A, one partial B unit, followed by 2 full B units (A-Ap-Bp-B-B). *Reticulitermes* differs only in having 2 full A units. In addition, the partial A and partial B units overlap in *Reticulitermes*, *Coptotermes* and *Heterotermes* but not in *Macrotermes* where they are separated by 13 bp which don't match either repeat unit. Non-macrotermitine termitids have lost the A repeats but retain the B repeat units in a variety of forms Bp-B (*Nasutitermes*), Bp-B-B (*Drepanotermes*) or B-B (*Macrognathotermes*). These repeat structures are thus synapomorphic for at least the clade Heterotermitinae+Termitidae, and the secondary loss of the A-type repeats is a synapomorphy for the non-macrotermine Termitidae (RGC 2a). While A+T-rich region repeats are found in other dictyopterans included in this study they are clearly not homologous to either the A- or B-type repeats (discussed below).

These two features were the only structural changes to the mt genomes found in the termite genera examined which were shared between two or more species. There were no genome rearrangements, duplications or pseudogenes identified in any of the study species.

Discussion

Termite Phylogeny and Mitochondrial Genomics

Mitochondrial phylogenomics resolves relationships between termite families and is insensitive to variations in phylogenetic inference method and different approaches to partitioning the genome data. Perhaps more significantly, unlike the majority of previous phylogenetic studies of termite relationships, mt genome data also provides high nodal support irrespective of analytical method or partitioning strategy. These results suggest that mt genome data is of high utility in firmly resolving termite relationships and that additional genomes would likely resolve the questions not addressed in this study e.g. the placement of serritermitids and stylotermitids, the number, composition and relationships of rhinotermitid subfamilies and possibly relationships between termitid subfamilies despite the low nodal supports found here. In contrast, relationships within the non-termite Dictyoptera which were strongly affected by the inclusion vs exclusion of third codon positions. This is likely due to the marked variation in base composition between termites+*Cryptocercus* and the remaining Dictyoptera, 67% vs 74% AT over the whole coding region, but greatest at third codon positions, where members of the *Cryptocercus*+termites clade were 14-15% lower A+T% than the non-termite dictyopterans (Table 3). Accounting for base compositional bias is a challenge in molecular systematics (Jermiin et al. 2004) and is acute in mt phylogenomics (c.f. Phillips & Penny, 2003; Sheffield et al. 2009). Of the computational methods proposed to deal with base compositional bias, partitioning by codon, PHASE and LogDet transforms all failed to correct for it and resulted in the same topology as all other analyses including third codon positions (Supp. Fig. S1). This is consistent with previous studies (e.g. Sheffield et al. 2009; Cameron et al. 2009; Song et al. 2010) which have found that methods of correcting for compositional biases vary in their effectiveness between different datasets.

Our study identified 4 major lineages within the termites – the Mastotermitidae, the Stolotermitidae + Hodotermitidae + Archotermopsidae (SHA clade), the Kalotermitidae and the Neoisoptera (=Rhinotermitidae+Termitidae). All modern phylogenetic analyses of termite evolution identify the Mastotermitidae as the sister of the remaining extant termites (=Euisoptera), whose monophyly is consistent with morphological (e.g. gain of the basal suture and loss of the anal lobe in the wings), biological (e.g. loss of ootheca; Nalepa & Lenz, 2000) and now genomic data. The sister-grouping of Mastotermitidae + Euisoptera is also one of the few well-supported clades across all previous studies; only in Donovan et al. (2000) does this node lack significant support due to the analysis being rooted on Mastotermitidae.

The second group we identify, the SHA clade, was also found by Thompson et al. (2000) and Legendre et al. (2008). In the remaining studies, the SHA families form an evolutionary grade which is not interspersed with any other termite families (except Kambhampati & Eggleton, 2000; and Lo & Eggleton, 2010). Nodal support in these studies for the SHA families forming a grade is, however, never significant. These families have all previously been considered as a single family, Hodotermitidae (Ahmad, 1950), however diverse life-histories found in the group have supported their division into multiple families. The Hodotermitidae are found in the

tropics and subtropics, from southern Africa to India, where they live in dry lands in soil, in large (populations of tens to hundreds of thousands of individuals) separate piece colonies, harvest grass for food, and have two developmental pathways, one producing true (i.e. obligatorily sterile) workers. The Stolotermitidae and Archotermopsidae have a quite contrasting biology; both live in cool temperate zones (Lacey et al. 2010) where they live within rotting wood in wet forested areas in small colonies (populations of ten to hundreds of individuals), and have one developmental pathway, with pseudergates (i.e. facultative sterile juveniles). They were long considered a single family, the Termopsidae, however their disjoint distribution – stolotermitids occur in the southern hemisphere whereas archotermopsids are northern – combined with their frequent non-monophyly in phylogenetic analyses led Engel et al. (2009) to raise the former to family status, Stolotermitidae, and to propose the Archotermopsidae as a new name for the northern termopsids as the type genus for this family was extinct and not closely related to the extant genera. The Archotermopsidae, however, is non-monophyletic in all studies where its monophyly is testable, including the one in which it is proposed (Engel et al. 2009), and it may therefore be taxonomically more conservative to revert to a single family, Hodotermitidae. Addition of the key taxon *Hodotermopsis* to this dataset would greatly advance the issue as this genus frequently rendered Archotermopsidae non-monophyletic in previous molecular analyses.

The third major lineage which we identify are the drywood termites, the Kalotermidae. Although not tested in the current study, kalotermitids are monophyletic in almost all previous phylogenetic hypotheses (paraphyletic with respect to Neoisoptera in Donovan et al. 2000). There is far less consensus, however, amongst these studies as to the position of this family within the termite tree. Most support the derived position found in the present study as sister to Neoisoptera (Kambhapati et al. 1996; Donovan et al., 2000; Thompson et al. 2000; Inward et al. 2007a; Engel et al. 2009; Ware et al. 2010), but several suggest a very basal position within an SHA grade (Kambhapati & Eggleton, 2000; Lo & Eggleton, 2010) or as sister of the remaining Euisoptera (Legendre et al. 2008). Crucially, there is limited nodal support for a basal position of kalotermitids in those studies (except Legendre et al. 2008), whereas the derived position is supported by both morphological studies including one of the few unambiguous synapomorphies found in Engel et al. (2009), forewing CuA elongate with 6 or more branches.

The Neoisoptera has been universally supported by previous and the present phylogenetic analyses with the major differences between studies concerning the placement of taxa not included in the present study (Serritermitidae and Stylotermitidae). Rhinotermitidae has been non-monophyletic in all studies except those with limited sampling e.g. Kambhapati et al. (1996) and Thompson et al. (2000) both use just 2 genera which are closely related and not fully representative of the family. The most comprehensive studies of rhinotermitid relationships, which included 10 of the 13 extant genera, both found evidence for the same 4 rhinotermitid clades which are collectively paraphyletic with respect to the Termitidae (Lo et al. 2004; Inward et al. 2007a).

The present study includes representatives of just 2 of these 4 rhinotermitid clades, however the same pattern of paraphyly with respect to the termitids is recovered. In both Lo et al. (2004) and Inward et al. (2007a) the sister group of the termitids is the Heterotermitinae (including Coptotermitinae as discussed below), however the nodal support for this relationship is poor in both, 62% parsimony bootstrap and 0.87 posterior probability in the former and a Bremer support of 7 in the later. The present study provides strong nodal support for this sister grouping and also identifies a strong molecular synapomorphy for the clade in the form of the unique double repeat units (see below). Further we find strong support for the sister grouping between *Coptotermes* and *Heterotermes* to the exclusion of *Reticulitermes*, thus rending the Heterotermitinae (sensu Engel et al. 2009) paraphyletic with respect to the Coptotermitinae (sensu Engel et al. 2009). Heterotermitinae was non-monophyletic in every previous analysis where it was testable except Engel et al. (2009) who found it to be the sister-group of the Coptotermitinae (i.e. a monophyletic clade composed of these three genera). Sinking the monogeneric subfamily Coptotermitinae into Heterotermitinae would thus reflect the results of all molecular phylogenies and be consistent with Engel et al.'s (2009) morphological phylogeny.

The present study includes three of the eight recognised termitid subfamilies and the relationships found are consistent with previous studies. Macrotermitinae is the sister of the remaining termitids which is consistent with the majority of published phylogenies, exceptions include Legendre et al. (2008) who favoured Sphaerotermitinae or Ware et al. (2010) who favoured Foraminitermitinae. Our finding of a monophyletic Termitinae (*Drepanotermes* + *Macrognathotermes*), to the exclusion of a Nasutitermitinae (*Nasutitermes*) is at odds with most previous studies which suggest that nasutes were derived from within the termitines (Donovan et al., 2000; Ohkuma et al., 2004; Inward et al. 2007a) although it is consistent with Legendre et al. (2008). This is the one node within the termite portion of this study which varied depending on analysis method, being strongly supported by Bayesian analyses and while found in the most likely tree, it was not significantly supported. Many additional taxa are needed to fully resolve termitid relationships and given the variability in the subfamily limits suggested by previous studies, caution should be exercised in choosing representative taxa to investigate termitid evolution.

Rare Genomic Changes in Termite Evolution.

In addition to the sequence based phylogeny, the study of termite mt genomes also revealed two RGCs based on complicated secondary structures within the A+T-rich region. RGC synapomorphies inferred from mt genomes have typically taken the form of gene rearrangements (e.g. Thao et al. 2004; Dowton et al., 2009b; Sheffield et al. 2010; Cameron et al., 2011), although repeat numbers, tRNA secondary structure changes, changes in the genetic code and even base mutations at conserved positions have also been proposed (Murrell et al., 2003; Lavrov, 2011). Avise (1994) coined the term "idiosyncratic markers" and Rokas & Holland (2000) "rare

genomic changes" to describe molecular features deemed unlikely to have evolved multiple times within a given group of organisms and are thus expected to be free of the problems of homoplasy which affect sequence or morphological data. Homoplasy has, however, been found in many instances when these types of markers have been investigated in depth e.g. a gene rearrangement found in bees and grasshoppers is a clear convergence (Flook et al. 1995) as are anticodon mutations within lice (Cameron et al 2007b). The RGCs identified here represent an opportunity to investigate a poorly studied class of phylogenetic markers within termites.

