

# MOLECULAR ECOLOGY

## Speciation processes in putative island endemic sister bat species: false impressions from mitochondrial DNA and microsatellite data

Journal:	<i>Molecular Ecology</i>
Manuscript ID	MEC-15-0787.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	<p>Kuo, Hao-Chih; Queen Mary, University of London, School of Biological and Chemical Sciences            Chen, Shiang-Fan; National Taipei University, Center for General Education            Fang, Yin-Ping; National Chiayi University, Department of Biological Resources            Cotton, James; Queen Mary, University of London, School of Biological and Chemical Sciences            Parker, Joe; Queen Mary, University of London, School of Biological and Chemical Sciences            Csorba, Gábor; Hungarian Natural History Museum, Department of Zoology            Lim, Burton; Royal Ontario Museum,            Eger, Judith; Royal Ontario Museum, Natural History;            Chen, Chia-Hong; Shei-Pa National Park Headquarters,            Chou, Cheng-Han; Endemic Species Research Institute, Division of Zoology            Rossiter, Stephen; Queen Mary, University of London, School of Biological and Chemical Sciences;</p>
Keywords:	Non-allopatric divergence, introgressive hybridization, gene flow, Taiwan, Murina

1 **Speciation processes in putative island endemic sister bat species: false**  
2 **impressions from mitochondrial DNA and microsatellite data**

3

4 HAO-CHIH KUO,<sup>1</sup> SHIANG-FAN CHEN,<sup>2</sup> YIN-PING FANG,<sup>3</sup> JAMES A COTTON,<sup>1\*\*</sup> JOE  
5 D PARKER,<sup>1</sup> GÁBOR CSORBA,<sup>4</sup> BURTON K LIM,<sup>5</sup> JUDITH L EGER,<sup>5</sup> CHIA-HONG  
6 CHEN,<sup>6</sup> CHENG-HAN CHOU<sup>7</sup> and STEPHEN J ROSSITER<sup>1\*</sup>

7 <sup>1</sup>*School of Biological and Chemical Sciences, Queen Mary University of London, London E1*  
8 *4NS, UK; <sup>2</sup>Center for General Education, National Taipei University, New Taipei City 23741,*  
9 *Taiwan; <sup>3</sup>Department of Biological Resources, National Chiayi University, Chiayi City 60004,*  
10 *Taiwan; <sup>4</sup>Department of Zoology, Hungarian Natural History Museum, 1088 Budapest, Hungary;*  
11 *<sup>5</sup>Department of Natural History, Royal Ontario Museum, Toronto, ON M5S 2C6, Canada; <sup>6</sup>Shei-*  
12 *Pa National Park Headquarters, Miaoli County 36443, Taiwan; <sup>7</sup>Division of Zoology, Endemic*  
13 *Species Research Institute, Nantou County 552, Taiwan*

14

15 *Keywords:* Non-allopatric divergence, introgressive hybridization, gene flow, Taiwan, *Murina*

16

17 \*Correspondence: Stephen Rossiter

18 Fax: +44 (0) 207 882 7732;

19 E-mail: s.j.rossiter@qmul.ac.uk

20

21 \*\*Current address: Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus,  
22 Cambridgeshire CB10 1SA, UK

23 **Abstract**

24 Cases of geographically restricted co-occurring sister taxa are rare and may point to potential  
25 divergence with gene flow. The two bat species *Murina gracilis* and *M. recondita* are both  
26 endemic to Taiwan and are putative sister species. To test for non-allopatric divergence and gene  
27 flow in these taxa, we generated sequences using Sanger and Next Generation Sequencing, and  
28 combined these with microsatellite data for coalescent-based analyses. MtDNA phylogenies  
29 supported the reciprocally monophyletic sister relationship between *M. gracilis* and *M. recondita*,  
30 however, clustering of microsatellite genotypes revealed several cases of species admixture  
31 suggesting possible introgression. Sequencing of microsatellite flanking regions revealed that  
32 admixture signatures stemmed from microsatellite allele homoplasy rather than recent  
33 introgressive hybridization, and also uncovered an unexpected sister relationship between *M.*  
34 *recondita* and the continental species *M. eleryi*, to the exclusion of *M. gracilis*. To dissect the  
35 basis of these conflicts between ncDNA and mtDNA, we analysed sequences from 10  
36 anonymous ncDNA loci with \*BEAST and isolation-with-migration (IM) and found two distinct  
37 clades of *M. eleryi*, one of which was sister to *M. recondita*. We conclude that Taiwan was  
38 colonized by the ancestor of *M. gracilis* first, followed by the ancestor of *M. recondita* after a  
39 period of allopatric divergence. After colonization, the mitochondrial genome of *M. recondita*  
40 was replaced by that of the resident *M. gracilis*. This study illustrates how apparent signatures of  
41 sympatric divergence can arise from complex histories of allopatric divergence, colonization and  
42 hybridization, thus highlighting the need for rigorous analyses to distinguish between such  
43 scenarios.

44

## 45 Introduction

46 The expectation that drift and selection will more easily establish genetic differentiation  
47 between populations with little or no gene flow has led to the overwhelming view that most cases  
48 of speciation must involve a period of geographical isolation (e.g. Barraclough & Vogler 2000;  
49 Turelli *et al.* 2001). Despite this, theoretical models have proposed routes by which parapatric  
50 and sympatric speciation (hereafter collectively referred to as ‘non-allopatric speciation’) can  
51 proceed (Slatkin 1982; Rice 1984; Dieckmann & Doebeli 1999; Gavrilets & Waxman 2002) and  
52 these studies have been complemented by several empirical examples that suggest non-allopatric  
53 speciation could be more common in nature than was previously assumed (reviewed in Via 2001).

54  
55 In evaluating potential cases of non-allopatric speciation, several researchers have  
56 advocated using biogeographical information. Lynch (1989) proposed that range overlap and  
57 relative range sizes of sister species or supra-specific taxa could be used to infer geographical  
58 modes of speciation, with substantial overlap pointing to sympatric speciation, and little overlap  
59 indicating allopatric speciation. Such reasoning has been used to argue for greater frequencies of  
60 sympatric speciation in nature than previously thought (e.g. Mattern & McLennan 2000), but has  
61 also received criticism given that a key assumption - that geographical ranges of natural  
62 organisms are constant through time - is probably seldom true (reviewed in Losos & Glor 2003;  
63 Coyne 2007).

64  
65 Extending the logic of studying range overlap, Coyne and Price (2000) emphasized the  
66 value of looking for ‘sister species’ (species that are each other’s closest relatives) that are both

67 vagile yet restricted geographically to the same small area, such as an oceanic island (also see  
68 White 1978). Under these conditions, they suggested that sympatric rather than allopatric  
69 divergence is a likely explanation for such overlapping ranges. Coyne and Price (2000) searched  
70 for oceanic islands (area 0.8 to 3500 km<sup>2</sup>) on which at least one endemic bird species occurred.  
71 By assessing whether or not each taxon's sister species co-existed on the same island or  
72 archipelago, consistent with *in situ* divergence, they found no clear evidence supporting non-  
73 allopatric speciation in the species pairs studied. Kisel and Barraclough (2010) extended this  
74 approach to diverse taxa and found that potential cases of *in situ* speciation in highly mobile  
75 animals such as moths and mammals were associated with larger island areas (140,000 km<sup>2</sup> and  
76 420,000 km<sup>2</sup>, respectively), while smaller areas were associated with possible examples from  
77 plants and less mobile animals (snails and lizards). Thus there appears little evidence for  
78 sympatric speciation when applying Coyne and Price's (2000) criteria for island-dwelling taxa, at  
79 least for highly mobile animals.

80  
81 A common criticism of focusing on geography in speciation biology is that it detracts  
82 from the underlying gene dynamics (see Hey 2006; Fitzpatrick *et al.* 2009). Indeed it is implicitly  
83 assumed, but rarely tested, that non-allopatric divergence typically occurs in the face of gene  
84 flow, especially where reproductive isolation is thought to arise via ecological shifts (e.g.  
85 Kingston & Rossiter 2004; Jackson 2008; Forbes *et al.* 2009) rather than by post-zygotic  
86 mechanisms (e.g. Husband & Sabara 2004). Recently several empirical studies have applied  
87 newly developed isolation-with-migration (IM) models (Hey & Nielsen 2004; Hey 2010b) to  
88 demonstrate historical gene flow among closely related taxa, attributable to either non-allopatric

89 divergence (e.g. Hey 2006; Nadachowska & Babik 2009) or, alternatively, allopatric divergence  
90 followed by secondary contact (e.g. Llopart *et al.* 2002; Geraldès *et al.* 2006). Generally,  
91 however, most such studies have been unable to date the occurrence of gene flow; simulations  
92 show that IM methods cannot provide reliable estimates of the timing of historical gene flow  
93 (Strasburg & Rieseberg 2011). Consequently, to distinguish between cases of taxa that have  
94 formed sympatrically with gene flow versus those that evolved while geographically isolated but  
95 which have subsequently undergone secondary gene flow, additional information can be  
96 informative, including knowledge of the broader patterns of genetic affiliations with other  
97 populations and taxa, and the potential for historical range shifts in light of climatic fluctuations.

98  
99 While interpretations of speciation based on either range overlap or isolation-with-  
100 migration each have their respective shortcomings, these approaches are complementary and  
101 together might offer a more powerful means of testing for non-allopatric divergence. Here we  
102 combine these methods to study the origin of two newly discovered bat species *Murina gracilis*  
103 and *M. recondita* on Taiwan (Kuo *et al.* 2009), which phylogenetic reconstructions using  
104 mitochondrial DNA (mtDNA) indicate are sister taxa (Kuo 2004, 2013). Such cases of  
105 geographically restricted co-occurring sister species are exceptionally rare in bats; in their review,  
106 Kisel and Barraclough (2010) identified two congeneric bat species on New Zealand as the best  
107 potential example. Given that Taiwan is a much smaller island (36,000 km<sup>2</sup>) with respect to the  
108 expected mobility of flying mammals, *M. gracilis* and *M. recondita* meet Coyne and Price's  
109 (2000) criteria and thus offer an excellent opportunity to test for non-allopatric divergence.  
110 Moreover, the fact that these taxa show contrasting altitudinal preferences, with the former

111 tending to occur at higher elevations (Kuo 2004, 2013), further suggests the potential for  
112 ecological speciation linked to niche differentiation. We generated datasets of nuclear and  
113 mitochondrial markers for both species and applied IM-based modeling approaches alongside  
114 other population genetic analyses for gene flow at different scales. Since a detailed dissection of  
115 the divergence process requires a thorough understanding of the relationship between the two  
116 focal taxa, we also generated and analysed sequences for the continental congeneric (and also  
117 recently described) species *M. eleryi* (Furey *et al.* 2009). We hypothesized that given their close  
118 relationship and endemic status on Taiwan, *M. gracilis* and *M. recondita* likely diverged *in situ*  
119 on this island and, therefore, would show evidence of a sister relationship across all markers, as  
120 well as evidence of divergence in the face of gene flow.

## 121 **Materials and methods**

122 We focused on three newly described species of tube-nosed bat from East Asia. *Murina gracilis*  
123 and *M. recondita* are considered endemic to Taiwan (Kuo *et al.* 2009), while *M. eleryi* has been  
124 recorded in Southern China, Vietnam and Laos (Furey *et al.* 2009; Eger & Lim 2011). Sibling  
125 relationships among these three taxa have been inferred from both morphometric comparisons  
126 and mtDNA phylogenetic reconstructions, with the latter unambiguously supporting a  
127 relationship of ((*M. gracilis*, *M. recondita*), *M. eleryi*) (see Systematic notes on the focal species,  
128 Supporting information). Although little is known about their respective ecologies, *M. gracilis*  
129 occurs at elevations of >1500 metres above-sea-level (asl), whereas *M. recondita* and *M. eleryi*  
130 both occur at lower elevations of <1500 metres asl (Kuo *et al.* 2014; Table S2, Supporting  
131 information).

