The Australian processionary caterpillar, *Ochrogaster lunifer* Herrich-Schäffer (Lepidoptera: Notodontidae), comprises cryptic species

Andrew Mather[1], Myron P. Zalucki[1], Julianne Farrell[1], Lynda E. Perkins[1], Lyn G. Cook[1]*

[1]*The University of Queensland, School of Biological Sciences, Brisbane, QLD, 4072, Australia*

*l.cook@uq.edu.au

Running Title

Ochrogaster lunifer is a cryptic species complex

Abstract

The bag shelter moth, *Ochrogaster lunifer* Herrich-Schäffer, 1855 (Thaumetopoeinae), is abundant and widespread throughout Australia where its larvae have been reported to feed mostly on *Acacia* and eucalypts. The larvae, known as processionary caterpillars, build silken nests on their host plants either on the ground at the base of the plant (*Acacia*) or above-ground on the trunk or among the canopy (*Acacia* and eucalypts). The caterpillars are medically important in that they shed tiny setae that can cause dermatitis and other health problems in humans and other mammals, including amnionitis and foetal loss in horses. Despite reports of behavioural, ecological and morphological differences between ground and canopy nesters, caterpillars of all nest types and hosts are currently considered to belong to one species. Here, we use DNA sequence data from the mitochondrial and nuclear genomes

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/aen.12410

of caterpillars taken from different nest types in eastern Australia to determine whether there is evidence for there being more than one species. We find significant genetic divergence between caterpillars from different nest types despite occurrence in sympatry at multiple sites, indicative of a lack of gene flow and the presence of at least two reproductively isolated species. Given the range of hosts and nest locations within hosts throughout Australia, further sampling is needed to determine just how many species there are under the current concept of *O. lunifer*.

Key words

Bag-shelter moth, *Teara contraria*, NUMT, *COI* pseudogene, Coalescence-based species delimitation, mPTP

INTRODUCTION

The bag-shelter moth, *Ochrogaster lunifer* Herrich-Schäffer, 1855, is a widespread and abundant moth native to Australia. It is commonly encountered in the larval form when caterpillars form single-line processions whilst moving between host trees or away from nests to overwinter then pupate in the soil. The gregarious larvae nest in silken “bags” in their host shrub or tree, or silk nests at the base of their host plant (mostly acacias and eucalypts, but also various other genera) (Fig. 1). The "bag" provides shelter to larvae, which in later instars, leave the nest at night to feed on leaves of their host. Their processionary behaviour, maintained by thigmotaxis (Steinbauer 2009), and laying of silk trails leads to branches and trunks acquiring a coating of silk. Their feeding can lead to defoliation of the host plant (Floater 1996) and, in outbreak years, feeding by larvae can have a significant ecological impact particularly in remnant vegetation patches (van Schagen *et al.* 1992). When complete defoliation occurs, the larvae move to a different tree (Mills 1951).

The location of the nest correlates with the oviposition site chosen by the adult female, with the nest enlarging as the larvae transition from neonates to later instars (Mills 1951; Floater 1996). Besides the location of oviposition, there are some differences in the ecology and morphology between above-ground and ground nesters. For example, the filamentous scales with which females cover their egg mass differs, with those of ground-nesters on *Acacia* being white (Floater 1996) whereas those on trunks of eucalypts and in canopies are generally golden (Perkins et al. unpubl.). Additionally, the size and colour of larvae and adults apparently differ between above-ground and ground nesters (Mills1951; Steinbauer 2018; Perkins et al. unpubl.) and, although both are found on *Acacia*, only above-ground nesters are found on eucalypts (*Eucalyptus*, *Corymbia*) (Floater 1996). Such differences between ground and above-ground nesters have led some authors (e.g., Mills 1951; Common 1990; Floater 1996; Floater & Zalucki 1999) to postulate that there might be more than a single species represented amongst the differing nest types, whereas others (e.g., Steinbauer 2018) have argued that nest location and colour polyphenism might be associated with ecology. Nevertheless, *Ochrogaster* is currently monotypic, with *O. lunifer* being the only recognised species (Australian Faunal Directory, 2018).

Given the suspicions that *O. lunifer* might represent a cryptic species complex (multiple species not originally recognised as distinct species on the basis of adult morphology), we here use DNA sequence data from larvae taken from ground and above-ground nests to test the species status of *O. lunifer*. If *O. lunifer* is a single biological species, we would expect that DNA sequences from both ground and above-ground nest types would be intermingled, and that any clustering of alleles would be by geographic location rather than by nest type. Strong geographic structuring is expected because adults of *O. lunifer* do not feed and live only a few days, and adult females lay only a single egg clutch (Floater & Zalucki 2000). Accurate knowledge of species status and identity is important in understanding the ecology of *Ochrogaster*, such as the equivocal results obtained by Steinbauer and Mitchell (2016) and Steinbauer (2018) with respect of nest location and host-use, and may have follow-on benefits for managing its populations: *Ochrogaster* has long been recognised as a cause of urticaria in humans (Southcott 1978; Floater 1996; Battisti *et al.* 2017), and for causing

miscarriages in horses (Cawell-Smith *et al.* 2012), due to tiny setae that are expelled when the caterpillar is disturbed (Perkins *et al.* 2016).