Determining if the secondary structures found in a non-coding portion of the genome constitute molecular synapomorphies has an additional complication in that to be putatively homologous the compared structures must also be positionally and structurally homologous. RGC 1, the large hairpin loop is found in the same general position within the A+T rich region and has a high degree of sequence conservation in all species in which it is found, thus satisfying both criteria. This structure, when mapped onto the phylogenetic tree derived from sequence data, provides support for the most weakly supported node on this tree, the clade *Eupolyphaga* +(*Cryptocercus*+termites). The support for this clade is slightly less than significant in both Bayesian analyses (posterior probability of 0.85 in both), and while present in the most likely tree, there is very limited bootstrap support for this node in either RaxML analysis (29 and 28% in the 5- and 16- partition datasets respectively). Relationships within the roaches are currently disputed and particularly so on the question of the sister group of *Cryptocercus* +termites (see Inward et al. 2007b; Lo et al. 2007). Most molecular studies favour blattids (including tryonicids) as the sister group (Lo et al. 2000; Inward et al. 2007b; Pellens et al. 2007; Ware et al. 2008; Murienne, 2009). The most recent and largest study to date, however, proposes the clade Polyphagidae +Nocticolidae, although this relationship is only significant in the Bayesian analysis (Djernæs et al. 2012). Morphological analyses mostly support the clade Polyphagidae +Lamproblattidae as the sister of *Cryptocercus* +termites (Klass, 1995, 1997; Klass & Meier, 2006) with the exceptions being studies that either did not include termites (McKittrick & Mackerras, 1965) or used termites as outgroups (Grandcolas, 1996). RGC 1 is thus consistent with past morphological studies and the most recent molecular analyses and forms an independent line of support for a clade whose nodal support is insignificant in both morphological and sequence-based analyses. Determining the presence or absence of RGC 1 in representatives of the Nocticolidae and Lamproblattidae would further help to resolve this question of the sister group of *Cryptocercus*+termites and allow a more direct comparison with previous cockroach phylogenies.

The second synapomorphy, the double repeat units, is consistent with our phylogeny and with the majority of other published molecular phylogenies of termites, however one must posit the secondary loss of the A-type repeats in the non-macrotermitine termitids (RGC 2a). This secondary loss is not just of duplicated regions of DNA but replacement of the region between *rrnS* and the B-type repeats with other non-coding DNA, as this

region in the non-macrotermitines has only limited sequence conservation with the A-type repeats in *Macrotermes* or the heterotermitines. The presence/absence of the A- and B-type repeats could be a useful character to resolve the early branching patterns within the Termitidae by determining in what set of taxa A-type repeats were lost and if they were lost on multiple occasions. As with sequence-based phylogenetic analysis of termitid evolution outlined above, many representatives will need to be examined to overcome uncertainties about classification schemes within this family. Similarly, it would be interesting to determine if these repeats occur in other rhinotermitid groups or if they are confined to the Heterotermitinae and more derived groups. A+T-rich region repeat units are also found in *Neotermes*, *Porotermes* and *Tamolanica*, however in each case they are clearly not homologous to either the A- or B-type repeats. *Neotermes* possesses 3 full (57 bp) and 1 partial (49 bp) repeat units however none include the long stem-loop structure (RGC 1) found in all termites and which forms the 5' end of the B-type repeats. *Porotermes* possesses 2 full (541 bp) and 1 partial (96 bp) repeat units and while the full repeats do contain the RGC 1 similar to a B-type repeat, it is located at a different position within the repeat and the repeats themselves are located much closer to *rrnS* than is usual for B-type repeats (103 bp vs a minimum of 252 bp in *Coptotermes*). The mantid *Tamolanica* has a complicated set of 63 bp repeats which occur in three regions of the genome: between *trnM* and *nad2* (1 full unit), at the *rrnS* end of the AT-rich region (1 full, 1 partial), and at the *trnI* end of the AT-rich region (3 full, 1 partial). None of these repeat units in *Tamolanica* include the hairpin loop structure conserved in other dictyopterans and repeats are absent from other mantid mt genomes (Cameron *unpublished data*).

Conclusion

The mt genome of termites provides two types of phylogenetic data, a large quantity of sequence data and rare genomic changes (RGC) which can be mapped to nodes in the termite tree. As a source of sequence data, the mt genome is a useful advance over the molecular datasets which have previously been applied to termite systematics. In particular the recovery of significant nodal support for the proposed interfamily relationships is an advance over the majority of previous studies and suggests that the mt genome can be used to increase our support for deep nodes across the termite tree. Obtaining full mt genomes for additional key termite genera, particularly within the largest family Termitidae, is necessary to completely test their phylogenetic potential. Comparatively few nuclear markers, only 18S and 28S, have been used in termite molecular systematics and so the testing of novel nuclear markers separately and in combination with the mt genomes of the exemplar genera included is also a necessary step. The two RGCs identified in termites and roaches and their four character states are congruent with the sequence-derived trees suggesting that they may be useful for testing the relationships of other genera which have been proposed to be closely related to the clades they appear to define. In particular RGC2, the AT-region repeat units, could be useful in understanding the early splitting relationships within Termitidae however many additional genera need to be tested to determine the evolution of this feature.

Acknowledgements

The authors would like to thank David Yeates (Australian National Insect Collection) for advice and support of this research, Cate Smith (CSIRO Plant Industry) for assistance with sequencing, and Jessica Ware (Rutgers University) and Michael Whiting (Brigham Young University) for donating specimens. This study was supported US National Science Foundation grants DEB0444972 (SLC), DEB1050569 (GJS), Australian Research Council grant DP1097265 (NL), the Taxonomy Research & Information Network (TRIN) and the Atlas of Living Australia. SLC, NL and TAE would like to acknowledge the more than 80 years of continuous research into termites at the CSIRO, which comprised the lion's share of termite research in Australia, and mourn the end of this research at CSIRO in 2011.

References

Abe, T., 1987. Evolution of life types in termites. In: Connell, J.H. & Hidaka, J. (Eds.), Evolution and coadaptation in biotic communities. University of Tokyo Press, Tokyo, pp 125-148.

Ahmad, M., 1950. The phylogeny of termite genera based on imago-worker mandibles. B. Am. Mus. Nat. Hist. 95, 37-86.

Altschul, S.F., Madden, T.L. Schäffer, A.A. Zhang, J. Zhang, Z. Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389-3402.

Avise, J.C., 1994. Molecular Markers, Natural History and Evolution. Chapman & Hall, New York, NY.

Bignell, D.E., Eggleton, P., 2000. Termites in ecosystems. In: Abe, T., Bignell, D.E., Higashi, M. (Eds.), Termites: Evolution, Sociality, Symbiosis, Ecology. Kluwer Academic Publishers, Dordrecht, pp. 363-387.

Brune, A., Ohkuma, M., 2011. Role of the Termite Gut Microbiota in Symbiotic Digestion. In: Bignell, D.E., Roisin, Y., Lo, N. (Eds.), Biology of Termites: A Modern Synthesis. Springer, Dordrecht, pp. 439-475.

Cameron, S.L., Whiting, M.F., 2007. Mitochondrial genomic comparisons of the subterranean termites from the Genus *Reticulitermes* (Insecta: Isoptera: Rhinotermitidae). Genome 50, 188-202.

Cameron, S.L., Whiting, M.F., 2008. The complete mitochondrial genome of the tobacco hornworm, *Manduca sexta*, (Insecta: Lepidoptera: Sphingidae), and an examination of mitochondrial gene variability within butterflies and moths. Gene 408, 112-123.

Cameron, S.L., Lambkin, C.L., Barker, S.C., Whiting, M.F., 2007a. Utility of mitochondrial genomes as phylogenetic markers for insect intraordinal relationships – A case study from flies (Diptera). Syst. Entomol. 32, 40-59.

Cameron, S.L., Johnson, K.P., Whiting, M.F., 2007b. The mitochondrial genome of the screamer louse *Bothriometopus* (Phthiraptera: Ischnocera): Effects of extensive gene rearrangements on the evolution of the genome as a whole. J. Mol. Evol. 65, 589-604.

Cameron, S.L., Dowton, M., Castro, L.R., Ruberu, K., Whiting, M.F., Austin, A.D., Diement, K., Stevens, J., 2008. The sequence of the mitochondrial genomes of two vespid wasps reveals a number of derived tRNA gene rearrangements. Genome 51, 800-808.

Cameron, S.L., Sullivan, J., Song, H., Miller, K.B., Whiting, M.F., 2009. A mitochondrial genome phylogeny of the Neuropterida (lace-wings, alderflies and snakeflies) and their relationship to the other holometabolous insect orders. Zool. Scr. 38, 575-590.

Cameron, S.L., Yoshizawa, K., Mizukoshi, A., Whiting, M.F., Johnson, K.P., 2011. Mitochondrial genome deletions and mini-circles are common in lice (Insecta: Phthiraptera). BMC Genomics 12, 394.

Cornwall, W.K., Corneliss, J.H.C., Allison, S.D., Bauhus, J., Eggleton, P., Peston, C.M., Scarff, F., Weedon, J.T., Wirth, C., Zanne, A.E., 2009. Plant traits and wood fates across the globe: rotted, burned or consumed? Glob. Change Biol. 15, 2431-2449.

Crozier, R.H., Crozier, Y.C. 1993. The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organisation. Genetics 133, 97-117.

Djernæs, M., Klass, K.-D., Picker, M.D., Damgaard, J. 2012. Phylogeny of cockroaches (Insecta: Dictyoptera, Blattodea), with placement of aberrant taxa and exploration of out-group sampling. Syst. Entomol. 37, 65-83.

Donovan, S.E., Jones, D.T., Sands, W.A., Eggleton, P., 2000. Morphological phylogenetics of termites (Isoptera). Biol. J. Linn. Soc. 70, 467-513.

Donovan, S.E., Eggleton, P., Bignell, D.E., 2001. Gut content analysis and a new feeding group analysis of termites. Ecol. Entomol. 26, 356-366.