132

### 133 *Assessment of contemporary gene flow between M. gracilis and M. recondita*

134 To test for contemporary gene flow between *M. gracilis* and *M. recondita* we screened  
135 populations for evidence of genetic admixture using the Bayesian clustering method implemented  
136 in STRUCTURE 2.3.2 (Pritchard *et al.* 2000). Genotypes of 106 *M. gracilis* and 144 *M.*  
137 *recondita* at 14 microsatellite loci (A4, A9, A10, A104, A109, A118, A122, B5, B9, B114, B121,  
138 B124, D110, and D117; see Kuo *et al.* 2013) were examined under a model that assumed  
139 independent allele frequencies among genetic clusters and allowed for mixed ancestries among  
140 individuals. We ran ten replicates of Markov chain Monte Carlo (MCMC) for between two and  
141 ten clusters (K) with each MCMC comprising 0.75 million iterations for sampling, and the same  
142 number for burn-in. To assess consistency across replicates we used the CLUMPP procedure

143 (Jakobsson & Rosenberg 2007) and visualized the resulting plots using DISTRUCT (Rosenberg  
144 2004). Plots under the K value that best fitted the data, as justified according to the rationale  
145 proposed by Pritchard *et al.* (2000), were inspected for signatures of genetic admixture.

146  
147 For each individual bat we estimated the proportion of its genetic composition, as  
148 measured by the ancestry coefficient  $q$  (Pritchard *et al.* 2000), that could be assigned to its own  
149 species, summing across multiple clusters where relevant. A few individuals of *M. recondita*  
150 showed  $<0.9$  assignment to their own species (see Results), potentially reflecting genetic  
151 introgression from *M. gracilis*. Alternatively, such signatures might arise through allele size  
152 homoplasy (SH), which can occur in microsatellites due to their hypervariable nature (Estoup *et*  
153 *al.* 2002). To identify those loci driving signatures of admixture, we repeated our STRUCTURE  
154 and CLUMPP analyses under the best-fitting K but excluded each locus one-by-one. Three  
155 replicate runs were conducted for each pruning scheme and, for each bat showing mixed ancestry,  
156 we recalculated  $q$  assigned into its own species under each pruning scheme. Loci that when  
157 pruned led to an increase in  $q$  (compared with values based on the full dataset) were regarded as  
158 likely candidates for introgression or SH. To distinguish between these scenarios, we amplified  
159 and sequenced the corresponding flanking regions of those loci contributing most to admixture,  
160 in order to check for parallel signatures of introgression that would be expected under tight  
161 linkage to the simple sequence tandem repeat units. We reasoned that apparent introgression at  
162 STRs but not at the adjacent flankers would be best explained by SH.

163

164           Using the approach we amplified and sequenced the flanking regions of two microsatellite  
165 loci (A9 and A122) in selected individuals of *M. gracilis* and *M. recondita*, including those with  
166 possible mixed species ancestries. In total, ten bats from each of the two species were selected  
167 (indicated by arrows in Fig. 1) plus one *M. eleryi* (HNHM 2007.28.2; see Table S2) for  
168 comparative purposes. We designed and paired the primers A9FRL (5'-GCA ATT TCA TTG  
169 TGT CCC TTG-3') and A9FRR (5'-GTC ATA GTT CTA GTC TCC CAG ATC C-3') for A9,  
170 and A122FRL (5'-CAT TCT ATC TGC CTA CCT TGA CA-3') and A122FRR (5'-GGC CTT  
171 CTC ACT AGG CAC AG-3') for A122. PCR cocktails of 15  $\mu$ L, contained 0.2  $\mu$ M each primer,  
172 2.0  $\mu$ L of template, and 7.5  $\mu$ L of the provided master mix of the Qiagen Type-it Microsatellite  
173 PCR Kit (QIAGEN). Reactions were performed on a BIO-RAD C1000 Thermal Cycler (BIO-  
174 RAD Laboratories) with the following thermal profile: 95  $^{\circ}$ C 5 min; 35 cycles of 95  $^{\circ}$ C 30 sec, 59  
175  $^{\circ}$ C 90 sec, 72  $^{\circ}$ C 30 sec; 60  $^{\circ}$ C 30 min. Sanger sequencing using primers A9FRR and A122FRR  
176 was undertaken by Eurofins MWG Operon (Ebersberg, Germany) on an ABI 3730 DNA  
177 Analyzer (Applied Biosystems).

178  
179           Sequences were trimmed and the chromatograms inspected by eye for double peaks,  
180 indicative of heterozygous sites. For reads with multiple heterozygous sites, we used the  
181 Bayesian statistical method in PHASE 2.1.1 (Stephens *et al.* 2001; Stephens & Scheet 2005) to  
182 infer haplotypes. PHASE analysis was applied separately to *M. gracilis* and *M. recondita*, using  
183 models that both allow or disallow intra-genic recombination. To facilitate phasing, known  
184 haplotypes obtained from 454 sequencing were used as references (see next section). Under each  
185 recombination model, we performed three replicated runs, each with 10,000 iterations (10

186 iterations per sample) following a burn-in of 1000 iterations. Haplotypes inferred by PHASE with  
187 a posterior probability of  $\geq 0.6$  were accepted (see Garrick *et al.* 2010). To investigate the  
188 segregation of haplotypes in relation to species identity, for each locus we constructed median-  
189 joining (MJ) network using the software NETWORK 4.6.1.1 (Fluxus Technology Ltd. 2010) in  
190 which the epsilon value was heuristically set as 10.

191

### 192 *Divergence within the M. gracilis complex revealed by mtDNA*

193 To obtain a more comprehensive picture of the divergence processes among *M. gracilis*, *M.*  
194 *recondita* and their relatives, we further included 12 additional individuals of *M. eleryi* collected  
195 from Southern China (n=7) and Central Vietnam (n=5) (see Table S2 for details). Genomic DNA  
196 of these bats was extracted using the DNeasy Blood & Tissue Kit (Qiagen).

197

198 We first investigated the divergence processes based on mtDNA by expanding our earlier  
199 published dataset (Kuo *et al.* 2014; see Table 1) by Sanger sequencing the partial cytochrome *b*  
200 (Cyt-*b*) and cytochrome c oxidase subunit 1 (COI) genes of these additional *M. eleryi* samples,  
201 which we also supplemented with published sequences (Francis *et al.* 2010; see Table 1). To  
202 infer the order of divergence among *M. gracilis*, *M. recondita* and *M. eleryi* taxa we used Heled  
203 and Drummond's (2010) Bayesian method implemented in \*BEAST. Assuming the absence of  
204 post-split gene flow between divergent taxa, \*BEAST can assess the stochastic process of  
205 coalescence by explicitly modeling effective population size along the course of divergence, and  
206 has also been shown to perform very well in estimating the order and timing of divergence events,  
207 even when applied to a single locus (see Drummond *et al.* 2012). The two mitochondrial genes

208 were concatenated and partitioned, into the first and second codon positions together (partition  
209 CP<sub>12</sub>) and the third codon position (CP<sub>3</sub>). For CP<sub>12</sub> and CP<sub>3</sub>, we used the nucleotide substitution  
210 models HKY+I (HKY, Hasegawa *et al.* 1985) and GTR+I+G (GTR, Tavaré 1986), respectively,  
211 selected based on the Akaike Information Criterion in jModelTest 2.1.3 (Posada 2008). The  
212 demography along the course of divergence was modeled by constant-sized populations between  
213 sequential splits, namely the piecewise-constant (PC) model. We used the extinction-free  
214 constant-rate birth process (Yule 1924) as a prior for divergence, and, under a strict clock model,  
215 we calibrated the root of the taxon tree with a lognormal prior of  $\log(\text{mean}) = 0.96$  and  
216  $\log(\text{standard deviation}) = 0.15$ , corresponding to an mtDNA-based estimate of the split time  
217 between *M. eleryi* and the common ancestor of *M. gracilis* and *M. recondita* as 2.62 (95% CI:  
218 1.86-3.48) million years ago (Ma) (see Systematic notes on the focal species, Supporting  
219 information). We performed two replicate MCMC runs in BEAST 1.7.4 (Drummond *et al.* 2012),  
220 each with 100 million iterations (8,000 iterations per sample) and the first 10% of runs discarded  
221 as burn-in. We examined the runs in Tracer 1.5 (Rambaut & Drummond 2007) as well as with  
222 the 'starbeast\_demog\_log' python script in the biopy 0.1.7 package (available from:  
223 <http://code.google.com/p/biopy/>); these two runs gave consistent results and so were combined to  
224 give an effective sample size (ESS) of > 2,600 per parameter.

225

### 226 *Nuclear markers for investigating divergence within the M. gracilis complex*

227 For insights into the divergence process in the nuclear genome among the focal taxa, we  
228 developed a set of ten anonymous autosomal markers based on microsatellite flanking regions  
229 (A4, A9, A104, A109, A118, A122, B5, B114, B124 and D117). For each locus we designed new

230 primer pairs based on clone sequences to amplify the flanker but to exclude the microsatellite  
231 motif (Table S3, Supporting information). Primers pairs designed for A9 and A122 amplified  
232 fragments that overlap with those described in the earlier section.

233

234 For *M. gracilis* and *M. recondita*, we obtained sequences for each of the ten loci by high-  
235 throughput 454-Pyrosequencing (Margulies *et al.* 2005). Given that both species show strong  
236 geographic substructure within Taiwan (Kuo *et al.* 2014), which would violate assumptions of  
237 coalescent-based analyses, sequences were only generated for 15 individuals of each taxon from  
238 part of their respective ranges (Table 1). To sort sequences, we incorporated unique 3-bp tags at  
239 the 5' end of the synthesized forward primer for each locus. For *M. gracilis*, we used the tags  
240 ACA, ACG, ACT, AGA, AGC, AGT, ATA, ATC, ATG, CGA, CGC, CGT, CTA, CTC and  
241 CTG, while for *M. recondita* we used CAT, GTA, GTC, GTG, GAC, GAG, GAT, GCA, GCG,  
242 GCT, TAC, TAG, TAT, TCA and TCG. PCR products of each locus were visualized on a gel and  
243 pooled for 454-Pyrosequencing at Mission Biotech (Taipei, Taiwan) on a 454 GS Junior System  
244 (Roche/454 Life Sciences).

245

246 We used the software CLC Genomics Workbench 4.9 (CLC bio, Denmark) to filter and  
247 sort 454-reads as follows. First, we used original clone sequences as references on which 454  
248 reads were mapped with a coverage parameter of 0.6 and a similarity parameter of 0.9. Second,  
249 we sorted mapped sequences into individual bats on the basis of the 3-bp tag plus 2-3 sites of the  
250 5' end of the forward primer. Sorted sequences were aligned using MUSCLE (Edgar 2004) in  
251 MEGA 5 (Tamura *et al.* 2011). For each individual, 50 reads were trimmed and aligned (in cases

252 where fewer reads were available, <50 were used), and, assuming that error rates per nucleotide  
253 were lower than correct base calls (Galan *et al.* 2010), we inferred the haplotypes of every bat for  
254 each of its two chromosomes. We adopted the following conservative criteria for acceptance of  
255 haplotypes for downstream genetic analyses: (1) for each individual, only inferred haplotypes  
256 with frequencies of  $\geq 5$  were accepted such that the occurrence of two haplotypes  $\geq 5$  signified a  
257 heterozygote, and (2) each individual was considered homozygous for a locus when only one  
258 haplotype was detected with a frequency of  $\geq 10$ . Bats that did not meet these criteria, because  
259 only one haplotype was detected at a frequency of  $\geq 5$  but  $< 10$ , were considered to be potentially  
260 heterozygous but with missing data for one of the two chromosomes.

261  
262 We also sequenced the same ten loci in *M. eleryi* using Sanger sequencing (Table 1), and  
263 confirmed in a subset of individuals of *M. gracilis* and *M. recondita* that both sequencing  
264 methods gave consistent results. PCR products were sequenced in both directions at Eurofins  
265 MWG Operon. In four of the ten loci, indels (1-9 bp) were observed, with individuals  
266 heterozygous for such indels showing characteristic double peaks in downstream stretches on  
267 their chromatograms. In these individuals, we were able to deduce the nucleotide sequences of  
268 each haplotype based on the rationale that the superposed signatures would correspond to two  
269 nucleotides that differed by a length equal to that of the indel (Flot *et al.* 2006). Where two  
270 haplotypes could not be inferred in this way we used PHASE (see earlier section), treating the  
271 two clades of *M. eleryi* that were identified and are hereafter referred to as *M. eleryi* 1 and 2 (see  
272 Results).