METHODS

Sample Collection, DNA extraction and PCR amplification

Larvae of *O. lunifer* were collected from nests using forceps or trapped on the trunk below their nest as they descended to pupate. A range of nest types, locations and host species were represented but all were from eastern Australia (Table 1) although the moth is widespread throughout Australia (Atlas of Living Australia, 2018). Samples were fixed immediately and stored whole in absolute ethanol until further processing. Only one specimen per nest was processed for DNA: when two samples are from the same GPS location, they are from nests on different plants.

Head capsules were removed from larvae, to avoid large quantities of gut contents and faecal material, and genomic DNA was extracted using a CTAB-Chloroform method, as per Lin *et al.* (2013). In the case of very small caterpillars, the entire body was processed for DNA extraction. The head or body was incubated in CTAB buffer with Proteinase K overnight at 55˚C, then removed and stored in 70% ethanol for future morphological analyses. Chloroform was then added to the CTAB mix, gently rocked for 15 mins, then centrifuged at high speed for 5 mins. Supernatant was removed and DNA was precipitated using 80% isopropanol. Samples were then centrifuged at high speed so that DNA formed a pellet, which was subsequently washed twice in 70% ethanol. The dried DNA pellet was resuspended in 100 µL of 10mM Tris-HCl pH8.0 and stored at -20˚C for future use.

Three independent gene regions were amplified using primers and conditions outlined in Table 2: the barcode region of the mitochondrial gene *COI* (Cytochrome c oxidase subunit I), and parts of the nuclear genes *EF1α* (Elongation factor 1-alpha) and *CAD* (Carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase). *EF1α* was

originally amplified using M44-1 and rcM53-2 but PCR failed for many samples. A new forward primer was designed specifically for *O. lunifer* from the few successful sequences using GENEIOUS R10 (www.geneious.com, Kearse *et al.* 2012). The CAD primers had been designed previously from alignments of sequences from Hemiptera, Coleoptera and Diptera in Genbank rather than specifically for *O. lunifer*. Successful PCR products were cleaned using Exonuclease I and Antarctic Phosphatase before being sequenced by Macrogen Inc. (Republic of Korea) using the Sanger method and the same primers used for the original PCR.

Sequences were edited in GENEIOUS and aligned using the GENEIOUS aligner. *COI* was checked for stop codons after translating with the invertebrate mitochondrial code and determining correct reading frame in GENEIOUS. Substitutions among inferred amino acids were compared using the polarity and hydrophobicity options in GENEIOUS. Base frequencies of all genes were determined and checked for bias among lineages (non-stationarity) using the Chi-square test implemented in PAUP* v4.0a (build 161) (Swofford 2002).

Coalescence-based species delimitation

We applied the Multi-rate Poisson Tree Processes method (mPTP) (Kapli *et al.* 2017) to a phylogram of our *COI* data (66 specimens) calculated using Tamura-Nei distances with Neighbor-Joining in GENEIOUS, and repeated the analysis with only one representative of each haplotype (as suggested by Kapli *et al.* 2017). The mPTP approach models transitions in branch lengths (substitutions) to identify putative species and allows independent rates (exponentially distributed) for each. In using simulated (Kapli *et al.* 2017) and empirical (Blair *et al.* 2017) datasets, mPTP has been found to be relatively robust compared with other coalescence methods, even when there is uneven sampling and/or differences between effective population sizes of the putative species.

The Barcode of Life database (BOLD) (Ratnasingham & Hebert 2007) holds about 58 *COI* sequences of *O. lunifer* from across Australia but all are from adults: they provide no information about nest type. Nevertheless, to assess the effect of sample size of our dataset,

we repeated the mPTP analyses incorporating the *COI* sequences of *O. lunifer* from BOLD, and with all members of the subfamily Thaumetopoeinae held in BOLD (Thaumetopoeinae dataset: 472 sequences, 121 from *O. lunifer*).

Phylogenetic analyses

Two concatenated datasets were made from the data for larvae with known nest type: 3GENE, which comprised all individuals for which all three genes had been obtained, and 2NUC, which comprised all individuals for which both nuclear genes had been obtained (i.e., no mtDNA included). Both datasets were analysed in a Bayesian framework with BEAST v1.8.4 (Drummond & Rambaut 2007). Each gene partition was assigned its own clock rate (strict) and substitution model (HKY for the nuclear genes, and HKY+invariants for *COI*). A Yule tree prior was applied because we were dealing with processes within species or among closely related putative species. The run comprised 10 million generations, with sampling every 1000, and the log was checked in TRACER v1.6 (Rambaut *et al.* 2014) to determine whether likelihoods had plateaued and that there had been sufficient sampling (ESS >200).

To assess congruence among gene partitions and to visualise uncertainty among phylogenetic trees, an unrooted network was calculated for both concatenated datasets using SPLITSTREE4 (Huson & Bryant 2006). Individual genes were compared using Neighbor-Joining phylograms calculated on Tamura-Nei distances in GENEIOUS and compared to determine if there was any supported (bootstrap > 70) conflict among genes.

RESULTS

There was no non-stationarity detected within *O. lunifer* for any gene, but there was in 3rd codon-positions of *COI* when other Thaumetopoeinae were included ($P \leq 10^{-8}$). For five individuals of *O. lunifer*, only polymorphic sequences of *COI* were obtained (Table 1) despite using two different primer pairs (LCO/HCO and LCO/LepR1) and these were excluded from further analyses. There was up to 6.8% divergence in *COI* among our samples of *O. lunifer*

and up to 9.2% when including those from BOLD. Specimens #1 and #4 were the most divergent in inferred amino acids (as translated in GENEIOUS using the invertebrate mitochondrial code), including a non-synonymous substitution (Glycine to Serine) that changed the polarity of that site relative to all other members of the subfamily.