Dowton, M., Cameron, S.L, Austin, A.D., Whiting, M.F., 2009a. Phylogenetic approaches for the analysis of mitochondrial genome sequence data in the Hymenoptera – a lineage with both rapidly and slowly evolving mitochondrial genomes. Mol. Phylogenet. Evol. 52, 512-519.

Dowton, M., Cameron, S.L., Dowavic, J.I., Austin, A.D., Whiting, M.F., 2009b. Characterisation of 67 mitochondrial gene rearrangements in the Hymenoptera reveal underlying trends in mitochondrial genome evolution. Mol. Biol. Evol. 26, 1607-1617.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792-1797.

Eggleton, P., 2001. Termites and trees: a review of recent advances in termite phylogenetics. Insect. Soc. 48, 187-193.

Eggleton, P., Tayasu, I., 2001. Feeding groups, lifestyles and the global ecology of termites. Ecol. Res. 16, 941-960.

Emerson, A.E., Krishna, K., 1975. The termite family Serritermitidae (Isoptera). Am. Mus. Novit. 2570, 1-31.

Engel, M.S., 2011. Family-group names for termites (Isoptera), redux. Zookeys 148, 171-184.

Engel, M.S., Grimaldi, D.A., Krishna, K., 2009. Termites (Isoptera): their phylogeny, classification and rise to ecological dominance. Am. Mus. Novit. 3650, 1-27.

Evans, T.E. Dawes, T.Z., Ward, P.R., Lo, N., 2011. Ants and termites increase crop yield in a dry climate. Nat. Commun. 2, 262.

Fenn, J.D., Song, H., Cameron, S.L., Whiting, M.F., 2008. A mitochondrial genome phylogeny of Orthoptera (Insecta) and approaches to maximizing phylogenetic signal found within mitochondrial genome data. Mol. Phylogenet. Evol. 49: 59-68.

Fischman, B.J., Woodard, S.H. & Robinson, G.E., 2011. Molecular evolutionary analyses of insect societies. P. Natl Acad. Sci. USA. 108, 10847-10854.

Flook, P.K., Rowell, C.H.F., Gellissen, G., 1995. Homoplastic rearrangements of insect mitochondrial tRNA genes. Naturwissenschaften 82, 336-337.

Grandcolas, P., 1996. The phylogeny of cockroach families: a cladistic appraisal of morpho-anatomical data. Can. J. Zoolog. 74, 508-527.

Gowri-Shankar, V., Rattray, M., 2007. A reversible jump method for Bayesian phylogenetic inference with a nonhomologous substitution model. Mol. Biol. Evol. 24, 1286-1299.

Hill, G.F., 1942. Termites (Isoptera) from the Australian Region. Council for Scientific & Industrial Research, Melbourne.

Hua, J., Li, M., Dong, P., Cui, Y., Xie, Q., Bu, W., 2008. Comparative and phylogenomic studies on the mitochondrial genomes of Pentatomorpha (Insecta: Hemiptera: Heteroptera). BMC Genomics 9, 610.

Hua, J., Li, M., Dong, P., Cui, Y., Xie, Q., Bu, W. 2009. Phylogenetic analysis of the true water bugs (Insecta: Hemiptera: Heteroptera: Neopomorpha): evidence from mitochondrial genomes. BMC Evol. Biol. 9, 134.

Hulsenbeck, J.P., Ronquist, F.R., 2001. MrBayes: Bayesian inference of phylogeny. Biometrics 17, 754-755.

Inward, D.J.G., Vogler, A.P., Eggleton, P., 2007a. A comprehensive phylogenetic analysis of termites (Isoptera) illuminates key aspects of their evolutionary history. Mol. Phylogenet. Evol. 44, 953-967.

Inward, D., Beccaloni, G., Eggleton, P., 2007b. Death of an order: a comprehensive molecular phylogeneitc study confirms that termites are eusocial cockroaches. Biol. Letters 3, 331-335.

Jermiin, L., Ho, S.Y., Ababneh, F., Robinson, J., Larkum, A.W., 2004. The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimated. Syst. Biol. 53, 638-643.

Kambhampati, S., Eggleton, P., 2000. Phylogenetics and taxonomy. In: Abe, T., Bignell, D.E., Higashi, M. (Eds.), Termites: Evolution, Sociality, Symbiosis, Ecology. Kluwer Academic Publishers, Dordrecht, pp 1-23.

Kambhampati, S., Kjer, K.M., Thorne, B.L., 1996. Phylogenetic relationship among termite families based on DNA sequence of mitochondrial 16S ribosomal RNA gene. Insect Mol. Biol. 5, 229-238.

Kim, M.-J., Kang, A.-R., Jeong, H.-C., Kim, K.-G., Kim, I., 2011. Reconstructing intraordinal relationships in Lepidoptera using mitochondrial genome data with the description of two newly sequenced lycaenids, *Spindasis takanonis* and *Protantigius superans* (Lepidoptera: Lycaenidae). Mol. Phylogenet. Evol. 61, 436-445.

Klass, K.-D., 1995. Die Phylogenie der Dictyoptera. Cuvillier Verlag Göttigen.

Klass, K.-D., 1997. The external male genitalia and the phylogeny of Blattaria and Mantodea. Bonn. Zool. Monogr. 42, 1-341.

Klass, K.-D., Meier, R., 2006. A phylogenetic analysis of Dictyoptera (Insecta) based on morphological characters. Entomol. Abh. 63, 3–50.

Komoto, N., Yukihiro, K., Ueda, K., Tomita, S., 2011. Exploring the molecular phylogeny of phasmids with whole mitochondrial genome sequences. Mol. Phylogenet. Evol. 58, 43-52.

Lacey, M.J., Lenz, M., Evans, T.A. 2010. Cryoprotection in dampwood termites (Termopsidae: Isoptera). J. Insect Physiol. 56, 1-7.

Lavrov, D.V., 2011. Key transitions in animal evolution: a mitochondrial DNA perspective. In Schierwater, B. & DeSalle, R. (Eds.), Key Transitions in Animal Evolution, CRC Press, pp. 34-54.

Legendre, F., Whiting, M.F., Bordereau, C., Cancello, E.M., Evans, T.A., Grandcolas, P., 2008. The phylogeny of termites (Dictyoptera: Isoptera) based on mitochondrial and nuclear genes: implications for the evolution of the worker and pseudergate castes, and foraging behaviors. Mol. Phylogenet. Evol. 48, 615-627.

Lo, N, Eggleton, P., 2011. Termite phylogenetics and co-cladogenesis with symbionts. In: Bignell, D.E., Roisin, Y., Lo, N. (Eds.), Biology of Termites: A Modern Synthesis. Springer, Dordrecht, pp 27-50.

Lo. N., Tokuda, G., Watanabe, H., 2000. Evidence from multiple gene sequences indicated that termites evolved from wood-feeding cockroaches. Curr. Biol. 10, 801-804.

Lo, N., Kitade, O., Miurua, T., Constantino, R., Matsumoto, T., 2004. Molecular phylogeny of the Rhinotermitidae. Insect. Soc. 51, 365-371.

Lo, N, Engel, M.S., Cameron, S.L., Nalepa, C.A., Tokuda, G., Grimaldi, D., Kitade, O., Krishna, K., Klass, K., Maekawa, K., Miura, T., Thompson, G.J., 2007. Save Isoptera: A comment on Inward et al. Biol. Letters 3, 562-563.

Lo, N., Tokuda, G., Watanabe, H., 2011. Evolution and function of endogenous termite cellulases. In: Bignell, D.E., Roisin, Y., Lo, N. (Eds.), Biology of Termites: A Modern Synthesis. Springer, Dordrecht, pp 51-68.

Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25, 955-964.

Maddison, W., Maddison, D., 2003. MacClade ver 4.06. Sinauer Associates, Sunderland.

Margam, V.M., Coates, B.S., Hellmich, R.L., Agunbiade, T., Seufferheld, M.J., Sun, W., Ba, M.N., Sanon, A., Binso-Dabire, C.L., Baoua, I., Ishiyaku M.F., Covasm F.G., Srinivasan, R., Armstrong, J., Murdock, L.L.,

Pittendrigh, B.R., 2011. Mitochondrial genome sequence and expression profiling for the legume pod bores *Maruca vitrata* (Lepidoptera: Crambidae). PLoS One 6, e16444.

McKittrick, F.A., Mackerras, M.J., 1965. Phyletic relationships within the Blattidae. Ann. Entomol. Soc. Am. 58, 224-230.

Murienne, J., 2009. Molecular data confirm family status for the *Tyronicus-Lauraesilpha* group (Insecta: Blattodea: Tyronicidae). Org. Divers. Evol. 9, 44-51.

Murrell, A., Campbell, N.J.H., Barker, S.C., 2003. The value of idiosyncratic markers and changes to conserved tRNA sequences from the mitochondrial genome of hard ticks (Acari: Ixodida: Ixodidae) for phylogenetic inference. Syst. Biol. 52, 296-310.

Nalepa, C.A., Lenz, M., 2000. The ootheca of *Mastotermes darwiniensis* Froggatt (Isoptera: Mastotermitidae): homology with cockroach oothecae. P. R. Soc. Lond. B 267, 1809–1813.

Noirot, C., Pasteels, J.M., 1988. The worker caste is polyphyletic in termites. Sociobiology 14, 15-20.

Ohkuma, M., Yuzawa, H., Amornsak, W., Sornnuwat, Y., Takematsu, Y., Yamada, A., Vongkaluang, C., Sarnthoy, O., Kirtibutr, N., Noparatnaraporn, N., Kudo, T., Inou, T., 2004. Molecular phylogeny of Asian termites (Isoptera) of the families Termitidae and Rhinotermitidae based on mitochondrial COII sequences. Mol. Phylogenet. Evol. 31, 701-710.

Perna, N.T., Kocher, T.D., 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J. Mol. Evol. 41, 353-359.

Pellens, R., D'Haese, C.A., Bellés, X., Piulachs, M.-D., Legendre, F., Wheeler, W.C., Grandcolas, P., 2007. The evolutionary transition from subsocial to eusocial behavior in Dictyoptera: phylogenetic evidence for modification of the 'shift-in-independent-care' hypothesis with a new subsocial cockroach. Mol. Phylogenet. Evol. 43, 616-626.