273

274 Final alignments of each of the ten nuclear markers contained samples of *M. gracilis*, *M.*  
275 *recondita*, *M. eleryi* 1 and *M. eleryi* 2. Excluding indels, we assessed the genetic variability of  
276 each taxon at each locus based on the number of segregating sites, Watterson's (1975) theta value  
277 and nucleotide diversity, all calculated in DnaSP 5.10.01 (Librado & Rozas 2009). We also used  
278 NETWORK to construct (epsilon = 10) locus-wise MJ genealogies for samples of all four taxa in  
279 order to gain information on genetic differences among them. Polzin and Daneshmand's (2003)  
280 maximum parsimony algorithm was used to remove unnecessary links in the network for one  
281 locus (A122) for which complex connections were seen.

282  
283 Prior to subsequent coalescent-based analyses, for each locus we tested for recombination  
284 using Hudson and Kaplan's (1985) four-gamete test assuming an infinite-site model of evolution,  
285 as implemented in the software IMGc (Woerner *et al.* 2007), in each case retaining the largest  
286 recombination-free block of sequences. This approach was not applied to locus A122, due to the  
287 detection of a tri-allelic segregating site that suggested this locus did not evolve in an infinite-site  
288 manner and might experience a faster mutation rate (also corroborated by analyses of genetic  
289 variability in the four focal taxa; see Results). Throughout this study, the HKY model was used  
290 for nucleotide substitutions for A122.

291  
292 *Divergence among taxa within the M. gracilis complex revealed by ncDNA*  
293 The order of divergence among focal taxa was inferred in \*BEAST, using the recombination-free  
294 sequences from ten nuclear loci, and with a JC nucleotide substitution model (Jukes & Cantor  
295 1969) applied to each except for A122. For demographic analysis throughout divergence, we

296 used the PC model (see Selection of a demographic model for the ncDNA \*BEAST analysis,  
297 Supporting information for details of model selection). For calibration of multi-locus coalescence  
298 events, we specified for each locus a strict clock with the clock rate sampled from the following  
299 lognormal prior:  $\log(\text{mean}) = -6.12$ ,  $\log(\text{standard deviation}) = 0.58$ . This prior was set for a mean  
300 estimate of the neutral mutation rate of mammalian nuclear genomes as 0.0022 substitutions per  
301 site per million years (Kumar & Subramanian 2002) while allowing rates to vary among loci to  
302 an order of magnitude. Two replicate \*BEAST runs were combined to give an ESS of  $> 200$  per  
303 parameter. Finally, a generation time for the focal taxa as two years was used to calculate the  
304 effective population sizes from the estimated demographic parameters.

305  
306 Using the phylogeny estimated in \*BEAST: (((*M. recondita*, *M. eleryi* 1), *M. eleryi* 2), *M.*  
307 *gracilis*), we then inferred the divergence process among these taxa in further detail, and  
308 specifically to estimate gene flow among them, we used isolation-with-migration (IM) models  
309 implemented in IMA2-8.26.11 (Hey & Nielsen 2007; Hey 2010b). An infinite site model was  
310 adopted for each of the ten loci analysed except for A122. Truncated uniform priors for entry  
311 demographic parameters of IMA2 were set following the author's recommendations (Hey 2011);  
312 specifically we set upper bounds of priors for the composite parameter of the split time (with the  
313 flag -t), the rescaled migration rate in either direction (-m), and the population mutation rate for  
314 each of the extant/ancestral taxa (-q) based on the genetic variability of *M. eleryi* 1, which had  
315 highest genetic variability of the four species (see Results). The geometric mean of Watterson's  
316 (1975) theta values over loci - an estimate for the population mutation rate - was 0.0055 per  
317 nucleotide site, corresponding to a value of 1.1 per locus given a geometric mean of sequence

318 lengths over the ten loci as 196. We set -t2 and -q5 for two and five times of the above estimates,  
319 respectively, and for post-split migrations we set -m2 for a moderately high upper bound for the  
320 2NM value, since  $2NM = q \times m/2 = 1.1$  under the set 'm' value. In addition, we explored a  
321 different m prior, modeled by an exponential distribution with a mean value of 0.5 (with flags -j8  
322 -m0.5), which circumvents truncation of the marginal estimate by an upper bound and is justified  
323 under the expectation of low migration between diverging taxa (Hey 2010b).

324

325 We also applied IM models, with identical prior settings to those described above, to  
326 pairwise datasets of the four focal taxa. Divergence times estimated independently for pairwise  
327 taxa could inform the splitting order (Hey 2010a) and thus corroborate the taxon phylogeny  
328 inferred by \*BEAST. Metropolis-coupling MCMC (MC<sup>3</sup>) for all above four- and two-taxon IM  
329 runs was conducted with a geometric heating scheme for 60 chains (heating parameters -ha0.95 -  
330 hb0.8), with one million iterations discarded as burn-in followed by five million iterations for  
331 sampling (100 iterations per sample). Three (for four-taxon analyses) or two (for two-taxon  
332 analyses) replicate runs were performed under each prior setting for assessment of convergence.  
333 Given that replicate runs appeared to converge, we combined results under each specific setting  
334 using the 'L mode' function of IMA2, and ran likelihood ratio tests (LRT) to assess evidence of  
335 post-split gene flow between focal taxa, following Nielsen and Wakeley (2001). Finally, to  
336 convert entry demographic parameters into interpretable scales, we applied a mutation rate of  
337 0.0022 per nucleotide site per million years (Kumar & Subramanian 2002), a geometric mean of  
338 sequence lengths of 196 bp, and a generation time of two years.

339

340 **Results**341 *Assessment of contemporary gene flow between M. gracilis and M. recondita*

342 Clustering analyses of 14 microsatellite loci yielded two different configurations at  $K=3$ , and  
343 also at  $K=4$  (Fig. 1). In each of these cases, lower ‘penalized log-likelihood values’ (Pritchard *et*  
344 *al.* 2000) were obtained for a configuration with an empty cluster that contained no substantial  
345 fraction of any individual bat of either species. By excluding runs with empty clusters (Guillot  
346 2008) the penalized log-likelihood was highest at  $K=4$  in which four *M. recondita* individuals  
347 (labeled as RA, RB, RC, and RD in Fig. 1) showed >10% of their genetic composition appearing  
348 to originate from *M. gracilis*. All other bats of both species were assigned almost exclusively to  
349 genetic clusters of their own species (each with a sum of ancestry coefficients  $q$  for its own  
350 species > 0.9).

351 Repeated STRUCTURE analyses with each locus sequentially removed revealed that  
352 apparent genetic admixture among some *M. gracilis* and *M. recondita* individuals was mainly  
353 attributable to loci A9 and A122 (Fig. 1; Table S4, Supporting information). To determine if  
354 these signatures stemmed from introgression, we sequenced the flanking regions of A9 and A122  
355 in selected bats (shown by arrows in Fig. 1) and constructed median-joining (MJ) networks. In  
356 the network based on A9, haplotypes of *M. gracilis* and *M. recondita* clustered by species,  
357 however, *M. recondita* showed an unexpected affiliation with *M. eleryi* (Fig. 2). Similarly, for the  
358 A122 network, all *M. recondita* haplotypes clustered together with the exception of one that  
359 grouped with *M. eleryi* (Fig. 2). It follows that apparent signatures of genetic admixture between  
360 *M. gracilis* and *M. recondita* seen in multi-locus microsatellite genotypes were not supported by  
361 analyses of these microsatellites’ flanker sequences, suggesting they reflect microsatellite allele

362 size homoplasy rather than true introgression. At the same time, however, we discovered an  
363 unexpectedly close relationship between *M. recondita* and *M. eleryi* that could be due to  
364 introgression. Based on this finding, we expanded sampling of *M. eleryi* in subsequent analyses  
365 to dissect the divergence history among all these *Murina* species.

366

#### 367 *Divergence within the M. gracilis complex revealed by mtDNA*

368 Using \*BEAST, the concatenated mitochondrial sequences of *Cyt-b* and COI (Table 1) recovered  
369 a mtDNA genealogy (not shown) with four maximally supported monophyletic groups (posterior  
370 probability = 1) representing *M. gracilis*, *M. recondita*, *M. eleryi* from Southern China (*M. eleryi*  
371 1) and *M. eleryi* from Central Vietnam (*M. eleryi* 2). Among these lineages, *M. gracilis* and *M.*  
372 *recondita* grouped together with maximal support, as did the two *M. eleryi* lineages. These  
373 relationships were also recovered in the inferred species tree (posterior probabilities > 0.99; see  
374 Fig. 3a). These two split events had median estimates of 0.79 (95% CI: 0.31-1.29) and 0.43 (95%  
375 CI: 0.10-0.77) Ma, respectively (Fig. 3a). Effective population sizes for extant and ancestral taxa  
376 ranged from 51,000 (*M. eleryi* 1) to 250,000 (*M. recondita*) as shown in Figure 3a.

377

#### 378 *ncDNA structure within the M. gracilis complex*

379 454-Pyrosequencing of autosomal microsatellite flanker sequences in larger sample sets of *M.*  
380 *gracilis* and *M. recondita* generated 73,395 reads, which were assembled to yield haplotypes for  
381 all ten target loci from both chromosomes in all but one individual of *M. gracilis* and 75% of  
382 individuals of *M. recondita*. Sanger sequencing of orthologous sequences in *M. eleryi* yielded  
383 haplotypes from both chromosomes in 70% of individuals (see Table 1).

384

385 Locus-specific MJ networks supported pilot sequencing of two loci, confirming that *M.*  
386 *recondita* ncDNA sequences were typically more related to those of *M. eleryi* than to those of *M.*  
387 *gracilis* (Fig. S1, Supporting information). For three of ten loci, *M. recondita* shared haplotypes  
388 with the two *M. eleryi* lineages but not with *M. gracilis*, and in cases where *M. recondita* shared  
389 haplotypes with *M. gracilis*, these haplotypes were also shared by the two *M. eleryi* lineages. At  
390 most loci, *M. eleryi* 1 and *M. eleryi* 2 had contrasting haplotype composition, indicative of  
391 divergence. Locus-wise measures of genetic variability for the four taxa are shown in Table 2. *M.*  
392 *recondita* had a lower level of genetic variability than all the others, particularly reflected by its  
393 lower  $\theta_S$  and  $\pi$  values at most of the ten markers. Four-gamete tests suggested potential intra-  
394 locus recombination at loci A4, A104, A109 and B124, for which relevant sections of alignments  
395 (for A4 and A109) or individual sequences (for A104 and B124) were removed using IMgc  
396 (shown in Fig. S1).

397

#### 398 *Divergence in M. gracilis complex revealed by ncDNA*

399 Supporting the haplotype networks, the species tree reconstructed using \*BEAST recovered a  
400 monophyletic clade containing *M. recondita* and both *M. eleryi* taxa (posterior probability = 1),  
401 which was estimated to have diverged from *M. gracilis* around 2.06 (95%CI: 1.13-3.19) Ma (Fig.  
402 3b). Within the (*M. recondita* + *M. eleryi*) clade, *M. recondita* formed a moderately well  
403 supported clade with *M. eleryi* 1 (posterior probability = 0.79) to the exclusion of *M. eleryi* 2.  
404 The split between *M. eleryi* 2 and the common ancestor of *M. recondita* and *M. eleryi* 1 was  
405 dated to 0.66 (95%CI: 0.28-1.14) Ma, while divergence between the latter two taxa was dated to

406 0.42 (95%CI: 0.16-0.77) Ma. Effective population sizes of lineages under the PC model ranged  
407 from 59,000 (*M. recondita*) to 632,000 (common ancestor of *M. recondita* and the two *M. eleryi*  
408 taxa) (see Fig. 3b).