Coalescence-based species delimitation

Analysis of the *COI* data from our specimens using mPTP identified four putative species (Fig. S1). These same four lineages were identified as putative species in mPTP analyses of all available *COI* sequences for *O. lunifer* (Fig. S2) and in the Thaumetopoeinae dataset (not shown). These latter mPTP analyses identified ten additional groups of *O. lunifer* as putative "species", along with several singletons.

Phylogenetic analysis

The 3GENE dataset comprised 2090 base pairs (618 bp *COI*, 657 bp *EF1α* and 816 bp *CAD*) with 94 variable sites. Most of the variable sites were in *COI* (72), with only 14 in *EF1α* and eight in *CAD*. Bayesian analysis of the 3GENE dataset found the same four clades identified as species in the mPTP analyses, each with posterior probabilities (PP) of 1.00. There was strong support (PP=1.00) for Group 1 being sister to the rest, and Group 2 being sister to Groups 3+4 (Fig. S3).

The SPLITSTREE graph of the 3GENE dataset showed potential conflict among partitions both within and among the four clades, with multiple parallel lines between main clusters (Fig. 2A). Phylograms of each gene showed supported conflict between *COI* and the nuclear genes (Fig. S4): in particular, individuals of *COI* Group 2 (specimens #1 and #4) were nested within Group 1 in analyses of both nuclear genes, and individuals of Group 3 were nested within Group 4.

There was no major conflict evidenced in the SPLITSTREE graph based on only nuclear loci, which grouped specimens mostly by nest type (ground nests on *Acacia* versus other nest types), with sequences from specimens of other nest types secondarily clustered by host

taxon (*Acacia* or eucalypt; Fig. 2B). This pattern was repeated in the Bayesian analysis of the nuclear loci (Fig. 3), with ground nesters on *Acacia* distinct from all others and other *Acacia*-nesters forming a clade amongst eucalypt-nesters albeit without support from posterior probabilities.

The four putative species identified in analyses of *COI* and the 3GENE dataset, and the two groups identified in analyses of the 2NUC dataset (ground *Acacia* and other), all co-occur with another at some sites (within metres of each other) and broadly overlap with each other at the continental scale (Fig. 4). In particular, ground-nesters on *Acacia* and eucalypt nesters co-occur near Gatton and Brisbane in southeast Queensland and in the Hunter Valley of NSW, and both also occur in Victoria (Fig. 4). Canopy-nesters on *Acacia* co-occur with ground-nesters on *Acacia* in the Hunter Valley (NSW).

DISCUSSION

Our aim was to test the hypothesis that *O. lunifer* is a single species. In order to interpret results and for our conclusions to be testable into the future, any assessment of species status needs to done within a framework of an explicit species concept. Given that *O. lunifer* is a sexually reproducing moth and that moths commonly use specific cues for mate attraction, we develop our arguments for its species status under a biological species concept. However, rather than conduct direct tests of attraction and reproductive compatibility, we use a lack of gene flow as indirect evidence for reproductive isolation.

Our results identify more than a single species within the current concept of *O. lunifer*. Coalescence species delimitation identified four species based on shifts in branching patterns of *COI*. However, species delineation methods using mtDNA, such as the mPTP analyses applied here, functionally identify population structure rather than species and any clades identified are best interpreted as needing further testing with other data (e.g., Kekkonen *et al.*, 2015). Deep coalescence in mtDNA can occur within populations that comprise admixed populations that have merged after being long separated geographically, or in species with

very high effective population sizes where haplotype diversity is maintained through deep time (e.g., Toon *et al.* 2016). Furthermore, mtDNA-only-based analyses provide little information about gene flow, which is important in interpreting whether populations are reproductively isolated and hence recognisable as species under the biological species concept. Nevertheless, mtDNA provides an independent locus from nuclear loci and congruence among clade membership provides additional confidence of reproductive isolation.

Nuclear DNA, being biparentally inherited, can provide direct evidence of gene flow and here it is the strongest evidence of reproductive isolation. Analyses of nuclear-only data (singly or concatenated) split the samples into two well-supported groups: ground nesters on *Acacia* (hereafter "*Acacia*-ground") and above-ground nesters on *Acacia* and eucalypt-nesters (hereafter "other") (Fig. 2B and Fig. S3). The split is not strictly by the nest position from which the larvae were collected, with three specimens from nest types that do not match those with which they cluster in phylogenetic analyses. Each of these might be from an atypical nest position, which we have observed occasionally in the field: canopy-nesters sometimes bivouac near the base of a host when preparing to procession away for pupation, and ground nesters sometimes form loose nests at branching points presumably by mistaking the axil as the ground when descending the host (Perkins et al. unpubl.). Specimen #10 was collected from a nest on the trunk of *A. stenophylla*. Typically, nests associated with *Acacia* are either relatively amorphous on the ground at the base of the plant (ground-nester, Fig. 1A) or formed of dense silk in the canopy among twigs and small branches (canopy-nester, Fig. 1B,C). The nest from which #10 was collected was amorphous and low on the trunk of the host *Acacia*: it may represent a ground nest that has formed a little above ground on the trunk. Specimens #6 and #8 were each collected from the base of a trunk of *Eucalyptus*, whereas all other nests on *Eucalyptus* and *Corymbia* are on trunks, branches or in the canopy. These two samples clustered with other specimens from these hosts.