Phillips, M.J., Penny, D., 2003. The root of the mammalian tree inferred from whole mitochondrial genomes. Mol. Phylogenet. Evol. 28, 171-185.

Posada, D., Crandall, K.A., 1998. ModelTest: Testing the best-fit model of nucleotide substitution. Bioinformatics 14, 817-818.

Rambaut A., Drummond A.J., 2007. Tracer v1.4. http://beast.bio.ed.ac.uk/Tracer

Rokas, A., Holland, P.W.H., 2000. Rare genomic changes as a tool for phylogenetics. Trends Ecol. Evol. 15, 454-459.

Saito, S., Tamura, K., Aotsuka, T., 2005. Replication origin of mitochondrial DNA in insects. Genetics 171, 1695-1705.

Sheffield, N.C., Song, H., Cameron, S.L., Whiting, M.F., 2008. A comparative analysis of mitochondrial genomes in Coleoptera (Arthropoda: Insecta) and genome descriptions of six new beetles. Mol. Biol. Evol. 25, 2499-2509.

Sheffield, N.C., Song, H., Cameron, S.L., Whiting, M.F., 2009. Nonstationary Evolution and Compositional Heterogeneity in Beetle Mitochondrial Phylogenomics. Syst. Biol. 58, 381-394.

Sheffield, N.C., Hiatt, K.D., Valentine, M.C., Song, H., Whiting, M.F., 2010. Mitochondrial genomics in Orthoptera using MOSAS. Mitochondr. DNA 21, 87-104.

Song, H., Sheffield, N.C. Cameron, S.L., Miller, K.B., Whiting, M.F., 2010. What happens when the phylogenetic assumptions are violated?: The effect of base compositional heterogeneity and among-site rate heterogeneity in beetle mitochondrial phylogenomics. Syst. Entomol. 35, 429-448.

Stamatakis, A., Hoover, J., Rougemont. J., 2008. A rapid bootstrap algorithm for the RaxML Web-servers. Syst. Biol. 75, 758-771.

Swofford D.L., 2002. PAUP* Phylogenetic Analysis using Parsimony (*and Other Methods) Ver. 4. Sinauer Associates, Sunderland.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.

Thao, M.L., Baumann, L., Baumann, P., 2004. Organization of the mitochondrial genomes of whiteflies, aphids and pysllids (Hemiptera: Sternorrhyncha). BMC Evol. Biol. 4: 25.

Thompson, G.J., Kitade, O., Lo, N., Crozier, R.H. 2000. Phylogenetic evidence for a single ancestral origin of a “true” worker caste in termites. J. Evol. Biol. 13, 869-881.

Timmermans, M.J.T.N., Dodsworth, S., Culverwell, C.L., Bocak, L., Ahrens, D., Littlewood, D.T.J., Pons, J., Vogler, A.P., 2010. Why barcode? High-throughput multiplex sequencing of mitochondrial genomes for molecular systematic. Nucleic Acids Res. 38: e197.

Ware, J.L., Litman, J., Klass, K.-D., Spearman, L.A., 2008. Relationships among the major lineages of Dictyoptera: the effect of outgroup selection on the dictyopteran tree topology. Syst. Entomol. 33, 429-450.

Ware, J.L., Grimaldi, D.A., Engel, M.S., 2010. The effects of fossil placement and calibration on divergence times and rates: An example from the termites (Insecta: Isoptera). Arthropod Struct. Dev. 39, 204-219.

Watanabe, H., Noda, H., Tokuda, G., Lo, N., 1998. A cellulose gene of termite origin. Nature 394, 330-331.

Wheeler, W.C., Aagesen, L., Arango, C.P., Faivovich, J., Grant, T., D’Haese, C., Janies, D., Smith, W.L., Varon, A., Giribet, G., 2006. Dynamic Homology and Phylogenetic Systematics: A Unified Approach Using POY. American Museum of Natural History, New York.

Yamada, A., Inoue, T., Wiwatwitaya, D., Ohkuma, M., Kudo, T., Abe, T., Sugimoto, A. 2005. Carbon mineralization by termites in tropical forests, with emphasis on fungus combs. Ecol. Res. 20, 453-460.

Yeates, D.K. 1995. Groundplans and exemplars: paths to the tree of life. Cladistics 11, 343-357.

Yoshizawa, K. 2010. Direct optimization overly optimizes data. Syst. Entomol. 35, 199-206.

Tables

Table 1. Samples used in this study, collection data and GenBank accession numbers.

Species	Family	Subfamily	Collecting locality / Colony source / Publication source	Date	Collector	Accession Number
Mastotermes darwiniensis	Mastotermitidae	N/A	Darwin, NT, Australia	1-Nov-1996		XXXXXX
Porotermes adamsoni	Stolotermitidae	N/A	Glen Elgin State Fores, NSW, Australia	1-May-2008	SLC & TAE	XXXXXX
Microhodotermes viator	Hodotermitidae	N/A	Worcester, South Africa	20-June-2005		XXXXXX
Zootermopsis augusticollis	Archaeotermopsidae	N/A	Triangle Mountain, British Columbia, Canada	11-Apr-2010	R. West	XXXXXX
Neotermes insularis	Kalotermitidae	N/A	Kalpowar State Forest, QLD, Australia	5-May-2008	SLC, TAE & NL	XXXXXX
Coptotermes lacteus	Rhinotermitidae	Coptotermitinae	Glen Elgin State Fores, NSW, Australia	1-May-2008	SLC & TAE	XXXXXX
Schedorhinotermes breinli	Rhinotermitidae	Rhinotermitinae	19km. Nth of Ayr, QLD, Australia	8-May-2008	SLC, TAE & NL	XXXXXX
Heterotermes sp.	Rhinotermitidae	Heterotermitinae	Pilliga Nature Reserve, NSW, Australia	19-Apr-2008	SLC & TAE	XXXXXX
Reticulitermes santonensis	Rhinotermitidae	Heterotermitinae	GenBank – Cameron & Whiting, 2007			EF206315
Macrotermes subhyalinus	Termitidae	Macrotermitinae	Colony – University of Dijon,			XXXXXX
Drepanotermes sp.	Termitidae	Amitermitinae	Abbot Point, QLD, Australia	8-May-2008	SLC, TAE & NL	XXXXXX
Macrognathotermes errator	Termitidae	Termitinae	Mt. Molloy, QLD, Australia	10-May-2008	SLC, TAE & NL	XXXXXX
Nasutitermes triodinae	Termitidae	Nasutitermitinae	Bilwon State Forest, QLD, Australia	10-May-2008	SLC, TAE & NL	XXXXXX
Cryptocercus russie	Cryptocercidae	N/A	No collection details			XXXXXX
Blatella germanica	Blattellidae	Blattellinae	GenBank – Jiang & Xiao unpublished			EU854321
Periplaneta fuliginosa	Blattidae	Blattinae	GenBank – Yamauchi et al. 2004			AB126004
Eupolyphaga sinensis	Polyphagidae	N/A	GenBank – Zhang et al. 2010			FJ830540
Tamolanica	Mantidae	Mantinae	GenBank – Cameron et al. 2006			DQ241797

tamolana						
Outgroups						
Sclerophasma paresiense	Mantophasmatidae	N/A	GenBank – Cameron et al. 2006			DQ241798
Megacrania alpheus	Phasmatidae	Platycraninae	GenBank – Komoto et al. 2011			AB477471
Locusta migratoria	Acrididae	Oedipodinae	GenBank – Flook et al. 1995			X80245

Table 2 Dictyoptera and termite general primers designed for this study.

Primer Name	Gene	Location[1]	Direction[2]	Sequence	Design Consensus
DICTY-1	*cox3*	353	N	AWG GRW TRA TWC CTR WWG GNG GTC	All Dictyoptera
DICTY-4	*nadh4*	631	J	ATT ATW GAW CCW GAW ACR GGR GC	All Dictyoptera
DICTY-7	*cob*	263	J	AGC AAC MYT MCA YGC AAA YGG RGC	All Dictyoptera
DICTY-8	*nad3*	134	J	MTT YGA RTG YGG RTT YGA YCC	All Dictyoptera
PORs4	*rrnL*	1066	N	ATT ACC TTA GGG ATA ACA GCG	All Termites
ISOs2	*trnL*	9	N	TTA AAT CCA VYG CAC TTA TCT GCC	All Termites
ISOs7	*cox1*	454	N	TCC YAR RAT TGA TGA WAC WCC TGC	Neoisoptera

[1] Location of 3’ base relative to the start of the gene in which the primer site occurs.
[2] Relative to majority strand of mt genome.

Table 3. Nucleotide composition statistics. Taxon groups are the average of all members of that group.

	Whole mt genome			Majority strand PCGs				Minority strand PCGs			
	A+T%	A-skew	C-skew	A+T% overall	A+T% 3rd codon	A-skew	C-skew	A+T% overall	A+T% 3rd codon	A-skew	C-skew
Non-termite Dictyoptera *	74.169	0.08893	0.20578	72.067	86.413	-0.04383	0.13118	76.148	88.434	-0.27342	-0.29186
Cryptocercus + termites clade	66.690	0.29712	0.28782	64.104	71.629	0.17034	0.25776	68.502	74.436	-0.5029	-0.34402
Cryptocercus	73.219	0.23815	0.23863	70.850	83.326	0.13104	0.18600	75.197	85.456	-0.40136	-0.30841
Mastotermes	67.794	0.16822	0.24595	65.220	72.290	0.03350	0.20268	70.091	75.314	-0.37401	-0.31469
Euisoptera	66.054	0.31278	0.29540	63.449	70.599	0.18502	0.26833	67.812	73.445	-0.52211	-0.34943
Neoisoptera	65.308	0.31845	0.30138	62.514	68.911	0.19379	0.27692	67.146	72.091	-0.52763	-0.34884

* *Tamolanica*, *Blatella*, *Periplaneta* and *Eupolyphaga*.