409  
410 To aid descriptions, hereafter we use the abbreviations 4PM2 and 4PME0.5 for four-taxon  
411 isolation-with-migration (IM) analyses specifying a truncated uniform prior (upper bound = 2)  
412 and an exponential prior (mean = 0.5), respectively. For focal demographic parameters estimated  
413 with the 4PM2 prior, we show summary statistics in Figure 4 and present marginal distributions  
414 in Figure S2 (Supporting information). Summary statistics of these demographic parameters from  
415 both 4PM2 and 4PME0.5 analyses are also presented in Table S5 (Supporting information). The  
416 marginal distribution for the post-split migration parameter,  $m$ , was more compressed toward  
417 zero with 4PME0.5 than with 4PM2 prior, leading to a smaller estimate from 4PME0.5 analysis  
418 (a trend also reflected in the population migration rate,  $2NM$ ; see Table S5); this was true for all  
419  $m$  parameters along the course of divergence although such compressions in marginal  
420 distributions were less apparent for those with peaks at zero. For all other demographic  
421 parameters including times of split and effective population sizes, similar estimates were obtained  
422 regardless of the prior used (Table S5).

423  
424 From both 4PM2 and 4PME0.5 analyses, modal estimates for times of split indicated  
425 divergence between *M. recondita* and *M. eleryi* 1 (T1), between the common ancestor of these  
426 two taxa and *M. eleryi* 2 (T2), and between the common ancestor of three taxa and *M. gracilis*  
427 (T3), at 1.2-1.4, 1.8-1.9, and around 2.3 Ma, respectively. Nevertheless, confidence sets for these

428 parameters were rather broad (see Fig. 4 for values from 4PM2 analysis); this was especially true  
429 for T1 and T3 for which flat marginal distributions were obtained (see Fig. S2a for curves from  
430 4PM2 analysis). Among a total of 18 m parameters, seven had non-zero modes under 4PM2: bi-  
431 directional migration between *M. recondita* and *M. eleryi* 1, and migration from *M. recondita* to  
432 *M. eleryi* 2, from *M. eleryi* 2 to *M. eleryi* 1, from *M. eleryi* 2 to the common ancestor of *M.*  
433 *recondita* and *M. eleryi* 1, from the common ancestor of *M. recondita* and *M. eleryi* 1 to *M.*  
434 *gracilis*, and from the common ancestor of *M. recondita* and both *M. eleryi* taxa to *M. gracilis*  
435 (Fig. S2b). Of these seven, only the former four also had non-zero modes under 4PME0.5. Based  
436 on either m or 2NM, LRTs were significant for migrations from *M. recondita* to *M. eleryi* 1 and  
437 from *M. eleryi* 2 to *M. eleryi* 1 under both priors except for migration from *M. recondita* to *M.*  
438 *eleryi* 1 based on m under 4PME0.5. Non-significant results were obtained for all other LRTs  
439 (see Table S5 for a summary of LRTs based on 2NM). Under the two priors used, effective  
440 population sizes for the extant taxa had modal estimates ranged from 27,000/33,000 (*M.*  
441 *recondita*) to 322,000/348,000 (*M. eleryi* 1) or 319,000/349,000 (*M. eleryi* 2) (Fig. 4; Table S5).  
442 For effective population sizes of ancestral populations, flat marginal distributions (Fig. S2c), and  
443 thus broad confidence intervals (Table S5), were obtained, suggesting little information from the  
444 data on these parameters.

445  
446 As in the four-taxon IM analyses, we used the abbreviations 2PM2 and 2PME0.5 for two-  
447 taxon IM analyses specifying a truncated uniform prior (upper bound = 2) and an exponential  
448 prior (mean = 0.5), respectively. Summary statistics for all six pairs of taxa analysed under 2PM2  
449 prior are shown in Figure 5 while their marginal distributions are presented in Figure S3

450 (Supporting information). A full list of summary statistics for all two-taxon analyses under the  
451 two priors used is given in Table S6 (Supporting information). Again, the marginal distribution  
452 for each  $m$  parameter was more compressed toward zero under 2PME0.5 than under 2PM2 prior,  
453 leaving smaller  $m$  estimates in the former analysis (see Table S6, Supporting information). For  
454 parameters of split times and effective population sizes, on the other hand, IM estimates were  
455 insensitive to the prior used (Table S6).

456

457 From two-taxon IM analyses, comparing modal estimates across datasets suggested that  
458 the order of splits among taxa was consistent with the results from ncDNA-based \*BEAST; thus  
459 *M. recondita* and *M. eleryi* 1 split most recently (1.1-1.2 Ma), then *M. eleryi* 2 and either *M.*  
460 *recondita* or *M. eleryi* 1 (1.4-1.5 Ma), and finally the most ancient split was between *M. gracilis*  
461 and either of the two *M. eleryi* lineages (1.5-1.7 Ma) (see Fig. 5 for values from 2PM2 analyses).  
462 Inconsistent with this inferred splitting order, modal estimates from two-taxon IM analyses  
463 suggested an unexpected recent divergence between *M. gracilis* and *M. recondita* (1.2-1.3 Ma).  
464 However, under either of the two priors used, the marginal distribution for the split time of *M.*  
465 *gracilis* and *M. recondita* was the flattest one among the six pairwise analyses (see Fig. S3a for  
466 curves from 2PM2 analyses), implying greater uncertainty underlying this estimate. When *M.*  
467 *gracilis* was paired with either of the other taxa, modal estimates of  $m$  were zero for both  
468 directions. In contrast, modal estimates of  $m$  for the remaining pairwise combinations of taxa  
469 were non-zero for both directions (Fig. S3b). Based on either  $m$  or 2NM, LRTs were significant  
470 for bi-directional migrations between *M. recondita* and *M. eleryi* 1 and migration from *M. eleryi*  
471 2 to *M. recondita* under both priors except for migration from *M. recondita* to *M. eleryi* 1 based

472 on  $m$  under 2PME0.5. LRTs were also significant for bi-directional migrations between *M. eleryi*  
473 1 and *M. eleryi* 2 based on 2NM under 2PME0.5 (Table S6). Two-taxon IM analyses gave  
474 effective population size estimates for *M. gracilis*, *M. recondita*, and *M. eleryi* 2 broadly  
475 consistent with those from four-taxon analyses. In contrast, the estimates for *M. eleryi* 1 were  
476 higher than obtained from four-taxon analyses (Table S6). As with the four-taxon models, two-  
477 taxon analyses gave diffuse marginal distributions for effective population sizes of ancestral  
478 populations (Fig. S3c), suggesting limited information about these parameters was available from  
479 the data.

480

For Review Only

## 481 Discussion

482 Cases of geographical confinement of sister taxa to islands are rare, and have been proposed as  
483 candidate systems in which non-allopatric speciation processes could have occurred (White 1978;  
484 Coyne & Price 2000; Coyne 2007). Here we examined the process of divergence between the  
485 tube-nosed bats *Murina gracilis* and *M. recondita*, both of which are endemic to Taiwan and  
486 which, based on phylogenetic reconstructions of mtDNA, appear to be sister species (Kuo 2004,  
487 2013). To determine whether these taxa did indeed form in the face of migration, consistent with  
488 non-allopatric speciation, we tested for gene flow at different temporal scales using a variety of  
489 molecular markers and statistical approaches. Tests for contemporary gene flow based on ncDNA  
490 uncovered an unexpected sister relationship between *M. recondita* and a continental species *M.*  
491 *eleryi*, to the exclusion of *M. gracilis*. The demographic process and putative mechanisms  
492 responsible for this mtDNA-ncDNA conflict are discussed below.

493

### 494 *Absence of recent introgressive hybridization between M. gracilis and M. recondita*

495 Although, for the most part, microsatellite-based clustering suggested little, if any, contemporary  
496 gene flow between *M. gracilis* and *M. recondita* (250 individual bats in total; Fig. 1), a small  
497 number of individuals of *M. recondita* did appear to show mixed ancestry with *M. gracilis* (Fig. 1  
498 and Table S4). Such cases of apparent genetic admixture based on microsatellite clustering are  
499 not unusual, and are typically attributed to introgression or the retention of ancestral  
500 polymorphism (e.g. Muir & Schlötterer 2005; Berthier *et al.* 2006; Randi 2008; Brown *et al.*  
501 2010; Bogdanowicz *et al.* 2012). In our study, however, sequencing and analyses of the  
502 respective *M. recondita* microsatellite flanking regions - which are expected to evolve more

503 slowly than their adjacent microsatellite motifs (Estoup *et al.* 2002) - did not support admixture.  
504 Instead, in networks based on these flanking regions, the two putative sister taxa were  
505 reciprocally monophyletic with respect to each other (Fig. 2). It follows that apparent admixture  
506 signatures based on microsatellite genotyping appear to stem from allele size homoplasies, a  
507 well-described phenomenon that is nonetheless infrequently tested for and described in studies of  
508 population structure (exceptions include Adams *et al.* 2004; Rossiter *et al.* 2007).

509

#### 510 *Divergence between M. gracilis and M. recondita*

511 Further clustering analyses of ten microsatellite flanking sequences, in which both focal taxa  
512 were included along with the mainland congener *M. eleryi*, recovered an unexpected grouping of  
513 the latter with *M. recondita*. Specifically, phylogenetic reconstruction performed in \*BEAST - an  
514 approach that simultaneously reconstructs the gene tree(s) and evaluates the support of different  
515 species trees in generating such a gene tree (or gene trees) through stochastic drift (Heled &  
516 Drummond 2010; Drummond *et al.* 2012) - indicated that *M. recondita* formed a monophyletic  
517 clade with *M. eleryi* 1 from Southern mainland China, which in turn formed a clade with *M.*  
518 *eleryi* 2 from Central Vietnam to the exclusion of *M. gracilis* (Fig. 3b). Additional supporting  
519 evidence for this history of divergence came from ncDNA-based two-taxon isolation-with-  
520 migration (IM) analyses (Fig. 5; Fig. S3a) by which split times of divergent taxa were estimated  
521 while accounting for the potential influences of post-split gene flow (Hey & Nielsen 2004; Hey  
522 & Nielsen 2007). It is important to recognize that these estimated split times had broad  
523 confidence intervals, indicative that the data contains limited information about these parameters.  
524 Further uncertainty in these estimates arises from the fact that they were solely informed by a

525 mean multi-locus mutation rate. Despite uncertainty about these split times, however, it is  
526 noteworthy that the IM and \*BEAST analyses of microsatellite flanking sequences both  
527 suggested the same order of divergence; thus the mitochondrial and nuclear genomes harbored  
528 conflicting signals about the divergence process of the focal taxa.

529  
530 By superimposing the mtDNA genealogy for the four focal taxa on the splitting order  
531 obtained by the ncDNA, we were able to infer a scenario of independent incursions of the  
532 Taiwanese species. *M. gracilis* appears to have colonized first, followed by *M. recondita* with  
533 massive mtDNA introgression from the former to the latter. The clustering of range-wide mtDNA  
534 haplotypes of both Taiwanese species to the exclusion of those from *M. eleryi* implies that the  
535 mitochondrial genomes of *M. recondita* have been completely replaced by those of *M. gracilis*.  
536 Meanwhile, the present reciprocally monophyletic relationship between the two Taiwanese  
537 species indicates that any introgression has since ceased, probably around 0.3-1.3 (median: 0.8)  
538 Ma (Fig. 3a: split between *M. gracilis* and *M. recondita*). In contrast, neither the two-taxon nor  
539 four-taxon IM analyses detected post-split nuclear gene flow between *M. gracilis* and *M.*  
540 *recondita* (Fig. 4 and Fig. 5, respectively).