Despite co-occurrence of *Acacia*-ground nesters and other nesters across multiple sites (sometimes within metres of each other) (Fig. 4), there is no evidence of recent gene flow

between them: alleles of both nuclear genes cluster based on *Acacia*-ground versus other, but not on geographic location. In contrast, moths of the same nest types separated by hundreds of kilometres share the same allele or have closely related alleles (Fig. 2B). This pattern is highly statistically unlikely (Rosenberg's $P_{(AB)} \leq 10^{-20}$; see Rosenberg 2007) if all belong to a single biological species. These results conflict with those of Steinbauer (2018) who found that captive individuals from canopy and ground nests in Victoria and southern NSW could mate and produce viable eggs. If there had been successful mating between moths from different nest types among our sampled populations, we would expect moths at the same location to be sharing alleles, but here they are completely different to the extent that there is no sharing of alleles. Furthermore, we included two populations sampled by Steinbauer (one canopy and one ground; N=3 and N=2 individuals respectively) and found no evidence of genetic mixing, with individuals from each clustering with specimens of the same nest type from far afield.

Although our assessment of gene flow and species boundaries needs to be extended to populations from other hosts and more broadly across Australia, the conflict with Steinbauer (2018) highlights the potential danger of lab-based mating trials within small enclosures. Normal behavioural or sensory preferences may be bypassed by the unusual conditions and lack of choice. Furthermore, mating alone might not be a good indicator of species status if post-zygotic reproductive isolation occurs. Eggs or larvae might not be fully viable, and the latter was not assessed by Steinbauer (2018).

Conflict between COI and nuclear genes

Our *COI* results conflict somewhat with those of Steinbauer and Mitchell (2016). We found no sharing of mtDNA haplotypes between *Acacia*-ground and other nest types but, similar to their study, the two nest types did not form reciprocally monophyletic groups. In our study, there was conflict in the number of groups and their membership depending on whether *COI* was included in the dataset or not. Some conflict between mitochondrial and nuclear genes might be expected because of their different modes of inheritance and population sizes leading to a difference in expected time to coalescence. Coalescence is usually shallower

(faster) for mtDNA than nuclear loci because it is haploid and transmitted only through the mother, in contrast to the biparentally inherited, diploid nuclear loci. The difference in coalescence rate means that mtDNA frequently differentiates animal species earlier in the speciation process than nuclear loci, but there can be variation depending on population size, mating systems, levels and direction of ongoing introgression, and sex-specific dispersal patterns (e.g., Hoelzer 1997; Hudson & Turelli 2003).

The distinction of two groups (*Acacia*-ground nesters and other nesters) was evident in analyses incorporating *COI*, with the exception of *Acacia*-ground-nest specimens #1 and #4 that formed a separate clade more closely related to the *Acacia*-canopy and eucalypt nesters than to other *Acacia*-ground-nesters (Fig. 2A, Figs S1-S3). There are several potential explanations for this conflict between nuclear and mtDNA loci. One possible explanation is that we amplified a pseudogene copy of *COI* from these two individuals (possibly a copy translocated to the nucleus: a nuclear mitochondrial copy, or NUMT): NUMTs of *COI* are common in insects (e.g., Sunnucks & Hales 1996; Richly & Leister 2004). The presence of NUMTs can sometimes lead to polymorphisms in *COI* sequences, which should be haploid, but here there was no evidence of a second copy of *COI* from specimens #1 and #4 (the electrophoresis traces were clean, with all peaks showing only a single base). However, there were more inferred amino acid changes including one that indicated a change in polarity relative to all other members of the subfamily. This sort of change is expected if random mutations are occurring in a defunct copy of a gene (such as a NUMT), but stop codons and indels are also expected over time and we found no evidence of these.

An alternative explanation for the conflict is that there has been an historical introgression event between *Acacia*-ground nesters and *Acacia*-canopy-nesters that led to the "capture" of a canopy-nester mtDNA lineage by a ground-nesting species, bringing that mtDNA lineage into the *Acacia*-ground nester nuclear background. Mitochondrial capture is well known in animals (e.g., Toews & Brelsford 2012), especially early in species divergence when there might be occasional introgression. In any case, such an introgression event must have taken place a long time ago because the *COI* lineage of #1 and #4 is now highly divergent from that

of *Acacia*-canopy-nesters and eucalypt-nesters: if it were recent, we would expect the haplotypes of #1 and #4 to be shared or nested among *COI* sequences from these nest types.

Species status of Ochrogaster lunifer

The lack of gene flow between *Acacia*-ground nesters and other nest types despite co-occurrence provides strong evidence that there are at least two species that have overlapping ranges in eastern Australia. Coalescence-based species delimitation using *COI* identified more putative species among the sequences available for *O. lunifer* in Genbank, but these results need to be treated with caution. For example, despite mPTP identifying the clade comprising #1 and #4 as a putative species in all analyses, the nuclear genes show that there has been recent gene flow between these and other *Acacia*-ground-nesting populations. This is more indicative of *Acacia*-ground nesters representing a single biological species consistent with their ecology.

There is some indication in our results that *Acacia*-canopy nesters, *Eucalyptus* nesters and *Corymbia* nesters might also be distinct species, given they form separate clades in analyses of nuclear DNA. However, there was no statistical support for these groups from posterior probabilities and too little variation in *CAD* to provide convincing resolution. Given that *O. lunifer* is widespread across the continent, is morphologically variable, and occurs on hosts not sampled during this study, there is a strong possibility that there are more than two cryptic species present. The presence of additional cryptic species could have confounded results of earlier studies, such as Steinbauer & Mitchell (2016) and Steinbauer (2018), which did not find clear differentiation between ground-nesters and canopy-nesters. For example, these two studies combined data for canopy nests from *Acacia* and other hosts, and did so similarly for ground-nesters.