Figure Legends:

Figure 1. Previous hypotheses of family level relationships within termites. A: Kambhampati et al. (1996); B: Kambhampati & Eggleton (2000); C: Donovan et al. (2000); D: Thompson et al. (2000); E: Inward et al. (2007); F: Legendre et al. (2008); G: Engel et al. (2009); H: Ware et al. (2010); I: Lo & Eggleton (2010). Family names are as per Engel et al. (2010) with terminals changed from the original publication to track this classification. Nodes marked with an * are significant for the data type and analyses used in the original study: morphology, at least one unambiguous synapomorphy (G), or decay index > 5 (C), molecular, bootstrap values >70% (A, D, F), decay index >10 (E), or posterior probability of >0.9 (H). Nodal significance not reported for B and I.

Figure 2. Phylogenetic tree of termites based on mt genomic data excluding 3rd codon positions. Nodal supports are the likelihood bootstrap percentages in the following form: ML results above, Bayesian results below; 5-partition/16 partition datasets; * indicates 100% bootstrap support and 1.0 posterior probability from all 4 analyses. Distribution of RGCs are marked with arrow, the structure of each RGC is shown, consensus secondary structure of RGC 1 (IUPAC symbols for consensus sequences), the repeat structure of RGC 2 (ncr: non-coding region, A: A-type repeat, B: B-type repeat, A(p), B(p): partial repeat units). Scale bar = 0.05 expected changes per site.

Figure 1. B&W in final.

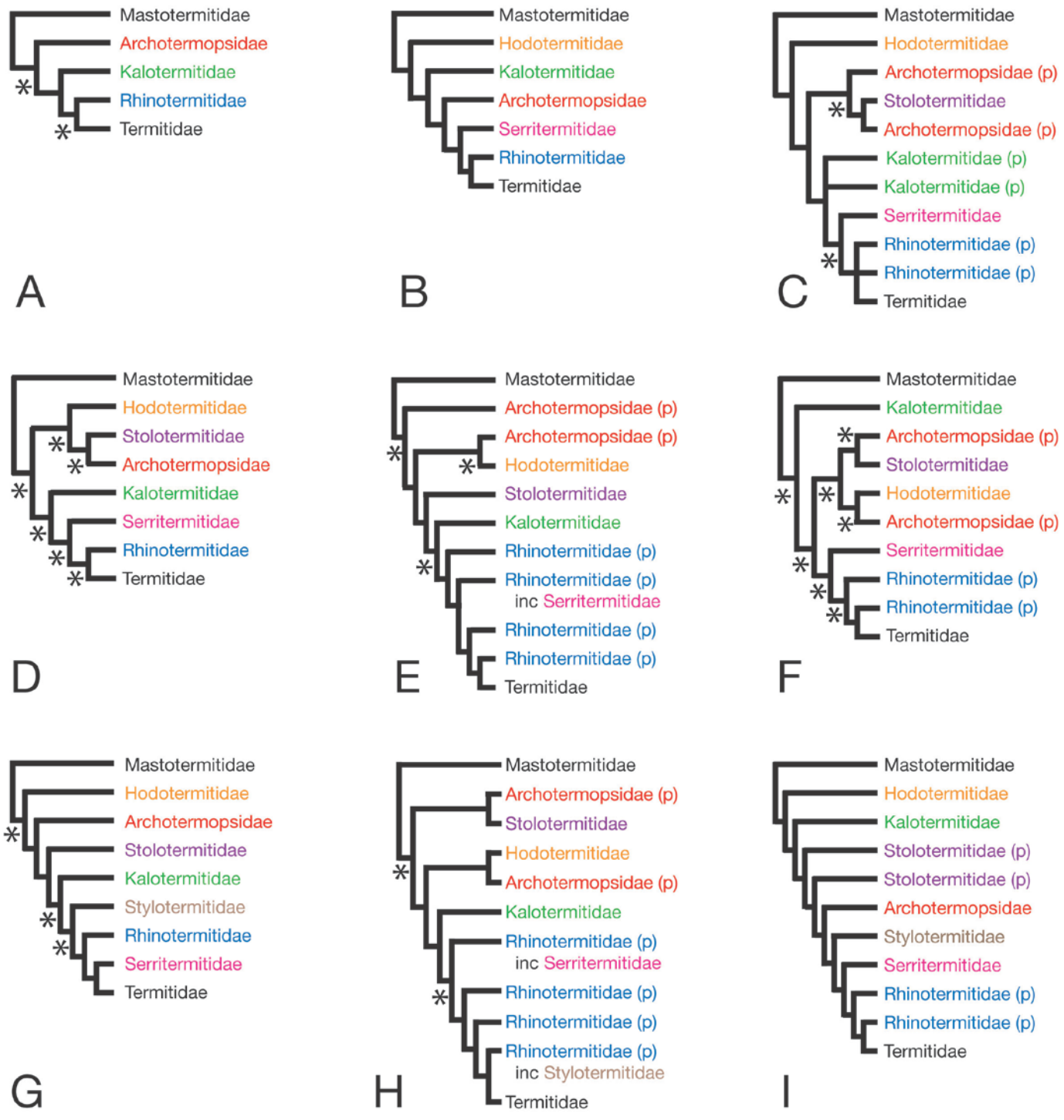
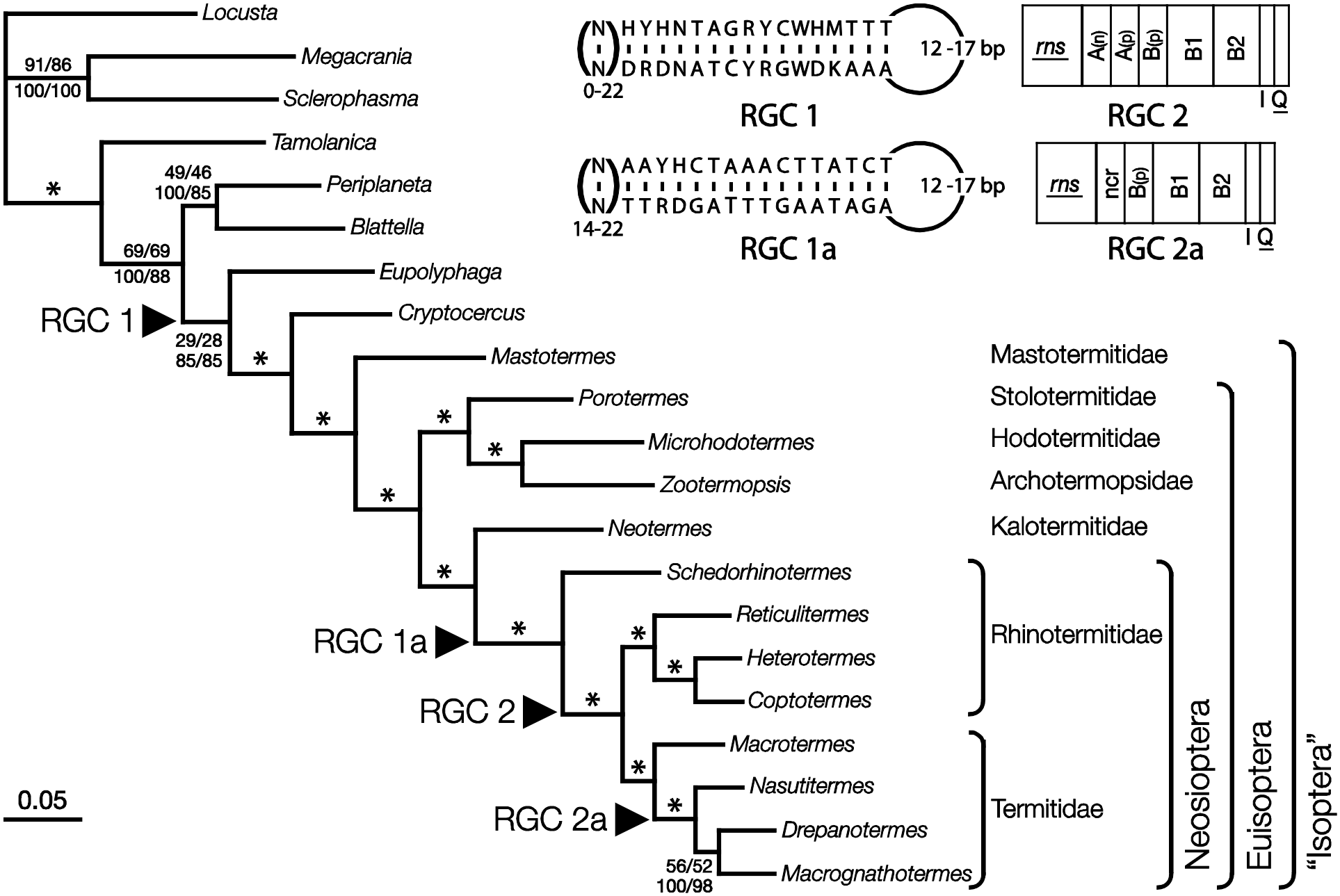


Figure 2.


Locusta
Megacrania
Sclerophasma
Tamolanica
Periplaneta
Blattella
Eupolyphaga
Cryptocercus
Mastotermes
Porotermes
Microhodotermes
Zootermopsis
Neotermes
Schedorhinotermes
Reticulitermes
Heterotermes
Coptotermes
Macrotermes
Nasutitermes
Drepanotermes
Macrognathotermes
91/86
100/100
49/46
100/85
69/69
100/88
29/28
85/85
56/52
100/98
RGC 1
RGC 1a
RGC 2
RGC 2a
0.05
(N)HYHNTAGRYCWHMTTT
(N)DRDNATCYRGWDKAAA
12 -17 bp
0-22
RGC 1
rns
A(n)
A(p)
B(p)
B1
B2
I
Q
RGC 2
(N)AAYHCTAAACTTATCT
(N)TTRDGATTTGAATAGA
12 -17 bp
14-22
RGC 1a
rns
ncr
B(p)
B1
B2
I
Q
RGC 2a
Mastotermitidae
Stolotermitidae
Hodotermitidae
Archotermopsidae
Kalotermitidae
Rhinotermitidae
Termitidae
Neosioptera
Euisoptera
"Isoptera"

Supplementary Tables

Supplementary Table S1. Amplification strategy and primers used in this study.