541  
542 Like all species of their genus, *M. gracilis* and *M. recondita* show a suite of eco-  
543 morphological adaptations for living in the forest interior, including broad wings and very large  
544 call bandwidths for detecting arthropods in dense clutter (Kingston *et al.* 1999; Kingston *et al.*  
545 2003; Schmieider *et al.* 2010). At the same time, however, such traits are associated with reduced  
546 gene flow (Struebig *et al.* 2011) and are likely to constrain the ability of these bats to move

547 across more open environments, including water bodies such as the Taiwan Strait. It follows that  
548 the two inferred incursions will almost certainly have occurred during glacial periods when the  
549 drop in sea-level exposed the continental shelf connecting Taiwan and mainland Asia (Vorisi  
550 2000). Consequently rather than diverging non-allopatrically on Taiwan, as we initially  
551 hypothesized based on their restricted island distributions, *M. gracilis* and *M. recondita* appear to  
552 have speciated during a period of geographical isolation, with their co-distribution arising via  
553 secondary contact. Similar histories of double incursions for pairs of congeneric Taiwanese taxa  
554 have been inferred in rodents (genera *Apodemus* and *Niviventer*; Yu 1995) and grass lizards  
555 (genus *Takydromus*; Lue & Lin 2008), both of which showed a paraphyletic relationships among  
556 island taxa with respect to continental forms.

557

558 Several scenarios may account for the observed massive introgression of mtDNA, yet  
559 little or no introgression of ncDNA, from *M. gracilis* to *M. recondita*. First, mitochondrial  
560 replacement can arise due to positive selection (e.g. reviewed in Melo-Ferreira *et al.* 2014).  
561 Despite finding no evidence of positive selection at mtDNA, we cannot rule this out (Tests of  
562 positive selection, Supporting information), especially given that mitochondrial protein-coding  
563 genes are subject to genetic hitch-hiking due to a lack of recombination that characterizes  
564 mitogenomes (Galtier *et al.* 2009). Alternatively, contrasting levels of mtDNA and ncDNA  
565 introgression can also reflect asymmetries between species in the extent to which they undergo  
566 assortative mating, or experience sex-biased reductions in hybrid fitness (Chan & Levin 2005;  
567 Mallet 2005). Perhaps the simplest explanation for the observed discordance between mtDNA  
568 and ncDNA markers lies in the fact that the two Taiwanese species would have experienced

569 different demographic conditions when they colonized Taiwan, as expected given the inferred  
570 scenario of two temporally distinct incursions. Specifically, it is probable that during the  
571 incursion of Taiwan by *M. recondita*, this colonizing taxon will have been in a phase of relative  
572 demographic growth compared to the earlier arriving *M. gracilis*. Simulations by Currat *et al.*  
573 (2008) show that under such conditions, introgression occurs from resident to the colonizing  
574 population. While massive mtDNA introgression from *M. gracilis* to *M. recondita* could have  
575 occurred via such ‘allele surfing’, a lack of parallel ncDNA introgression could be attributed to  
576 male-biased gene flow among conspecific populations of the latter species (Tests for sex-biased  
577 dispersal in *Murina gracilis* and *M. recondita*, Supporting information), so that *M. gracilis*  
578 nuclear alleles would have faced competition from those of conspecific populations, and thus  
579 reduced their chance of fixation. Indeed, evidence for the influence of sex-biased dispersal on  
580 differential rates of introgression among markers with different modes of inheritance has been  
581 reported in a wide range of taxa (Petit & Excoffier 2009). Further studies are needed to  
582 disentangle the contributions of these potential mechanisms between *M. gracilis* and *M.*  
583 *recondita*.

584

#### 585 *Introgression in tube-nosed bats and other bat species*

586 Introgression was not only inferred for *M. gracilis* to *M. recondita*. Our four-taxon IM analyses  
587 also suggested post-split nuclear gene flow from *M. recondita* to the continental *M. eleryi*, as well  
588 as between divergent clades of *M. eleryi* (Fig. 4). We therefore speculate that levels of mtDNA  
589 introgression might be even higher in these cases, based again on insights and expectations from  
590 simulations (Currat *et al.* 2008). In the former of these cases, more sampling of *M. eleryi*

591 populations close to Taiwan is needed to test for mtDNA. In the latter case, on the other hand,  
592 complete mitochondrial replacement between the two *M. eleryi* clades is implied by the  
593 observation that the estimated split times based on both two- and four-taxon models were  
594 substantially earlier than those estimated from mitochondrial genes. Unlike the situation between  
595 *M. gracilis* and *M. recondita* where mitochondrial replacement occurred over only tens of  
596 kilometres, any replacement between groups of *M. eleryi* would have involved much larger  
597 geographic distances of up to 900 km. Given that levels of introgression into invading species are  
598 known to fall with distance from the front wave of hybridization (Currat *et al.* 2008), it is likely  
599 that mitochondrial introgression in our focal bats is very efficient, and probably facilitated by  
600 very low mtDNA gene flow among local populations (demonstrated in Kuo *et al.* 2014 for the  
601 two Taiwanese species).

602

603         Our study of *Murina* sister taxa adds to the mounting evidence for introgression in bats.  
604 Moreover, the inferred patterns - namely unidirectional rampant mtDNA introgression with little  
605 or no nuclear introgression - show similarities to those reported for species of the genera *Myotis*  
606 (Berthier *et al.* 2006) and *Rhinolophus* (Mao *et al.* 2013) as well as members of the family  
607 Pteropodidae (Nesi *et al.* 2011). Higher introgression of mtDNA than of ncDNA has also been  
608 implicitly suggested in other bat species (Larsen *et al.* 2010; Vallo *et al.* 2011), whereas higher  
609 levels of introgression at ncDNA than at mtDNA in bats appears to be much rarer (but see Hulva  
610 *et al.* 2010; Mao *et al.* 2010). Like in the case of *Murina*, Berthier *et al.* (2006) attributed  
611 elevated mtDNA introgression in *Myotis* to contrasting demographic dynamics between the  
612 invading versus resident species, together with male-based gene flow among populations of the

613 former taxon. However, it is important to recognize that multiple alternative scenarios might  
614 account for such patterns, leading Toews and Brelsford (2012) to call for a shift away from  
615 documenting further mito-nuclear discordance towards hypothesis testing to rule out some  
616 explanations of common patterns.

617

### 618 *Conclusions*

619 Our analyses of gene flow at different temporal and spatial scales indicate that rather than  
620 diverging on Taiwan, the two endemic and closely related *Murina* species are a product of  
621 vicariant speciation from a period of isolation. Consequently the current range overlap stems  
622 from multiple incursions, while the previously accepted sister-relationship based on mtDNA is  
623 incorrect and was obscured by mitochondrial introgression coupled with a lack of wider sampling  
624 of geographically-distant continental taxa. This example reveals how historical processes of  
625 vicariance, colonization and hybridization can readily lead to misleading signatures of non-  
626 allopatric speciation, thus highlighting the need for careful analyses to distinguish among such  
627 scenarios. Further caution comes from simulations of bird range data that show that even true  
628 sister taxa that have speciated in allopatry can readily show partial range overlap and,  
629 occasionally, complete range overlap (Phillimore *et al.* 2008). It is perhaps unsurprising,  
630 therefore, that many of the most plausible cases of divergence with gene flow come from plants,  
631 attributed to their greater tendency for fine-scale niche divergence (Anacker & Strauss 2014).  
632 Nonetheless, even in the most convincing and well-studied such examples it is still not trivial to  
633 rule out the alternative explanation of double colonisations, particularly where introgression,  
634 extinction and/or incomplete sampling are also possibilities (see Papadopoulos *et al.* 2011).

**Table 1** Sample sizes, sources and sequencing methods (Sanger or 454-Pyrosequencing) of genetic markers used for reconstruction of divergence processes among the focal taxa.

Taxon	Mitochondrial markers			Nuclear markers	
	n	Cyt- <i>b</i>	COI	n	FR
<i>M. gracilis</i>	87 <sup>c</sup>	II <sup>c</sup> Sanger	II <sup>c</sup> Sanger	15 <sup>e</sup>	I 454
<i>M. recondita</i>	113 <sup>c</sup>	II <sup>c</sup> Sanger	II <sup>c</sup> Sanger	15 <sup>e</sup>	I 454
<i>M. eleryi</i> 1 <sup>a</sup>	8	I Sanger	I, III <sup>d</sup> Sanger	7	I Sanger
<i>M. eleryi</i> 2 <sup>b</sup>	5	I Sanger	I, III <sup>d</sup> Sanger	5	I Sanger

N, sample size; FR, flanking region sequence obtained for ten microsatellite markers; I, this study; II, Kuo *et al.* (2014); III, Francis *et al.* (2010).

<sup>a</sup> *M. eleryi* 1 refers to the single sample from North Vietnam (HNHM 2007.28.2) and samples from Southern China, listed in Table S2 (Supporting information).

<sup>b</sup> *M. eleryi* 2 refers to samples from Central Vietnam, listed in Table S2.

<sup>c</sup> Range-wide samples for which concatenated Cyt-*b* and COI sequences are available with Dryad entry doi:10.5061/dryad.f5th5.

<sup>d</sup> Available with accession numbers HM540936-7, JQ601475, JQ601483, JQ601503 and JQ601510 for *M. eleryi* 1, and HM540933, HM540938, JQ601543 and JQ601545 for *M. eleryi* 2.

<sup>e</sup> Selected samples from sites #49 and #54 for *M. gracilis* and those from sites #48, #51, #52, and #54 for *M. recondita* (see Kuo *et al.* 2014 for details of site locations).

635

**Table 2** Genetic variability of four taxa in the *Murina gracilis* complex based on ten microsatellite flanking regions. Values are calculated for each alignment showing no signature of intra-locus recombination except for the locus A122. Estimates of  $\theta_s$  and  $\pi$  are presented as corresponding raw values multiplied by 100.

Locus	L	<i>M. gracilis</i>				<i>M. recondita</i>				<i>M. eleryi 1</i>				<i>M. eleryi 2</i>			
		Nseq	S	$\theta_s$	$\pi$	Nseq	S	$\theta_s$	$\pi$	Nseq	S	$\theta_s$	$\pi$	Nseq	S	$\theta_s$	$\pi$
A4	216	30	4	0.47	0.48	27	2	0.24	0.22	14	2	0.29	0.19	10	7	1.15	1.12
A9	102	29	3	0.75	1.22	15	0	0	0	14	3	0.93	0.97	10	2	0.69	0.55
A104	156	30	3	0.49	0.93	25	0	0	0	9	4	0.94	0.68	10	1	0.23	0.13
A109	137	30	2	0.37	0.45	30	1	0.18	0.05	14	3	0.69	0.40	6	2	0.64	0.49
A118	180	30	4	0.56	0.59	21	1	0.15	0.24	12	4	0.74	0.37	8	4	0.86	1.03
A122	184	30	5	0.69	0.63	24	7	1.02	0.86	6	10	2.38	2.86	10	7	1.35	1.88
B5	286	30	0	0	0	24	0	0	0	12	1	0.12	0.06	10	1	0.12	0.17
B114	318	30	0	0	0	23	1	0.09	0.16	12	2	0.21	0.27	8	4	0.49	0.60
B124	249	30	0	0	0	20	0	0	0	13	4	0.52	0.45	10	2	0.28	0.27
D117	234	30	4	0.43	0.42	29	0	0	0	12	5	0.71	0.86	8	5	0.82	0.93

L, sequence length in number of base pairs; Nseq, sample sizes of sequences; S, number of segregating sites;  $\theta_s$ , Watterson's (1975) theta;  $\pi$ , nucleotide diversity.

636 **Figure captions**

637 **Figure 1** Bayesian clustering analyses implemented in STRUCTURE. The upper panel shows  
638 penalized log-likelihood values (Pritchard *et al.* 2000) of ten replicated runs under each of  
639 successive numbers of clusters (K) for the full dataset (14 microsatellite loci). The lower panel  
640 shows DISTRUCT plots for analyses based on the full dataset and based on 13 loci with either  
641 locus A9 or locus A122 removed. Selected individuals for sequencing of flanking regions of loci  
642 A9 and A122 are indicated by arrows.

643 **Figure 2** Median-joining networks based on flanking regions of microsatellite loci A9 and A122.  
644 Circles are coloured to represent haplotypes of *Murina gracilis* (yellow), *M. eleryi* (light blue),  
645 four individuals of *M. recondita* labeled as RA, RB, RC, and RD in Figure 1 (grey), and  
646 remaining individuals of *M. recondita* (dark blue). Sizes of circles are proportional to sample  
647 sizes of unique haplotypes. The scale bar applies to both networks.