Future taxonomic studies should use newly emerging methods, such as targeted gene capture or ddRADseq, that can obtain many nuclear markers to test patterns of gene flow, especially among co-occurring but ecologically different populations. Comprehensive molecular data from across the geographic and host range of *O. lunifer* could also provide a framework for

identifying the morphological characteristics of larvae and adults, particularly male genitalia, that might allow morphological identification of the putative species.

ACKNOWLEDGEMENTS

We would like to thank Martin Steinbauer for providing specimens from Victoria and southern NSW, and landowners for allowing collection of specimens from their property. Partial funding for the project was provided by Australian Research Council Linkage Project grant LP140100687 (*Assessing animal exposure to urticating caterpillar hairs and developing management strategies to reduce the consequence of foetal abortion in mares*).

REFERENCES

Atlas of Living Australia. 2018. Available from URL https://www.ala.org.au [Accessed 25 June 2018]

Australian Faunal Directory. 2018. Names List for *Ochrogaster* Herrich-Schäffer, [1855]. Available from URL https://biodiversity.org.au/afd/taxa/Ochrogaster/names [Accessed 25 June 2018]

Battisti A, Larsson S, Roques A. 2017. Processionary moths and associated urtication risk: Global change-driven effects. *Annual Review of Entomology*, **62**, 323-342.

BOLD 2018. Available from URL http://www.boldsystems.org/index.php/databases [accessed 13 April 2018].

Blair C & Bryson RW. 2017. Cryptic diversity and discordance in single-locus species delimitation methods within horned lizards (Phrynosomatidae: *Phrynosoma*). *Molecular Ecology Resources*, **17**, 1168-1182.

Cawdell-Smith AJ, Todhunter KH, Anderson ST, Perkins NR & Bryden WL. 2012. Equine amnionitis and fetal loss: Mare abortion following experimental exposure to processionary caterpillars (*Ochrogaster lunifer*). *Equine Veterinary Journal*, **44**, 282-288.

Cho S, Mitchell A, Regier JC, Mitter C, Poole RW, Friedlander TP & Zhao S. 1995. A highly conserved nuclear gene for low-level phylogenetics: *Elongation-Factor 1a* recovers morphology-based tree for Heliothine moths. *Molecular Biology and Evolution*, **12**, 650-656.

Common IFB. 1990. Moths of Australia. CSIRO Publishing, Melbourne.

Drummond AJ & Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.

Floater GJ. 1996. Life history comparisons of ground- and canopy-nesting populations of *Ochrogaster lunifer* Herrich-Schaffer (Lepidoptera: Thaumetopoeidae): Evidence for two species? *Australian Journal of Entomology*, **35**, 223-230.

Floater GJ & Zalucki MP. 1999. Life tables of the processionary caterpillar *Ochrogaster lunifer* Herrich-Schaffer (Lepidoptera: Thaumetopoeidae) at local and regional scales. *Australian Journal of Entomology*, **38**, 330-339.

Folmer O, Black M, Hoeh W, Lutz R & Vrijenhoek R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**, 294-299.

Hebert PDN, Penton EH, Burns JM, Janzen DH & Hallwachs W. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 14812-14817.

Hoelzer GA. 1997. Inferring phylogenies from mtDNA variation: Mitochondrial-gene trees versus nuclear-gene trees revisited. *Evolution*, **51**, 622-626.

Hudson RR & Turelli M. 2003. Stochasticity overrules the "three-times rule": Genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution*, **57**, 182-190.

Huson DH & Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254-267. Software available from www.splitstree.org

Kapli P, Lutteropp S, Zhang J, Kobert K, Pavlidis P, Stamatakis A & Flouri T. 2017. Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. *Bioinformatics*, **33**, 1630-1638.

Kass RE & Raftery AE. 1995. Bayes Factors. *Journal of the American Statistical Association*, **90**, 773-795.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P & Drummond A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647-1649.

Kekkonen M, Mutanen M, Kaila L, Nieminen M & Hebert PDN. 2015. Delineating species with DNA barcodes: A case of taxon dependent method performance in moths. *PLoS One*, **10**, 32.

Lin YP, Kondo T, Gullan P & Cook LG. 2013. Delimiting genera of scale insects: molecular and morphological evidence for synonymising *Taiwansaissetia* Tao, Wong and Chang with *Coccus* Linnaeus (Hemiptera: Coccoidea: Coccidae). *Systematic Entomology*, **38**, 249-264.

Mills MB. 1951. Bag shelter caterpillars and their habits. *Western Australian Naturalist*, **3**, 61-67.

Perkins LE, Zalucki MP, Perkins NR, Cawdell-Smith AJ, Todhunter KH, Bryden WL & Cribb BW. 2016. The urticating setae of *Ochrogaster lunifer*, an Australian processionary caterpillar of veterinary importance. *Medical and Veterinary Entomology*, **30**, 241-245.

Rambaut A, Suchard MA, Xie D & Drummond A. 2014. Tracer v1.6. Available from http://beast.bio.ed.ac.uk/Tracer.

Ratnasingham S & Hebert PDN. 2007. BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Molecular Ecology Notes*, **7**, 355-364.