Region	Primers (F & R)	Sequence (5' →3')
Mastotermes darwiniensis		
cox1	C1-J-1718[1]	GGA GGA TTT GGA AAT TGA TTA GTT CC
	C1-N-2329[1]	ACT GTA AAT ATA TGA TGA GCT CA
cox1 → *cox2*	MSs8[2]	ATC ACC ATA CTA TTA ACA GAC CGC
	MSs1[2]	ATG ATT ATC AAG GCG TGA TCA TGG
cox2	F Leu[3]	TCT AAT ATG GCA GAT TAG TGC
	R Lys[3]	GAG ACC AGT ACT TGC TTT CAG TCA TC
cox2 → *nadh4*	MSs34[2]	ATC TAC AAG AAC ATT TAA CCT CCC
	MSs33[2]	ATA TGA ACG TCT TGG AAG ACG GAG
subPCR *cox3* → *nadh5*	MSs34[2]	ATC TAC AAG AAC ATT TAA CCT CCC
	MSs42[2]	ATC GCT TCT TAT TTG AGG TAG ACC
subPCR *nadh5* → *nadh4*	MSs46[2]	AAA GGT AAA AAA GTA ATC ACG GG
	MSs33[2]	ATA TGA ACG TCT TGG AAG ACG GAG
nadh4 → *cob*	N4-J-8944[1]	GGAGCTTCAACATGAGCTTT
	CB-N-10920[1]	CCCTCAGAATGATATTTGTCCTCA
cob → *rrnL*	CB-J-10612[1]	CCATCCAACATCTCAGCATGATGAAA
	MSs4[2]	TAA ATT ACC TTA GGG ATA ACA GCG
rrnL	16SB[4]	CTC CGG TTT GAA CTC AGA TCA
	16SA[4]	CGC CTG TTT ATC AAA AAC AT
rrnL → *rrnS*	MSs3[2]	TGC TCA AAC CAT TCA TTC CAG CCC
	MSs5[2]	TGA TAA TAT TTC AGG TCA AGG TGC
rrnS → *trnM*	MSs6[2]	AAA AGA TCT TCG TTA TAA CGG CGG
	TM-N-193[1]	TGG GGT ATG AAC CCA GTA GC
trnM → *cox1*	MSs12[2]	TAT TCT ATC ACA ATG AAG TGC CTG
	MSs7[2]	ATG GGC AAT CCC TCT CGC CAA GGG
Porotermes adamsoni		
cox1 → *cox2*	C1-J-1718[1]	GGA GGA TTT GGA AAT TGA TTA GTT CC
	R Lys[3]	GAG ACC AGT ACT TGC TTT CAG TCA TC
cox2 → *cox3*	F Leu[3]	TCT AAT ATG GCA GAT TAG TGC
	DICTY-1[6]	AWG GRW TRA TWC CTR WWG GNG GTC
cox3 → *nadh4*	PORs8[2]	TCT CAG TAT TTG ATC CAT CGA C
	PORs13[2]	AGC CTG AAC GTC TCC AAG CTG G
nadh4 → *rrnL*	DICTY-4[6]	ATT ATW GAW CCW GAW ACR GGR GC
	PORs4[6]	ATT ACC TTA GGG ATA ACA GCG
rrnL → *rrnS*	16SB[4]	CTC CGG TTT GAA CTC AGA TCA
	SR-N-14594[5]	AAA CTA GGA TTA GAT ACC C
rrnS → *cox1*	PORs7[2]	ACC AAG TAA GGT CCA ACG CGG
	PORs1[2]	TAG GAT TGA GGA TAC ACC AGC
Microhodotermes viator		
cox1 → *cox2*	C1-J-1718[1]	GGA GGA TTT GGA AAT TGA TTA GTT CC
	R Lys[3]	GAG ACC AGT ACT TGC TTT CAG TCA TC
cox2 → *cox3*	F Leu[3]	TCT AAT ATG GCA GAT TAG TGC
	DICTY-1[6]	AWG GRW TRA TWC CTR WWG GNG GTC
cox3 → *nadh4*	MICs3[2]	ATT CAA GCA TAC GTA TTC GCC G
	MICs22[2]	TGA TGG TTA TGG CCA GTG AGC C
nadh4 → *rrnL*	DICTY-4[6]	ATT ATW GAW CCW GAW ACR GGR GC
	PORs4[6]	ATT ACC TTA GGG ATA ACA GCG
rrnL → *rrnS*	16SB[4]	CTC CGG TTT GAA CTC AGA TCA

	SR-N-14594[5]	AAA CTA GGA TTA GAT ACC C
rrnS → trnL	MICs6[2]	ACC AAG TAA GGT CCA ACG TGG
	ISOs2[2]	TTA AAT CCA VYG CAC TTA TCT GCC
Zootermopsis augusticollis		
cox1 → nadh4	ZOOs7[2]	ACC AAT CCT AAT TGG AGG TTT CGG
	ZOOs8[2]	ACC TGA GCG TCT TCA GGC TGG
nadh3 → nadh4	DICTY-8[6]	MTT YGA RTG YGG RTT YGA YCC
	ZOOs8[2]	ACC TGA GCG TCT TCA GGC TGG
nadh4 → rrnL	DICTY-4[6]	ATT ATW GAW CCW GAW ACR GGR GC
	PORs4[6]	ATT ACC TTA GGG ATA ACA GCG
cob→ rrnL	DICTY-7[6]	AGC AAC MYT MCA YGC AAA YGG RGC
	PORs4[6]	ATT ACC TTA GGG ATA ACA GCG
rrnL → rrnS	16SB[4]	CTC CGG TTT GAA CTC AGA TCA
	SR-N-14594[5]	AAA CTA GGA TTA GAT ACC C
rrnS → cox1	ZOOs1[2]	TAA AAG CTG CAC CTT GAC CTG
	ISOs7[6]	TCC YAR RAT TGA TGA WAC WCC TGC
Neotermes insularis		
cox1 → cox2	C1-J-1718[1]	GGA GGA TTT GGA AAT TGA TTA GTT CC
	R Lys[3]	GAG ACC AGT ACT TGC TTT CAG TCA TC
cox2 → cox3	F Leu[3]	TCT AAT ATG GCA GAT TAG TGC
	DICTY-1[6]	AWG GRW TRA TWC CTR WWG GNG GTC
cox3 → nadh4	NEOs4[2]	AGT AAC AGG AGC CAT TGG AGC
	NEOs5[2]	TGG CTA GAG AGT CGG TTT TGC GC
nadh4 → rrnL	N4-J-8944[1]	GGA GCT TCA ACA TGA GCT TT
	16SA[4]	CGC CTG TTT ATC AAA AAC AT
rrnL → rrnS	16SB[4]	CTC CGG TTT GAA CTC AGA TCA
	SR-N-14594[5]	AAA CTA GGA TTA GAT ACC C
rrnS → cox1	NEOs10[2]	AGA TCC TCA CCA CAA CGG CGG
	NEOs1[2]	AGT TAA CTG CTC CTA GGA TGG
Coptotermes lacteus		
cox1 → cox2	C1-J-1718[1]	GGA GGA TTT GGA AAT TGA TTA GTT CC
	R Lys[3]	GAG ACC AGT ACT TGC TTT CAG TCA TC
cox2 → cox3	COPs2[2]	ACA GAT GCC ACA CCA GGA CG
	DICTY-1[6]	AWG GRW TRA TWC CTR WWG GNG GTC
cox3 → nadh4	COPs8[2]	TGC TGG CCA TCT ACT ACT CAC CC
	COPs16[2]	TGA GCG TAT TCA GGC TGG CG
nadh4 → rrnL	DICTY-4[6]	ATT ATW GAW CCW GAW ACR GGR GC
	COPs3[2]	AGG AAT GAT TTA ACT CCT CTT GG
rrnL → rrnS	16SB[4]	CTC CGG TTT GAA CTC AGA TCA
	SR-N-14594[5]	AAA CTA GGA TTA GAT ACC C
rrnS → cox1	COPs6[2]	ACG GCG GTA TAC AAA CCA TAG C
	COPs1[2]	TAG GTG TAG GGA GAA GAT GGC
Schedorhinotermes breinli		
cox1 → cox2	C1-J-1718[1]	GGA GGA TTT GGA AAT TGA TTA GTT CC
	R Lys[3]	GAG ACC AGT ACT TGC TTT CAG TCA TC
cox2 → cox3	F Leu[3]	TCT AAT ATG GCA GAT TAG TGC
	DICTY-1[6]	AWG GRW TRA TWC CTR WWG GNG GTC
cox3 → nadh4	SCHs7[2]	ATC CGA CCA GGA ACC CTG GC
	SCHs12[2]	AGC CTG AGC GTG TTC AGG CTG G
nadh4 → rrnL	DICTY-4[6]	ATT ATW GAW CCW GAW ACR GGR GC
	PORs4[6]	ATT ACC TTA GGG ATA ACA GCG
rrnL → rrnS	16SB[4]	CTC CGG TTT GAA CTC AGA TCA
	SR-N-14594[5]	AAA CTA GGA TTA GAT ACC C