648 **Figure 3** Divergence within the *Murina gracilis* complex inferred from \*BEAST analyses based  
649 on (a) mtDNA and (b) ncDNA. In each panel, extant taxa are labeled as G for *M. gracilis*, R for  
650 *M. recondita*, E1 for *M. eleryi* 1 and E2 for *M. eleryi* 2; the upper left corner shows the maximum  
651 clade credibility (MCC) topology with posteriors for specific taxon grouping, and the lower right  
652 corner shows demographic estimates obtained under corresponding MCC topology. In the  
653 demographic plots, horizontal and vertical dimensions are scaled to represent effective population  
654 sizes (boxes) and split times (horizontal lines; present time at the top), respectively. Black boxes  
655 and black horizontal lines are scaled to represent median estimates, while grey ones including  
656 lines with double arrows are scaled to present 95% confidence intervals of corresponding  
657 variables. Scale bars for 100,000 individuals and for one million years apply to both panels.

658 **Figure 4** Divergence within the *Murina gracilis* complex inferred from four-taxon IM analyses  
659 under the 4PM2 prior. Extant taxa are labeled as in Figure 3 while ancestral populations are  
660 labeled as A1 for *M. recondita* plus *M. eleryi* 1, A2 for *M. recondita* plus both *M. eleryi* taxa, and  
661 A3 for all extant taxa. Split times are numbered to present successive divergence events from  
662 recent to ancient ones under the specified phylogeny - (((R, E1), E2), G). Horizontal and vertical  
663 dimensions are scaled as in Figure 3 except that black boxes and black horizontal lines are scaled  
664 to present modal estimates. Black curved arrows mark directions of post-split migrations, each of  
665 which was significant using the population migration rate (2NM) based on Nielsen and  
666 Wakeley's (2001) likelihood ratio test; modal estimates of these 2NM values are also given.

667 **Figure 5** Divergence within the *Murina gracilis* complex recovered by pairwise two-taxon IM  
668 analyses under the 2PM2 prior. For an explanation of the labels used and format of the plots, see  
669 Figure 4. Scale bars for 100,000 individuals and for one million years apply to all six plots.

670

671 **Acknowledgements**

672 We thank Ian Barnes, Michael Bruford, Kalina Davies and Richard Nichols for helpful comments  
673 on this work. HCK was supported by the Overseas Research Students Awards Scheme and the  
674 University of London Central Research Fund as well as by the Studying Abroad Scholarship  
675 from the Taiwanese Ministry of Education. GC received support from the SYNTHESYS Project,  
676 which is financed by the European Community Research Infrastructure Action under the FP7  
677 "Capacities" Program, and from the Hungarian Scientific Research Fund (OTKA) K112440. This  
678 research was also funded by a grant to SFC, YPF and SJR from the Taiwanese Ministry of  
679 Science and Technology (NSC 100-2621-B-126 -001).

680

681 **References**

- 682 Adams RI, Brown KM, Hamilton MB (2004) The impact of microsatellite electromorph size  
683 homoplasy on multilocus population structure estimates in a tropical tree (*Corythophora*  
684 *alta*) and an anadromous fish (*Morone saxatilis*). *Molecular Ecology*, **13**, 2579-2588.
- 685 Anacker BL, Strauss SY (2014) The geography and ecology of plant speciation: range overlap  
686 and niche divergence in sister species. *Proceedings of the Royal Society B-Biological*  
687 *Sciences*, **281**.
- 688 Barraclough TG, Vogler AP (2000) Detecting the geographical pattern of speciation from  
689 species-level phylogenies. *American Naturalist*, **155**, 419-434.
- 690 Berthier P, Excoffier L, Ruedi M (2006) Recurrent replacement of mtDNA and cryptic  
691 hybridization between two sibling bat species *Myotis myotis* and *Myotis blythii*.  
692 *Proceedings of the Royal Society B-Biological Sciences*, **273**, 3101-3109.
- 693 Bogdanowicz W, Piksa K, Tereba A (2012) Hybridization hotspots at bat swarming sites. *Plos*  
694 *One*, **7**, e53334.
- 695 Brown RM, Nichols RA, Faulkes CG, *et al.* (2010) Range expansion and hybridization in Round  
696 Island petrels (*Pterodroma* spp.): evidence from microsatellite genotypes. *Molecular*  
697 *Ecology*, **19**, 3157-3170.
- 698 Chan KMA, Levin SA (2005) Leaky prezygotic isolation and porous genomes: Rapid  
699 introgression of maternally inherited DNA. *Evolution*, **59**, 720-729.
- 700 Coyne JA (2007) Sympatric speciation. *Current Biology*, **17**, R787-R788.
- 701 Coyne JA, Price TD (2000) Little evidence for sympatric speciation in island birds. *Evolution*, **54**,  
702 2166-2171.
- 703 Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The hidden side of invasions: massive  
704 introgression by local genes. *Evolution*, **62**, 1908-1920.
- 705 Dieckmann U, Doebeli M (1999) On the origin of species by sympatric speciation. *Nature*, **400**,  
706 354-357.
- 707 Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti  
708 and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969-1973.
- 709 Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high  
710 throughput. *Nucleic Acids Research*, **32**, 1792-1797.
- 711 Eger JL, Lim BK (2011) Three new species of *Murina* from southern China (Chiroptera:  
712 Vespertilionidae). *Acta Chiropterologica*, **13**, 227-243.

- 713 Estoup A, Jarne P, Cornuet JM (2002) Homoplasmy and mutation model at microsatellite loci and  
714 their consequences for population genetics analysis. *Molecular Ecology*, **11**, 1591-1604.
- 715 Fitzpatrick BM, Fordyce JA, Gavrilets S (2009) Pattern, process and geographic modes of  
716 speciation. *Journal of Evolutionary Biology*, **22**, 2342-2347.
- 717 Flot J-F, Tillier A, Samadi S, Tillier S (2006) Phase determination from direct sequencing of  
718 length-variable DNA regions. *Molecular Ecology Notes*, **6**, 627-630.
- 719 Fluxus Technology Ltd. (2010) Network. Available from: [http://www.fluxus-](http://www.fluxus-engineering.com/sharenet.htm)  
720 [engineering.com/sharenet.htm](http://www.fluxus-engineering.com/sharenet.htm).
- 721 Forbes AA, Powell THQ, Stelinski LL, Smith JJ, Feder JL (2009) Sequential sympatric  
722 speciation across trophic levels. *Science*, **323**, 776-779.
- 723 Francis CM, Borisenko AV, Ivanova NV, *et al.* (2010) The role of DNA barcodes in  
724 understanding and conservation of mammal diversity in Southeast Asia. *Plos One*, **5**,  
725 e12575.
- 726 Furey NM, Thong VD, Bates PJJ, Csorba G (2009) Description of a new species belonging to the  
727 *Murina* 'suilla-group' (Chiroptera: Vespertilionidae: Murininae) from north Vietnam. *Acta*  
728 *Chiropterologica*, **11**, 225-236.
- 729 Galan M, Guivier E, Caraux G, Charbonnel N, Cosson J-F (2010) A 454 multiplex sequencing  
730 method for rapid and reliable genotyping of highly polymorphic genes in large-scale  
731 studies. *BMC Genomics*, **11**, 296.
- 732 Galtier N, Nabholz B, Glemin S, Hurst GDD (2009) Mitochondrial DNA as a marker of  
733 molecular diversity: a reappraisal. *Molecular Ecology*, **18**, 4541-4550.
- 734 Garrick RC, Sunnucks P, Dyer RJ (2010) Nuclear gene phylogeography using PHASE: dealing  
735 with unresolved genotypes, lost alleles, and systematic bias in parameter estimation. *BMC*  
736 *Evolutionary Biology*, **10**, 118.
- 737 Gavrilets S, Waxman D (2002) Sympatric speciation by sexual conflict. *Proceedings of the*  
738 *National Academy of Sciences of the United States of America*, **99**, 10533-10538.
- 739 Geraldes A, Ferrand N, Nachman MW (2006) Contrasting patterns of introgression at X-linked  
740 loci across the hybrid zone between subspecies of the European rabbit (*Oryctolagus*  
741 *cuniculus*). *Genetics*, **173**, 919-933.
- 742 Guillot G (2008) Inference of structure in subdivided populations at low levels of genetic  
743 differentiation-the correlated allele frequencies model revisited. *Bioinformatics*, **24**, 2222-  
744 2228.
- 745 Hasegawa M, Kishino H, Yano TA (1985) Dating of the human-ape splitting by a molecular  
746 clock of mitochondrial DNA. *Journal of Molecular Evolution*, **22**, 160-174.
- 747 Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data.  
748 *Molecular Biology and Evolution*, **27**, 570-580.
- 749 Hey J (2006) Recent advances in assessing gene flow between diverging populations and species.  
750 *Current Opinion in Genetics & Development*, **16**, 592-596.
- 751 Hey J (2010a) The divergence of chimpanzee species and subspecies as revealed in  
752 multipopulation isolation-with-migration analyses. *Molecular Biology and Evolution*, **27**,  
753 921-933.
- 754 Hey J (2010b) Isolation with migration models for more than two populations. *Molecular Biology*  
755 *and Evolution*, **27**, 905-920.
- 756 Hey J (2011) Documentation for IMA2. Available from:  
757 [https://bio.cst.temple.edu/~hey/program\\_files/IMA2/Using\\_IMA2\\_8\\_24\\_2011.pdf](https://bio.cst.temple.edu/~hey/program_files/IMA2/Using_IMA2_8_24_2011.pdf).

- 758 Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and  
759 divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D.*  
760 *persimilis*. *Genetics*, **167**, 747-760.
- 761 Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain  
762 Monte Carlo methods in population genetics. *Proceedings of the National Academy of*  
763 *Sciences of the United States of America*, **104**, 2785-2790.
- 764 Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in  
765 the history of a sample of DNA sequences. *Genetics*, **111**, 147-164.
- 766 Hulva P, Fornuskova A, Chudarkova A, *et al.* (2010) Mechanisms of radiation in a bat group  
767 from the genus *Pipistrellus* inferred by phylogeography, demography and population  
768 genetics. *Molecular Ecology*, **19**, 5417-5431.
- 769 Husband BC, Sabara HA (2004) Reproductive isolation between autotetraploids and their diploid  
770 progenitors in fireweed, *Chamerion angustifolium* (Onagraceae). *New Phytologist*, **161**,  
771 703-713.
- 772 Jackson DE (2008) Sympatric speciation: perfume preferences of orchid bee lineages. *Current*  
773 *Biology*, **18**, R1092-R1093.
- 774 Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for  
775 dealing with label switching and multimodality in analysis of population structure.  
776 *Bioinformatics*, **23**, 1801-1806.
- 777 Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: *Mammalian protein metabolism*  
778 (ed Munro HH), pp. 21-132. Academic Press, New York.
- 779 Kingston T, Francis CM, Akbar Z, Kunz TH (2003) Species richness in an insectivorous bat  
780 assemblage from Malaysia. *Journal of Tropical Ecology*, **19**, 67-79.
- 781 Kingston T, Jones G, Akbar Z, Kunz TH (1999) Echolocation signal design in Kerivoulineae and  
782 Murinineae (Chiroptera: Vespertilionidae) from Malaysia. *Journal of Zoology*, **249**, 359-  
783 374.
- 784 Kingston T, Rossiter SJ (2004) Harmonic-hopping in Wallacea's bats. *Nature*, **429**, 654-657.
- 785 Kisel Y, Barraclough TG (2010) Speciation has a spatial scale that depends on levels of gene  
786 flow. *American Naturalist*, **175**, 316-334.
- 787 Kumar S, Subramanian S (2002) Mutation rates in mammalian genomes. *Proceedings of the*  
788 *National Academy of Sciences of the United States of America*, **99**, 803-808.
- 789 Kuo H-C (2004) *Systematics of bats of genus Murina in Taiwan (Chiroptera: Vespertilionidae)*.  
790 MS Thesis, National Taiwan University, Taipei.
- 791 Kuo H-C (2013) *Phylogeography and diversification of Taiwanese bats*. PhD Thesis, Queen  
792 Mary, University of London, London.
- 793 Kuo H-C, Chen S-F, Fang Y-P, Flanders J, Rossiter SJ (2014) Comparative rangewide  
794 phylogeography of four endemic Taiwanese bat species. *Molecular Ecology*, **23**, 3566-  
795 3586.
- 796 Kuo H-C, Chen S-F, Rossiter SJ (2013) Development and characterisation of microsatellite loci  
797 for *Murina gracilis* (Vespertilionidae) and cross-species amplification in two other  
798 congeneric species. *Conservation Genetics Resources*, **5**, 1117-1120.
- 799 Kuo H-C, Fang Y-P, Csorba G, Lee L-L (2009) Three new species of *Murina* (Chiroptera:  
800 Vespertilionidae) from Taiwan. *Journal of Mammalogy*, **90**, 980-991.