Richly E & Leister D. 2004. NUMTs in sequenced eukaryotic genomes. *Molecular Biology and Evolution*, **21**, 1081–1084.

Rosenberg NA. 2007. Statistical tests for taxonomic distinctiveness from observations of monophyly. *Evolution*, **61**, 317-323.

Southcott RV. 1978. Lepidopterism in the Australian region. *Records of the Adelaide Children's Hospital*, **2**, 87–173.

Steinbauer MJ. 2009. Thigmotaxis maintains processions of late-instar caterpillars of *Ochrogaster lunifer*. *Physiological Entomology*, **34**, 345-349.

Steinbauer M & Mitchell A. 2016. Explaining phenotypic and behavioural variability in *Ochrogaster lunifer* (Lepidoptera: Notodontidae) from an evolutionary ecology perspective [version 1; not peer reviewed]. *F1000Research*, **5**, 351 (poster: 14th International Symposium on Insect-Plant Relationships 2011).

Steinbauer MJ. In press. Wing pattern polyphenism in two behavioural forms of *Ochrogaster lunifer* (Lepidoptera: Notodontidae). *Austral Entomology* https://doi.org/10.1111/aen.12329

Sunnucks P & Hales DF. 1996. Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). *Molecular Biology and Evolution*, **13**, 510-524.

Swofford DL. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Toews DPL & Brelsford A. 2012. The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, **21**, 3907-3930.

Toon A, Daglish, GJ, Ridley AW, Emery RN, Holloway JC & Walter GH. 2016. Random mating between two widely divergent mitochondrial lineages of *Cryptolestes ferrugineus* (Coleoptera: Laemophloeidae): A test of species limits in a phosphine-resistant stored product pest. *Journal of Economic Entomology*, **109**, 2221-2228.

van Schagen JJ, Hobbs RJ & Majer JD. 1992. Defoliation of trees in roadside corridors and remnant vegetation in the Western Australian wheatbelt. *Journal of the Royal Society of Western Australia*, **75**, 75-81.

SUPPORTING INFORMATION

Figure S1. Figure S1. Output of mPTP species delimitation using the phylogram derived from *COI* data. A. With all *COI* sequences generated for this study. B. With only one representative of each haplotype. Clades identified as species in mPTP are coloured in red and labelled species 1-4. Coloured bars next to terminals match those used in figures in the main text.

Figure S2. Species delimitation using mPTP on a phylogram derived from *COI* sequence data from *Ochrogaster lunifer* in our dataset and those available in BOLD. Clades identified as species are coloured red. Coloured bars next to some clades indicate groups referred to in the main text.

Figure S3. Maximum Clade Credibility tree from Bayesian Inference analysis in BEAST using all three genes (*COI*, *CAD* and *EF1α*). The topology is similar to that for *COI* alone. *Acacia*-ground nesters are in maroon and red, above-ground nesters on *Acacia* are in orange, and eucalypt-nesters in pale blue and dark blue. Numbers above relevant nodes represent posterior probabilities and the scale bar represents substitutions per site. Specimen codes are detailed in Table 1.

Figure S4. Neighbour-joining dendrograms using DNA sequence data from *COI* (A), *CAD* (B) and *EF1α* (C). Numbers above branches represent bootstrap values ≥70 derived from 1000 pseudoreplicates. Terminals are colour coded as per clades in the main text.

Tables

Table 1. Details for the samples used during this study. The names of the nest type follow Perkins *et al.* (2016). In some figures, "#" in names is replaced by "hash". A Genbank accession number indicates that a sequence was obtained for that gene for that specimen, and a P indicates that only polymorphic traces were obtained. Collectors: JF (Julianne Farrell), LP (Lynda Perkins), MS (Martin Steinbauer).

Sample Name	**Nest Type**	**Host**	**Sample**	***COI*[c]**	***EF1α***	***CAD***	**Latitude**	**Longitude**	**Collector**	**Collection Date**
Ground *Acacia*										
G1516TA	Ground	*A. holosericea*	head	X	X	X	-27.55	152.34	LP	08.iii.2016
G1516TD	Ground	*A. salicina*	head	X	X	-	-27.55	152.34	LP	01.iv.2016

G1516TJ	Ground	*A. fimbriata*	head	X	X	X	-27.55	152.33	LP	08.iii.2016
G1516TK	Ground	*A. fimbriata*	head	X	X	X	-27.55	152.33	LP	08.iii.2016
G1516TL	Ground	*A. aneura*	head	X	-	X	-27.55	152.34	LP	08.iii.2016
G1516TO	Ground	*A. concurrens*	head	X	X	X	-27.55	152.34	LP	08.iii.2016
G1516TP	Ground	*A. pendula*	head	X	X	X	-27.55	152.34	LP	08.iii.2016
G1516TQ	Ground	*A. concurrens*	head	X	-	X	-27.55	152.33	LP	08.iii.2016
G1516TU	Ground	*A. podalyriifolia*	head	X	X	-	-27.55	152.34	LP	08.iii.2016
G1516TV	Ground	*A. holosericea*	head	X	-	X	-27.55	152.34	LP	08.iii.2016