rrnS → *cox1*	SCHs4[2]	AGT AAG GTA CAA CGC GGA TTA TCG
	SCHs1[2]	TGA TGA TAC ACC TGC TAG ATG G
Heterotermes sp.		
cox1 → *cox2*	C1-J-1718[1]	GGA GGA TTT GGA AAT TGA TTA GTT CC
	R Lys[3]	GAG ACC AGT ACT TGC TTT CAG TCA TC
cox2 → *cox3*	F Leu[3]	TCT AAT ATG GCA GAT TAG TGC
	DICTY-1[6]	AWG GRW TRA TWC CTR WWG GNG GTC
cox3 → *nadh4*	HETs3[2]	TCG AAC ACC TTG TAC CAC AAG G
	HETs8[2]	TAC TTT GTT AGC GTC CCT TCC
nadh4 → *rrnL*	DICTY-4[6]	ATT ATW GAW CCW GAW ACR GGR GC
	PORs4[6]	ATT ACC TTA GGG ATA ACA GCG
rrnL → *rrnS*	16SB[4]	CTC CGG TTT GAA CTC AGA TCA
	SR-N-14594[5]	AAA CTA GGA TTA GAT ACC C
rrnS → *cox1*	HETs6[2]	AGA AAC AAG CTG CAC CTT GAC C
	HETs7[2]	TGA TGA TAC ACC TGC TAG GTG G
Macrotermes subhyalinus		
cox1 → *cox2*	C1-J-1718[1]	GGA GGA TTT GGA AAT TGA TTA GTT CC
	R Lys[3]	GAG ACC AGT ACT TGC TTT CAG TCA TC
cox2 → *cox3*	F Leu[3]	TCT AAT ATG GCA GAT TAG TGC
	DICTY-1[6]	AWG GRW TRA TWC CTR WWG GNG GTC
cox3 → *nadh4*	MTEs5[2]	ACC GCA AAC ATG ATC GCA GG
	MTEs10[2]	ACC TGA GCG TGT TCA GGC TGG
nadh4 → *nadh1*	DICTY-4[6]	ATT ATW GAW CCW GAW ACR GGR GC
	MTEs26[2]	TGT CCT GTT AGG ATG TAT GGG
cob → *rrnL*	MTEs25[2]	TGC CGA GAC GTA AAC TAC GG
	MTEs6[2]	ACC CTA TAG AGT TTA ACA TTC GG
rrnL → *rrnS*	16SB[4]	CTC CGG TTT GAA CTC AGA TCA
	SR-N-14594[5]	AAA CTA GGA TTA GAT ACC C
rrnS → *cox1*	MTEs5[2]	ACC GCA AAC ATG ATC GCA GG
	MTEs10[2]	ACC TGA GCG TGT TCA GGC TGG
Macrognathotermes errator		
cox1 → *cox2*	C1-J-1718[1]	GGA GGA TTT GGA AAT TGA TTA GTT CC
	R Lys[3]	GAG ACC AGT ACT TGC TTT CAG TCA TC
cox2 → *cox3*	F Leu[3]	TCT AAT ATG GCA GAT TAG TGC
	DICTY-1[6]	AWG GRW TRA TWC CTR WWG GNG GTC
cox3 → *nadh4*	MACs4[2]	AGA CCT TGA CCA CTA ACA GGG
	MACs8[2]	TGA GCG GGT TCA GGC TGG
nadh4 → *rrnL*	DICTY-4[6]	ATT ATW GAW CCW GAW ACR GGR GC
	PORs4[6]	ATT ACC TTA GGG ATA ACA GCG
rrnL → *rrnS*	16SB[4]	CTC CGG TTT GAA CTC AGA TCA
	SR-N-14594[5]	AAA CTA GGA TTA GAT ACC C
rrnS → *cox1*	MACs7[2]	AGA AAC AAG CTG CAC CTT GAC C
	MACs1[2]	AGA AGT AGT AGG GCA GTA ATG GC
Drepanotermes sp.		
cox1 → *cox2*	C1-J-1718[1]	GGA GGA TTT GGA AAT TGA TTA GTT CC
	R Lys[3]	GAG ACC AGT ACT TGC TTT CAG TCA TC
cox2 → *cox3*	F Leu[3]	TCT AAT ATG GCA GAT TAG TGC
	DICTY-1[6]	AWG GRW TRA TWC CTR WWG GNG GTC
cox3 → *nadh4*	DREs4[2]	TGA ACA CTT AGT GCC ACA AGG
	DREs16[2]	AGA ATC GTG TGA GTG TGG CG
nadh4 → *rrnL*	DICTY-4[6]	ATT ATW GAW CCW GAW ACR GGR GC
	PORs4[6]	ATT ACC TTA GGG ATA ACA GCG
cob→ *rrnL*	DICTY-7[6]	AGC AAC MYT MCA YGC AAA YGG RGC

	PORs4[6]	ATT ACC TTA GGG ATA ACA GCG
rrnL → rrnS	16SB[4]	CTC CGG TTT GAA CTC AGA TCA
	SR-N-14594[5]	AAA CTA GGA TTA GAT ACC C
rrnS → trnM	DREs8[2]	ACG GCG GTA TAC AAA CCA AAG C
	TM-N-193[1]	TGGGGTATGAACCCAGTAGC
trnM → cox1	DREs14[2]	TGC ATT CAC TCT AAG AAT CAT CC
	DREs12[2]	ACT GAT GCT CCG GCA TGG GC
Nasutitermes triodinae		
cox1 → cox2	C1-J-1718[1]	GGA GGA TTT GGA AAT TGA TTA GTT CC
	R Lys[3]	GAG ACC AGT ACT TGC TTT CAG TCA TC
cox2 → cox3	F Leu[3]	TCT AAT ATG GCA GAT TAG TGC
	DICTY-1[6]	AWG GRW TRA TWC CTR WWG GNG GTC
cox3 → nadh4	NASs4[2]	ACC ATG GCC TCT TAC AGG TGC
	NASs9[2]	ACC TGA ACG GGT TCA GGC TGG
nadh4 → rrnL	DICTY-4[6]	ATT ATW GAW CCW GAW ACR GGR GC
	NASs5[2]	ATT ACC TTA GGG ATA ACA GCG
rrnL → rrnS	16SB[4]	CTC CGG TTT GAA CTC AGA TCA
	SR-N-14594[5]	AAA CTA GGA TTA GAT ACC C
rrnS → cox1	NASs8[2]	ACA AGC TGC ACC TTG ACC TG
	NASs1[2]	AGG ATG GAT GAT ACT CCT GC
Cryptocercus russei		
cox1 → cox2	C1-J-1718[1]	GGA GGA TTT GGA AAT TGA TTA GTT CC
	R Lys[3]	GAG ACC AGT ACT TGC TTT CAG TCA TC
cox2 → nadh4	CRYs6[2]	AGC AGA TGC TAC ACC AGG ACG
	CRYs7[2]	AGA TCT TGT AAT ATA GCC GCT CCC
subPCR *cox2 → nadh5*	CRYs6[2]	AGC AGA TGC TAC ACC AGG ACG
	CRYs25[2]	ATT GAC TGT TTG TTA TTC ATT TCG
subPCR *nad5 → nadh4*	CRYs24[2]	TAT ATC TCA ATC TAC TGA TGA GG
	CRYs13[2]	TCC TTC TTT AGT GCT GTT TAT ACA C
nadh4		Sequence from Svenson & Whiting (2009)
nadh4 → rrnL	CRYs8[2]	AGT AGG AAT CAA GCT ACC CTC
	CRYs2[2]	ACT AAA TTA CCT TAG GGA TAA CAG CG
rrnL		Sequence from Svenson & Whiting (2009)
rrnL → cox1	CRYs1[2]	ATT ATG CTA CCT TTG CAC GGT C
	CRYs3[2]	ACT AAT CAG TTA CCA AAT CCT CCG

[1] Primers taken from Simon *et al*. (1994) *Annals of the Entomological Society of America* **87**: 651-701.
[2] Primers specifically designed for sequencing this genome
[3] Primers taken from Whiting (2002) *Zoologica Scripta* **31**: 93-104.
[4] Primers taken from Bybee *et al*. (2004) *Molecular Phylogenetics & Evolution* **30**: 789-797.
[5] Primers taken from Skerratt et al. (2002) *Parasitology Research* **88**: 376-379.
[6] Primers designed to work across Dictyoptera (DICTY-) or termites (ISOs)
Svenson & Whiting (2009) *Cladistics* **25**: 468-514.

Supplementary Table S2 Genome annotations

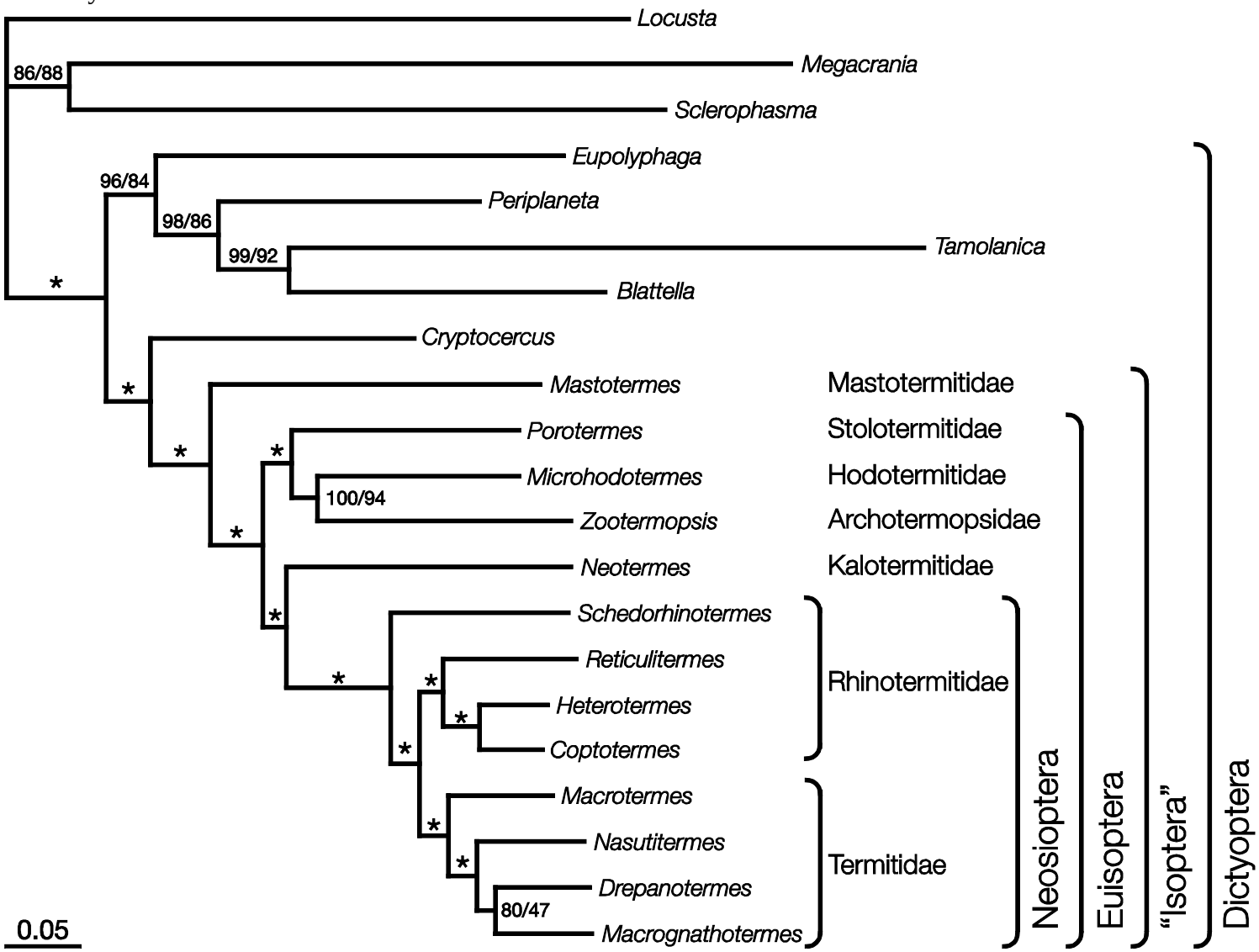
See separate Excel spreadsheet.