- 801 Larsen PA, Marchán-Rivadeneira MR, Baker RJ (2010) Natural hybridization generates  
802 mammalian lineage with species characteristics. *Proceedings of the National Academy of*  
803 *Sciences of the United States of America*, **107**, 11447-11452.
- 804 Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA  
805 polymorphism data. *Bioinformatics*, **25**, 1451-1452.
- 806 Llopart A, Elwyn S, Lachaise D, Coyne JA (2002) Genetics of a difference in pigmentation  
807 between *Drosophila yakuba* and *Drosophila santomea*. *Evolution*, **56**, 2262-2277.
- 808 Losos JB, Glor RE (2003) Phylogenetic comparative methods and the geography of speciation.  
809 *Trends in Ecology & Evolution*, **18**, 220-227.
- 810 Lue K-Y, Lin S-M (2008) Two new cryptic species of *Takydromus* (squamata: lacertidae) from  
811 taiwan. *Herpetologica*, **64**, 379-395.
- 812 Lynch JD (1989) The gauge of speciation: on the frequencies of modes of speciation. In:  
813 *Speciation and its Consequences* (eds Otte D, Endler JA), pp. 527-553. Sinauer  
814 Associates, Sunderland, MA.
- 815 Mallet J (2005) Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, **20**,  
816 229-237.
- 817 Mao X, He G, Zhang J, Rossiter SJ, Zhang S (2013) Lineage divergence and historical gene flow  
818 in the Chinese horseshoe bat (*Rhinolophus sinicus*). *Plos One*, **8**, e56786.
- 819 Mao X, Zhang J, Zhang S, Rossiter SJ (2010) Historical male-mediated introgression in  
820 horseshoe bats revealed by multilocus DNA sequence data. *Molecular Ecology*, **19**, 1352-  
821 1366.
- 822 Margulies M, Egholm M, Altman WE, *et al.* (2005) Genome sequencing in microfabricated high-  
823 density picolitre reactors. *Nature*, **437**, 376-380.
- 824 Mattern MY, McLennan DA (2000) Phylogeny and speciation of felids. *Cladistics*, **16**, 232-253.
- 825 Melo-Ferreira J, Vilela J, Fonseca MM, *et al.* (2014) The elusive nature of adaptive  
826 mitochondrial DNA evolution of an arctic lineage prone to frequent introgression.  
827 *Genome Biology and Evolution*, **6**, 886-896.
- 828 Muir G, Schlötterer C (2005) Evidence for shared ancestral polymorphism rather than recurrent  
829 gene flow at microsatellite loci differentiating two hybridizing oaks (*Quercus* spp.).  
830 *Molecular Ecology*, **14**, 549-561.
- 831 Nadachowska K, Babik W (2009) Divergence in the face of gene flow: the case of two newts  
832 (Amphibia: Salamandridae). *Molecular Biology and Evolution*, **26**, 829-841.
- 833 Nesi N, Nakoune E, Cruaud C, Hassanin A (2011) DNA barcoding of African fruit bats  
834 (Mammalia, Pteropodidae). The mitochondrial genome does not provide a reliable  
835 discrimination between *Epomophorus gambianus* and *Micropteropus pusillus*. *Comptes*  
836 *Rendus Biologies*, **334**, 544-554.
- 837 Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: A Markov chain Monte  
838 Carlo approach. *Genetics*, **158**, 885-896.
- 839 Papadopulos AST, Baker WJ, Crayn D, *et al.* (2011) Speciation with gene flow on Lord Howe  
840 Island. *Proceedings of the National Academy of Sciences of the United States of America*,  
841 **108**, 13188-13193.
- 842 Petit RJ, Excoffier L (2009) Gene flow and species delimitation. *Trends in Ecology & Evolution*,  
843 **24**, 386-393.

- 844 Phillimore AB, Orme CDL, Thomas GH, *et al.* (2008) Sympatric speciation in birds is rare:  
845 Insights from range data and simulations. *American Naturalist*, **171**, 646-657.
- 846 Polzin T, Daneshmand SV (2003) On Steiner trees and minimum spanning trees in hypergraphs.  
847 *Operations Research Letters*, **31**, 12-20.
- 848 Posada D (2008) jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution*,  
849 **25**, 1253-1256.
- 850 Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus  
851 genotype data. *Genetics*, **155**, 945-959.
- 852 Rambaut A, Drummond AJ (2007) Tracer v1.4. Available from  
853 <http://tree.bio.ed.ac.uk/software/tracer/>.
- 854 Randi E (2008) Detecting hybridization between wild species and their domesticated relatives.  
855 *Molecular Ecology*, **17**, 285-293.
- 856 Rice WR (1984) Disruptive selection on habitat preference and the evolution of reproductive  
857 isolation - a simulation study. *Evolution*, **38**, 1251-1260.
- 858 Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure.  
859 *Molecular Ecology Notes*, **4**, 137-138.
- 860 Rossiter SJ, Benda P, Dietz C, Zhang S, Jones G (2007) Rangewide phylogeography in the  
861 greater horseshoe bat inferred from microsatellites: implications for population history,  
862 taxonomy and conservation. *Molecular Ecology*, **16**, 4699-4714.
- 863 Schmieder DA, Kingston T, Hashim R, Siemers BM (2010) Breaking the trade-off: rainforest  
864 bats maximize bandwidth and repetition rate of echolocation calls as they approach prey.  
865 *Biology Letters*, **6**, 604-609.
- 866 Slatkin M (1982) Pleiotropy and parapatric speciation. *Evolution*, **36**, 263-270.
- 867 Stephens M, Scheet P (2005) Accounting for decay of linkage disequilibrium in haplotype  
868 inference and missing data imputation. *American Journal of Human Genetics*, **76**, 449-  
869 462.
- 870 Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction  
871 from population data. *American Journal of Human Genetics*, **68**, 978-989.
- 872 Strasburg JL, Rieseberg LH (2011) Interpreting the estimated timing of migration events between  
873 hybridizing species. *Molecular Ecology*, **20**, 2353-2366.
- 874 Struebig MJ, Kingston T, Petit EJ, *et al.* (2011) Parallel declines in species and genetic diversity  
875 in tropical forest fragments. *Ecology Letters*, **14**, 582-590.
- 876 Tamura K, Peterson D, Peterson N, *et al.* (2011) MEGA5: molecular evolutionary genetics  
877 analysis using maximum likelihood, evolutionary distance, and maximum parsimony  
878 methods. *Molecular Biology and Evolution*, **28**, 2731-2739.
- 879 Tavaré S (1986) Some probabilistic and statistical problems in the analysis of DNA sequences. In:  
880 *Some mathematical questions in biology: DNA sequence analysis* (ed Miura RM), pp. 57-  
881 86. American Mathematical Society, Providence, Rhode Island.
- 882 Toews DPL, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in  
883 animals. *Molecular Ecology*, **21**, 3907-3930.
- 884 Turelli M, Barton NH, Coyne JA (2001) Theory and speciation. *Trends in Ecology & Evolution*,  
885 **16**, 330-343.
- 886 Vallo P, Benda P, Cerveny J, Koubek P (2013) Conflicting mitochondrial and nuclear paraphyly  
887 in small-sized West African house bats (*Vespertilionidae*). *Zoologica Scripta*, **42**, 1-12.

- 888 Via S (2001) Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology &*  
889 *Evolution*, **16**, 381-390.
- 890 Voris HK (2000) Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and  
891 time durations. *Journal of Biogeography*, **27**, 1153-1167.
- 892 Watterson GA (1975) On the number of segregating sites in genetical models without  
893 recombination. *Theoretical Population Biology*, **7**, 256-276.
- 894 White MJD (1978) *Modes of speciation*. W. H. Freeman and Company, San Francisco, CA.
- 895 Woerner AE, Cox MP, Hammer MF (2007) Recombination-filtered genomic datasets by  
896 information maximization. *Bioinformatics*, **23**, 1851-1853.
- 897 Yu HT (1995) Patterns of diversification and genetic population structure of small mammals in  
898 Taiwan. *Biological Journal of the Linnean Society*, **55**, 69-89.
- 899 Yule GU (1924) A mathematical theory of evolution based on the conclusions of Dr J. C. Willis.  
900 *Philosophical Transactions of the Royal Society B-Biological Sciences*, **213**, 21-87.

901

### 902 **Data accessibility**

903 GenBank accessions for new *Murina eleryi* sequences are KT762291-KT762293 for COI,  
904 KT762294-KT762306 for *Cyt-b* and KT778876-KT779034 (unphased) for microsatellite  
905 flanking regions. GenBank accessions for microsatellite flanking sequences of *M. gracilis* and *M.*  
906 *recondita* generated via 454-Pyrosequencing are KT794560-KT795100. For all focal taxa studied,  
907 we created an archive comprising (1) alignments of phased sequences at two microsatellite  
908 flanking regions used to produce Fig. 2, (2) an alignment of concatenated *Cyt-b* and COI genes  
909 used to produce Fig. 3a, (3) alignments of phased sequences at 10 microsatellite flanking regions  
910 used to produce Fig. 3b and (4) input data for the four-taxon IM analyses. This archive is  
911 accessible with the Dryad entry: XXX.

912

### 913 **Authors' contributions**

914 H.-C.K., S.-F.C., Y.-P.F, C.-H. Chen and C.-H. Chou collected tissue samples from Taiwanese  
915 taxa; G.C., B.K.L. and J.L.E. provided tissue samples of *M. eleryi*; H.-C.K. conducted laboratory  
916 work under the supervision of SJR, and performed statistical analyses with input from S.J.R,  
917 J.A.C. and J.D.P. H.-C.K. wrote the paper with S.J.R and all authors participated in discussions.

918

### 919 **Supporting information**

920 Systematic notes on the focal species.

921 Selection of a demographic model for the ncDNA \*BEAST analysis.

922 Tests of positive selection.

923 Tests for sex-biased dispersal in *Murina gracilis* and *M. recondita*.

924 **Figure S1** Median-joining networks built for the *Murina gracilis* complex based on ten  
925 microsatellite flanking regions.