G1516TW	Ground	*A. fimbriata*	head	X	X	X	-27.55	152.34	LP	08.iii.2016
G1516TX	Ground	*A. concurrens*	head	X	X	X	-27.55	152.33	LP	08.iii.2016
G1516TY	Ground	*A. concurrens*	head	X	X	X	-27.55	152.33	LP	08.iii.2016
G1516TZ	Ground	*A. salicina*	head	X	X	X	-27.55	152.34	LP	15.iii.2016
S1516#1	Ground	*A. concurrens*	head	X	X	X	-27.37	152.89	LP	23.iii.2016
S1516#2	Ground	*A. concurrens*	head	X	-	X	-27.37	152.89	LP	23.iii.2016
S1516#3	Ground	*A. concurrens*	head	X	X	X	-27.37	152.89	LP	23.iii.2016
D1516#3	Ground	*A. leiocalyx*	head	X	X	X	-27.34	152.90	LP	14.iii.2016

D1516#4	Ground	*A. concurrens*	head	X	X	X	-27.34	152.90	LP	23.iii.2016
D1516#6	Ground	*A. concurrens*	head	X	X	X	-27.34	152.90	LP	14.iii.2016
D1516#8	Ground	*A. concurrens*	head	X	X	X	-27.34	152.90	LP	14.iii.2016
D1516#11	Ground	*A. concurrens*	head	X	X	X	-27.34	152.90	LP	14.iii.2016
MS#25	Ground	*A. pycnantha*	head	X	X	X	-36.21	143.97	MS	18.ii.2010
MS#26	Ground	*A. pycnantha*	head	P	X	X	-36.21	143.97	MS	18.ii.2010
#1	Ground	*Acacia* sp.	head	X	X	X	-32.11	150.98	JF	16.ii.2016
#4	Ground	*Acacia* sp.	whole	X	X	X	-32.12	150.95	JF	16.ii.2016

#10	Trunk (loose)	*A. stenophylla*	whole	X	X	X	-27.34	151.55	JF	05.ii.2016
#11	Ground	*A. harpophylla*	head	X	X	X	-27.56	152.55	JF	11.iii.2016
#24	Ground	*Acacia* sp.	head	X	X	-	-36.60	143.25	JF	24.iv.2016
#30	Ground	*Acacia* sp.	head	X	X	-	-36.60	143.25	JF	24.iv.2016
Above ground *Acacia*										
MS#33	Canopy	*A. pendula*	head	X	X	X	-35.51	145.18	MS	18.ii.2010
MS#34	Canopy	*A. pendula*	head	X	X	X	-35.51	145.18	MS	18.ii.2010

MS#35	Canopy	*A. pendula*	head	X	X	X	-35.51	145.18	MS	18.ii.2010
#2	Canopy	*Acacia* sp.	head	X	X	X	-32.11	150.98	JF	16.ii.2016
#9	Canopy	*Acacia* sp.	whole	X	X	-	-27.34	151.55	JF	05.ii.2016
#15	Canopy	*A. pendula*	head	X	X	X	-32.34	150.58	JF	24.iii.2016
#16	Canopy	*A. pendula*	head	X	X	X	-32.33	150.82	JF	24.iii.2016
#17	Canopy	*A. pendula*	head	X	X	X	-32.47	150.86	JF	25.iii.2016
#19	Canopy	*Acacia* sp.	head	X	X	-	-29.00	150.02	JF	22.iv.2016
#20	Canopy	*Acacia* sp.	head	P	X	X	-27.39	151.62	JF	17.iv.2016

#21	Canopy	*Acacia* sp.	head	X	X	X	-27.32	151.51	JF	17.iv.2016
#22	Canopy	*Acacia* sp.	head	X	X	X	-29.21	150.01	JF	22.iv.2016
#23	Canopy	*Acacia* sp.	head	X	X	-	-27.36	151.60	JF	17.iv.2016
#26	Canopy	*A. pendula*	head	P	X	X	-31.31	148.27	JF	14.iii.2016
#27	Canopy	*Acacia* sp.	head	X	X	-	-29.00	150.02	JF	22.iv.2016
#33	Canopy	*Acacia* sp.	head	X	X	-	-27.34	151.55	JF	22.iv.2016
Eucalypt										
G1516Ta	Tree-hugger	*C. tessellaris*	head	X	X	X	-27.56	152.34	LP	03.iii.2016

G1516Tb	Tree-hugger	*C. tessellaris*	head	X	X	X	-27.56	152.34	LP	18.iv.2016
G1516Tc	Tree-hugger	*C. tessellaris*	head	X	X	X	-27.56	152.34	LP	08.iii.2016
G1516Td	Tree-hugger	*C. tessellaris*	head	X	X	X	-27.56	152.34	LP	08.iii.2016
G1516Te	Tree-hugger	*C. tessellaris*	head	X	X	X	-27.56	152.34	LP	08.iii.2016
G1516Tg	Tree-hugger	*C. tessellaris*	head	X	X	X	-27.56	152.34	LP	08.iii.2016
G1516Th-1	Tree-hugger	*C. tessellaris*	head	X	X	X	-27.56	152.34	LP	08.iii.2016
G1516Th-2	Tree-hugger	*C. tessellaris*	head	X	X	X	-27.56	152.34	LP	08.iii.2016
G1516To-1	Tree-hugger	*C. tessellaris*	head	X	X	-	-27.56	152.34	LP	09.vi.2016