Supplementary Table S3 Skew statistics for all genome included in this study.

	Whole coding region			Majority strand PCGs			Minority strand PCGs		
	A-skew	C-skew	A+T%	A-skew	C-skew	A+T%	A-skew	C-skew	A+T%
Locusta	0.18234	0.17845	0.74669	0.06933	0.16622	0.72663	-0.38757	-0.19881	0.76539
Megacrania	0.20446	0.23329	0.76226	0.10913	0.18421	0.74317	-0.37433	-0.30256	0.77373
Sclerophasma	0.10477	0.15116	0.74630	-0.02771	0.10604	0.72531	-0.29169	-0.19250	0.76502
Tamolanica	0.05739	0.23866	0.75307	-0.06996	0.16577	0.72975	-0.23924	-0.31590	0.77658
Eupolyphaga	0.12784	0.24501	0.71700	-0.01026	0.17028	0.69728	-0.32187	-0.33048	0.72894
Periplaneta	0.11999	0.16499	0.75122	-0.00876	0.08528	0.73183	-0.28775	-0.25577	0.76825
Blattella	0.05049	0.17447	0.74546	-0.08634	0.10338	0.72382	-0.24481	-0.26531	0.77215
Cryptocercus	0.23815	0.23863	0.73219	0.13104	0.18600	0.70850	-0.40136	-0.30841	0.75197
Mastotermes	0.16822	0.24595	0.67794	0.03350	0.20268	0.65220	-0.37401	-0.31469	0.70091
Porotermes	0.28402	0.28128	0.66418	0.14195	0.24694	0.64248	-0.51027	-0.36093	0.67954
Microhodotermes	0.33287	0.29379	0.67100	0.20826	0.27371	0.64852	-0.51787	-0.33233	0.68993
Zootermopsis	0.32530	0.29558	0.68878	0.19051	0.24957	0.66428	-0.54920	-0.40557	0.69953
Neotermes	0.26349	0.26320	0.67787	0.12918	0.23437	0.65749	-0.46689	-0.30361	0.69681
Schedorhinotermes	0.33960	0.31553	0.65448	0.20948	0.29146	0.62016	-0.55143	-0.35817	0.67558
Coptotermes	0.34718	0.32787	0.63984	0.21462	0.28058	0.61619	-0.54850	-0.36121	0.66022
Heterotermes	0.33403	0.30260	0.64371	0.23120	0.31535	0.61716	-0.55382	-0.35632	0.65557
Reticulitermes	0.30866	0.29430	0.65510	0.18642	0.26523	0.62976	-0.52497	-0.35315	0.67117
Macrotermes	0.35769	0.33256	0.65206	0.23669	0.30719	0.62107	-0.56200	-0.39786	0.67326
Nasutitermes	0.29324	0.29066	0.65503	0.15835	0.24627	0.63886	-0.48305	-0.29804	0.67930
Macrognathotermes	0.28584	0.27964	0.66074	0.16305	0.26124	0.62429	-0.50397	-0.33619	0.67419
Drepanotermes	0.28137	0.26785	0.66371	0.15054	0.24801	0.63361	-0.49329	-0.32978	0.68237
Majority Strand	Pos1			Pos2			Pos3		
Locusta	0.16197	-0.12840	0.66420	-0.33868	0.24656	0.65200	0.30610	0.68690	0.86368
Megacrania	0.26701	-0.09524	0.68574	-0.33992	0.26035	0.66961	0.32896	0.68421	0.87428
Sclerophasma	0.04064	-0.18631	0.65546	-0.37575	0.23522	0.65253	0.18228	0.52980	0.86801
Tamolanica	0.02689	-0.14583	0.66507	-0.34902	0.25963	0.64878	0.06334	0.74035	0.87555
Eupolyphaga	0.11777	-0.08088	0.64289	-0.34969	0.25000	0.64258	0.15824	0.48753	0.80658
Periplaneta	0.10168	-0.14888	0.64819	-0.35758	0.21295	0.64904	0.16358	0.45690	0.89851
Blattella	0.01205	-0.17233	0.65269	-0.36778	0.24847	0.64292	0.04691	0.45775	0.87587
Cryptocercus	0.19813	-0.10943	0.65269	-0.31921	0.23786	0.63970	0.42437	0.69029	0.83326
Mastotermes	0.13056	-0.04461	0.59869	-0.35169	0.24551	0.63505	0.29141	0.50473	0.72290
Porotermes	0.21127	-0.03507	0.58908	-0.32075	0.24942	0.62516	0.49080	0.64885	0.71335
Microhodotermes	0.24726	-0.03913	0.59843	-0.31915	0.25909	0.61572	0.62030	0.76260	0.73144
Zootermopsis	0.21293	-0.08585	0.62276	-0.29837	0.26835	0.61805	0.57368	0.73145	0.75208
Neotermes	0.21624	-0.03640	0.59196	-0.33852	0.24857	0.61844	0.44119	0.67647	0.76213
Schedorhinotermes	0.26571	-0.01545	0.57635	-0.31789	0.27473	0.60279	0.62844	0.72055	0.68136
Coptotermes	0.25578	-0.01505	0.56558	-0.31766	0.27109	0.61230	0.66580	0.68212	0.67074
Heterotermes	0.27077	0.00101	0.56694	-0.31237	0.28027	0.61065	0.69041	0.77510	0.67394

Reticulitermes	0.24102	-0.03975	0.58290	-0.31335	0.28764	0.61152	0.58040	0.65379	0.69489
Macrotermes	0.27412	-0.00916	0.57055	-0.30873	0.27873	0.61582	0.70155	0.76184	0.67687
Nasutitermes	0.22264	-0.05797	0.57835	-0.32394	0.27356	0.62009	0.52311	0.66512	0.71822
Macrognathotermes	0.24249	-0.02725	0.56725	-0.32054	0.28345	0.61468	0.52815	0.63791	0.69100
Drepanotermes	0.22231	-0.05296	0.57948	-0.31157	0.29318	0.61555	0.49474	0.61961	0.70586
Minority Strand	Pos1			Pos2			Pos3		
Locusta	-0.29310	-0.36269	0.73007	-0.43537	0.06494	0.67670	-0.42880	-0.56962	0.88943
Megacrania	-0.23562	-0.34262	0.75017	-0.44655	-0.08966	0.69708	-0.43586	-0.73481	0.87396
Sclerophasma	-0.16635	-0.33158	0.73574	-0.44534	0.01114	0.68754	-0.27636	-0.40217	0.87187
Tamolanica	-0.20503	-0.47583	0.72460	-0.44097	-0.08966	0.69495	-0.11248	-0.59375	0.91024
Eupolyphaga	-0.22429	-0.46154	0.71051	-0.44075	-0.08861	0.66992	-0.30915	-0.54676	0.80641
Periplaneta	-0.21588	-0.43641	0.72036	-0.46051	-0.03704	0.67992	-0.21511	-0.45985	0.90446
Blattella	-0.19846	-0.43434	0.72385	-0.44742	-0.06034	0.67643	-0.13176	-0.50000	0.91626
Cryptocercus	-0.29332	-0.42365	0.71786	-0.48932	-0.05495	0.68359	-0.42182	-0.63636	0.85456
Mastotermes	-0.26273	-0.42163	0.68432	-0.47170	-0.04583	0.66527	-0.38889	-0.54237	0.75314
Porotermes	-0.33831	-0.40726	0.65387	-0.53912	-0.11022	0.65154	-0.63810	-0.62827	0.73324
Microhodotermes	-0.36809	-0.38337	0.65597	-0.51012	-0.09717	0.65527	-0.65409	-0.59538	0.75855
Zootermopsis	-0.42560	-0.49260	0.67015	-0.53029	-0.07724	0.65666	-0.67269	-0.77370	0.77181
Neotermes	-0.32841	-0.38843	0.66177	-0.51299	-0.07905	0.64615	-0.54602	-0.53698	0.78252
Schedorhinotermes	-0.38363	-0.43320	0.65575	-0.52660	-0.05263	0.64226	-0.72440	-0.66581	0.72873
Coptotermes	-0.38727	-0.43762	0.64735	-0.53560	-0.06564	0.63802	-0.71055	-0.62385	0.69532
Heterotermes	-0.37242	-0.40509	0.64316	-0.54405	-0.05927	0.63452	-0.73225	-0.64944	0.68903
Reticulitermes	-0.36285	-0.41223	0.64620	-0.53595	-0.08171	0.64106	-0.65962	-0.63265	0.72626
Macrotermes	-0.36730	-0.43902	0.65690	-0.53443	-0.08494	0.63852	-0.76301	-0.75696	0.72435
Nasutitermes	-0.31629	-0.35758	0.65481	-0.51410	-0.06458	0.64341	-0.60377	-0.53887	0.73971
Macrognathotermes	-0.32479	-0.37349	0.65272	-0.51940	-0.06139	0.64759	-0.65217	-0.63819	0.72226
Drepanotermes	-0.29412	-0.42604	0.65650	-0.49843	-0.02609	0.64903	-0.66529	-0.61417	0.74161

Supplementary Figure S1 Phylogenetic tree of termites based of mt genomic data with 3rd codon positions included. Nodal supports are the likelihood bootstrap percentages in the following form: 6-partitions/16 partitions; * indicates 100/100 support; all nodes received 1.0 posterior probabilities from both 6-partition and 16-partition analyses.



Supplementary Figure S2 Alignment of stem-loop structures in the A+T rich regions of dictyopteran insects (RGC 1).