- 926 **Figure S2** Marginal densities for demographic parameters from a four-taxon IM analysis of the  
927 *Murina gracilis* complex.
- 928 **Figure S3** Marginal densities for demographic parameters from IM analyses for the six two-  
929 taxon pairs of the *Murina gracilis* complex.
- 930 **Figure S4** A mitochondrial gene tree used in tests of positive selection.
- 931 **Figure S5** Samples of *Murina gracilis* and *M. recondita* analysed for sex-biased dispersal.
- 932 **Figure S6** Correlograms of individual-level analyses of sex-biased dispersal in *Murina gracilis*  
933 and *M. recondita*.
- 934 **Table S1** External measurements showing body sizes of *M. gracilis* and its relatives.
- 935 **Table S2** Details of 13 voucher specimens of *Murina eleryi* sampled for genetic analyses.
- 936 **Table S3** Primer pairs for amplification of flanking regions of ten microsatellite loci.
- 937 **Table S4** Estimated proportions of 'foreign' genetic composition for four individual bats of  
938 *Murina recondita* as inferred from STRUCTURE analyses.
- 939 **Table S5** IM estimates based on all four taxa from the *Murina gracilis* complex.
- 940 **Table S6** IM estimates for six pairs of two taxa from the *Murina gracilis* complex.
- 941 **Table S7** Sources of mitochondrial sequences used in maximum likelihood analyses in PAML to  
942 test for signatures of positive selection.
- 943 **Table S8** Predictions and results of the population-level tests - based on four statistics - for  
944 evidence of male-biased dispersal in *M. gracilis* and *M. recondita*.
- 945

Figure 1

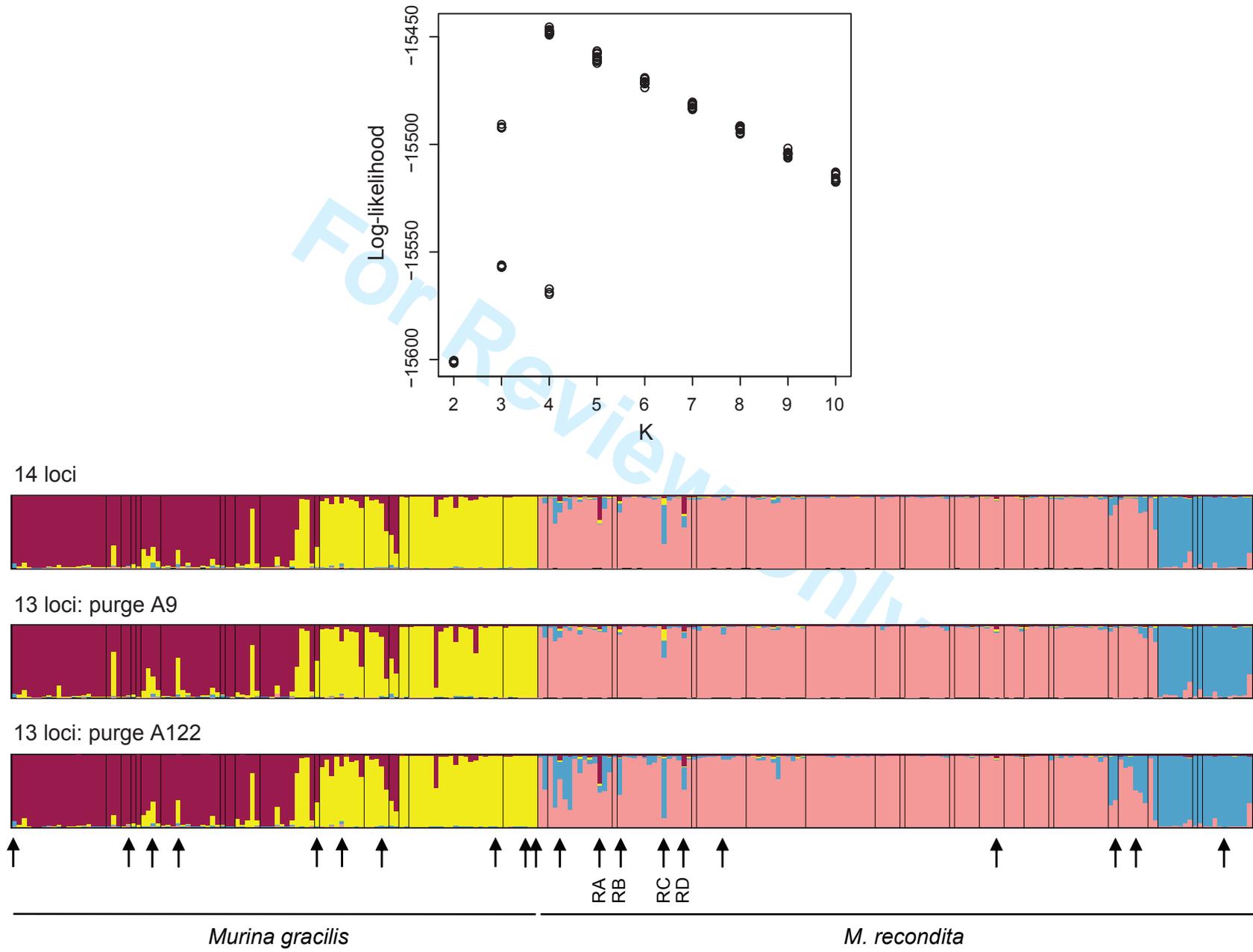




Figure 3

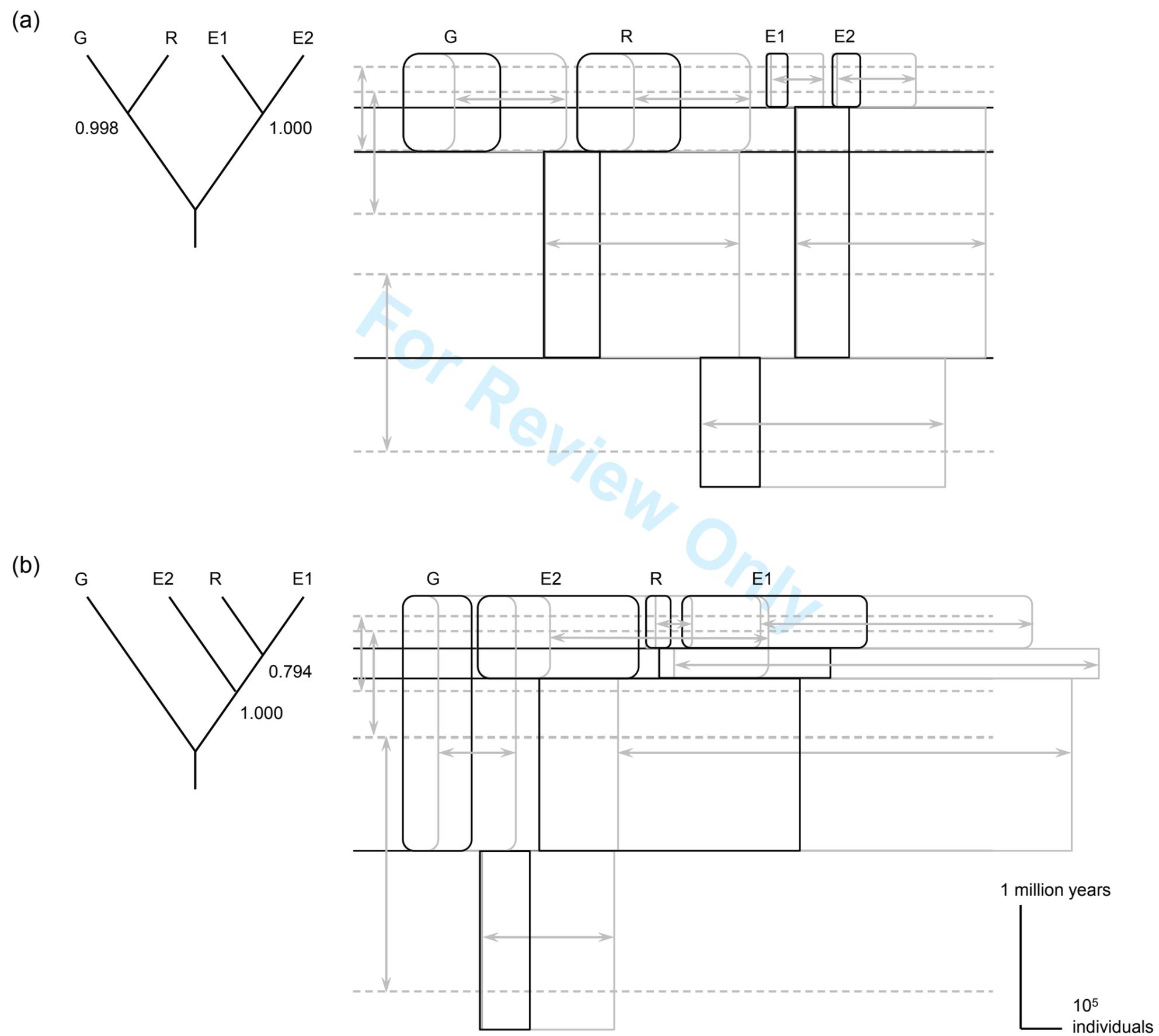


Figure 4

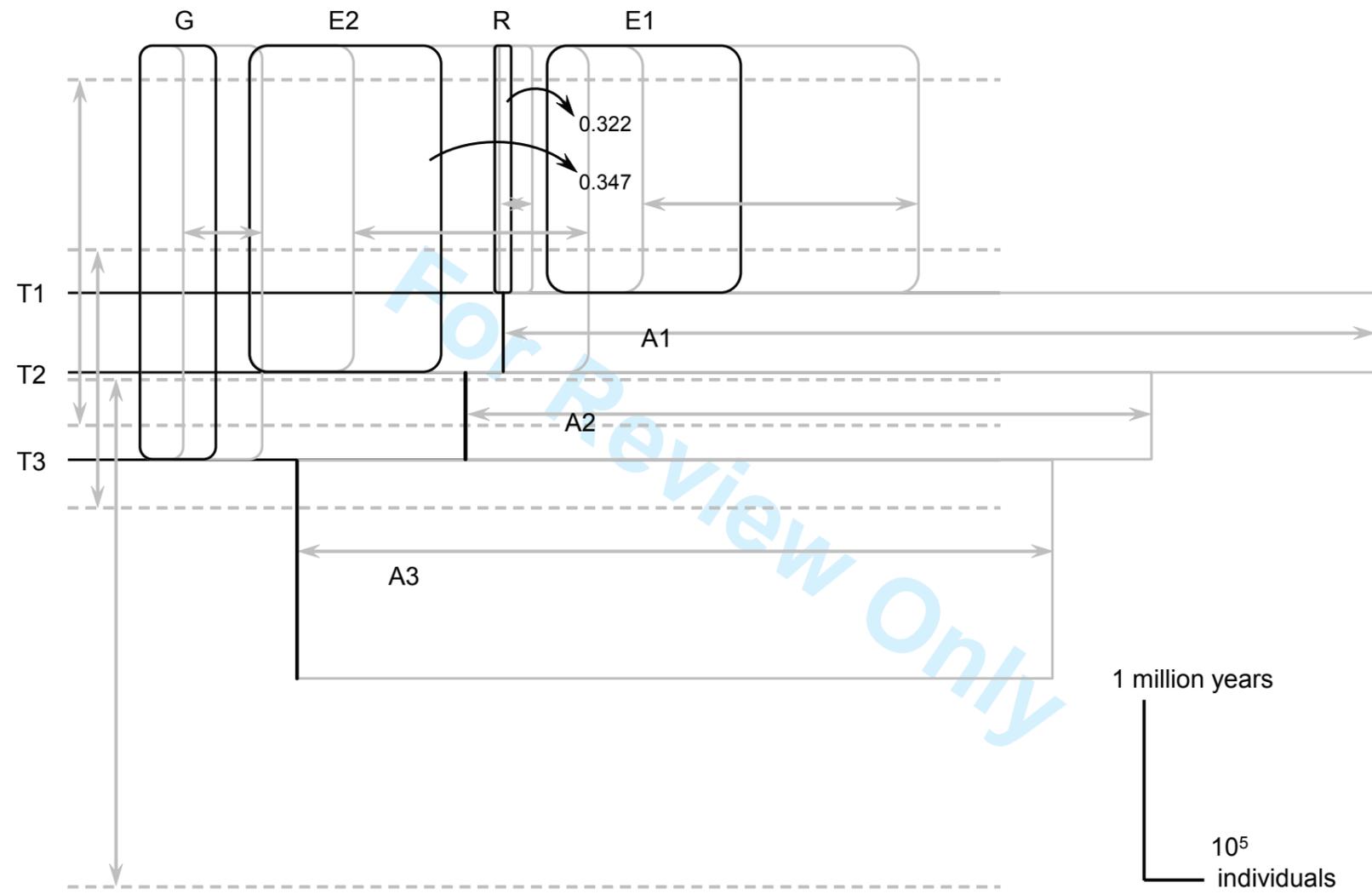


Figure 5