G1516Tn	Tree-hugger	*C. tessellaris*	head	X	X	X	-27.56	152.34	LP	04.v.2016
G1516#01676	Tree-hugger	*C. tessellaris*	head	X	X	-	-27.55	152.34	LP	04.v.2016
G1516#00261	Tree-hugger	*C. tessellaris*	head	X	-	-	-27.55	152.34	LP	04.v.2016
#01686	Tree-hugger	*C. tessellaris*	head	X	X	X	-27.55	152.34	LP	09.vi.2016
#0220	Tree-hugger	*C. tessellaris*	head	X	X	X	-27.55	152.33	LP	09.vi.2016
S1516Ta	Tree-hugger	*C. tessellaris*	head	X	X	-	-27.36	152.89	LP	16.v.2016
#3	Trunk	eucalypt	whole	X	X	X	-32.12	150.95	JF	16.v.2016
#5	Trunk	*Eucalyptus* sp.	whole	X	X	X	-32.47	150.97	JF	17.ii.2016

#6	Ground	*E. sideroxylon*	whole	X	X	X	-32.47	150.97	JF	17.ii.2016
#7	Tree-hugger	*Eucalyptus* sp.	whole	X	X	X	-32.47	150.97	JF	17.ii.2016
#8	Ground	*Eucalyptus* sp.	whole	X	X	X	-32.47	150.97	JF	17.ii.2016
#25	Canopy	*E. orgadophila*	head	X	X	X	-27.70	151.66	JF	29.iii.2016
#28	Trunk	*C. tessellaris*	head	P	X	-	-27.56	152.33	JF	15.iii.2016
#29	Canopy	eucalypt	head	P	X	X	-27.70	151.66	JF	29.iii.2016
#31	Trunk	*C. tessellaris*	head	X	-	-	-27.56	152.33	JF	12.iv.2016
#34	Canopy	eucalypt	head	-	X	-	-27.70	151.66	JF	16.iv.2016

#35	Canopy	eucalypt	head	X	X	-	-27.70	151.66	JF	16.iv.2016

Table 2. Primers and PCR conditions used to amplify the three genes.

Gene/primer name	Sequence 5'-3'	Source	Annealing temperature
Cytochrome c oxidase subunit I (*COI*)			
LCO	GGTCAACAAATCATAAAGATATTGG	1	
HCO	TAAACTTCAGGGTGACCAAAAAATCA	1	
LepR1	TAAACTTCTGGATGTCCAAAAAATCA	2	45-51˚C
Elongation factor 1-alpha (*EF1α*)			
M44-1	GCTGAGCG(CT)GA(GA)CGTGGTATCAC	3	
rcM53-2	GCAATGTGRGCIGTGTGGCA	3	
EF1a_Olun_F	ATAGAGATTTCATCAAGAACATGATC	4	54˚C
Carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase (*CAD*)			
CAD_F1	GATAAYTGYATTACWGTTTGTAAYATGGAAAA	4	
CAD_R1	TCWGCWGCAACTGTATCTATTTGTTT	4	50˚C

1 Folmer *et al*. (1994)

2 Hebert *et al*. (2004)

3 Cho *et al.* (1995)

4 This study

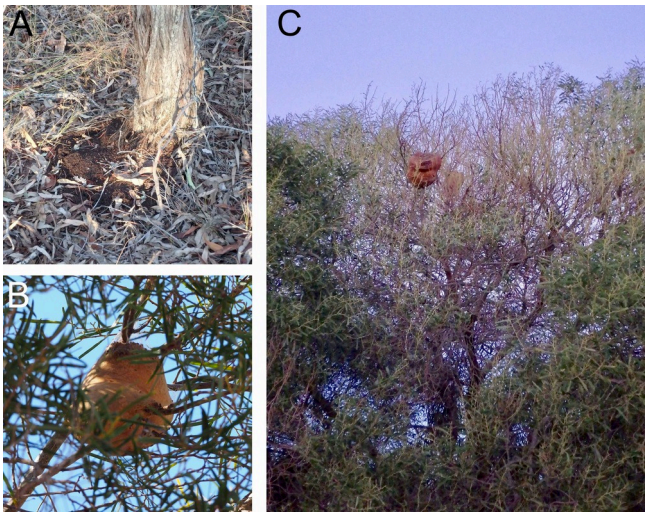
Figure legends

Figure 1. Nests of processionary caterpillars of *Ochrogaster lunifer*. A) Ground nest at base of trunk of *Acacia*. B) Canopy nest in crown of *Acacia*. C) Canopy nest in crown of *Acacia* showing defoliation from feeding in the vicinity of the nest.

Figure 2. Output from SPLITSTREE showing alternative pathways in the network. A) Combined analysis of *COI* (mitochondrial) and the two nuclear genes (*CAD* and *EF1*α). B) Incorporating only the two nuclear genes. Distinct clades, nest type and host use are colour-coded: *Acacia*-ground nesters (maroon and red), above-ground nesters on *Acacia* (orange) and eucalypt-nesters (pale blue and dark blue).

Figure 3. Maximum Clade Credibility tree from Bayesian Inference analysis in BEAST of two nuclear genes (*CAD* and *EF1*α) from *Ochrogaster lunifer*. *Acacia*-ground nesters (maroon and red) are clearly separated from caterpillars of other nest types (above-ground nesters on *Acacia*, in orange, and eucalypt-nesters in pale blue and dark blue). Numbers above relevant nodes represent posterior probabilities and the scale bar represents substitutions per site.

Figure 4. Map showing location of caterpillars of *Ochrogaster lunifer* sampled from nests included in this study. Points are colour-coded as in other figures: *Acacia*-ground nesters (maroon and red), above-ground nesters on *Acacia* (orange) and eucalypt-nesters (pale blue and dark blue). Points are presented at varying sizes so that different colours at the same location can be seen.



AEN_12410_F1.jpg