



Ecological speciation in sympatric palms: 4. Demographic analyses support speciation of *Howea* in the face of high gene flow

Papadopoulos, Alexander S. T.; Igea, Javier; Smith, Thomas P.; Hutton, Ian; Baker, William J.; Butlin, Roger K.; Savolainen, Vincent

Evolution

DOI:

[10.1111/evo.13813](https://doi.org/10.1111/evo.13813)

[10.1111/evo.13813](https://doi.org/10.1111/evo.13813)

Published: 01/09/2019

Peer reviewed version

[Cyswllt i'r cyhoeddiad / Link to publication](#)

Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):

Papadopoulos, A. S. T., Igea, J., Smith, T. P., Hutton, I., Baker, W. J., Butlin, R. K., & Savolainen, V. (2019). Ecological speciation in sympatric palms: 4. Demographic analyses support speciation of *Howea* in the face of high gene flow. *Evolution*, 73(9), 1996-2002.
<https://doi.org/10.1111/evo.13813>, <https://doi.org/10.1111/evo.13813>

Hawliau Cyffredinol / General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

BRIEF COMMUNICATION

Ecological speciation in sympatric palms: 4. Demographic analyses support speciation of *Howea* in the face of high gene flow

Alexander S.T. Papadopoulos^{1,2}, Javier Igea^{1,3}, Thomas P. Smith¹, Ian Hutton⁴, William J. Baker⁵, Roger K. Butlin^{6,7} and Vincent Savolainen^{1,5,*}

¹Department of Life Sciences, Silwood Park Campus, Imperial College London, Buckhurst Road, Ascot, SL5 7PY, UK.

²Molecular Ecology and Fisheries Genetics Laboratory, Environment Centre Wales, School of Biological Sciences, Bangor University, Bangor, LL57 2UW, UK.

³Department of Plant Sciences, University of Cambridge, Cambridge, CB2 3EA, UK.

⁴Lord Howe Island Museum, Lord Howe Island, New South Wales, Australia.

⁵Royal Botanic Gardens, Kew, Richmond, TW9 3AB, UK.

⁶Department of Animal and Plant Sciences, University of Sheffield, Sheffield, S10 2TN, UK.

⁷Department of Marine Sciences, University of Gothenburg, Gothenburg, SE-405 30, Sweden

*Correspondence: email v.savolainen@imperial.ac.uk

Running title (40 characters) Gene flow and speciation in *Howea* palms

Keywords Coalescence, sympatry, speciation, ddRAD

AUTHOR CONTRIBUTIONS

VS designed the research with contributions from ASTP, JI, and RKB. ASTP and TPS collected data. ASTP and JI analysed the data. IH and WB contributed to field collections. ASTP and VS wrote the manuscript. All authors commented on the manuscript.

ACKNOWLEDGMENTS

We thank the Lord Howe Island Board and the New South Wales National Park and Wildlife Services for granting research permits, Hank and Sue Bower, Christo Haselden, Peter Weston, Larry Wilson for their help on LHI, Laszlo Csiba and Helen Hipperson for help in the lab, Matt Hahn for comments, and the European Research Council, NERC and the Leverhulme Trust for funding.

DATA ACCESSIBILITY

The sequence data are available at the Sequence Reads Archive under accession numbers PRJNA386480 and SRP063985.

1 **BRIEF COMMUNICATION**

2

3 **Ecological speciation in sympatric palms: 4. Demographic analyses support speciation of**
4 ***Howea* in the face of high gene flow**

5

6 **Abstract**

7 The idea that populations must be geographically isolated (allopatric) to evolve into separate
8 species has persisted for a long time. It is now clear that new species can also diverge despite
9 ongoing genetic exchange, but few accepted cases of speciation in sympatry have held up when
10 scrutinised using modern approaches. Here, we examined evidence for speciation of the *Howea*
11 palms of Lord Howe Island, Australia, in light of new genomic data. We used coalescence-based
12 demographic models combined with double digest restriction-site associated DNA sequencing of
13 multiple individuals and provide support for previous claims by Savolainen et al. (Nature 441: 210–
14 213, 2006) that speciation in *Howea* did occur in the face of gene flow.

15

16 **Introduction**

17 Sympatric speciation has re-emerged as a controversial topic, with recent analyses of genome-wide
18 data casting doubt on some of the most best-known examples, such as cichlid fish in Cameroonian
19 crater lakes (Martin *et al.* 2015). Due to a lack of confidence in ascertaining whether speciation has
20 taken place in the face of gene flow, our understanding of the genomic underpinning of such
21 processes has also remained piecemeal (Renaut *et al.* 2013; Cruickshank & Hahn 2014). Here, we
22 examined a case of speciation in *Howea* palms, a genus that comprises only two species, both
23 endemic to the subtropical Lord Howe Island (LHI; Savolainen *et al.* 2006). The island is isolated
24 (600 km from mainland Australia) and minute (<16 km²). Furthermore, modelling of the geological
25 history and sizes of LHI and nearby Ball's Pyramid rock showed that, for any pair of endemic sister
26 species that have diverged within the lifetime of the island, an allopatric phase in their divergence is
27 unlikely (Papadopulos *et al.* 2011). This was critical in promoting *Howea* as a prime example of
28 sympatric speciation under a biogeographic definition (Savolainen *et al.* 2006; Mallet *et al.* 2009;
29 Coyne 2011; Papadopulos *et al.* 2011). Marked flowering time differences between the species
30 indicate that prezygotic isolation is now strong and current levels of gene flow are likely to be low
31 (Savolainen *et al.* 2006; Babik *et al.* 2009; Dunning *et al.* 2016; Hipperson *et al.* 2016). Indirect
32 evidence of post-zygotic isolation due to selection against juvenile hybrids supports the hypothesis
33 that divergent selection has influenced the speciation process (Hipperson *et al.* 2016). Given that the
34 distributions of *Howea* palms overlap extensively and that they are wind pollinated, Savolainen *et*
35 *al.* (2006) argued that speciation is likely to have occurred in the face of high gene flow, which may
36 have reduced quickly as divergence progressed (Savolainen *et al.* 2006; Babik *et al.* 2009;

37 Papadopulos *et al.* 2011, 2013b, 2014). However, previous efforts were based on a limited number
38 of markers (two gene sequences and amplified fragment length polymorphism; Savolainen *et al.*
39 2006, Babik *et al.* 2009), which leaves room for doubt about the precise timing of divergence and
40 did not allow for the quantification of the extent of the gene exchange between *H. forsteriana* and
41 *H. belmoreana*. Here, we characterise the demographic history of the *Howea* palms with a genome-
42 wide dataset to evaluate (i) whether genomic data support speciation with gene flow and (ii) how
43 genetic exchange progressed during speciation in sympatry.

44

45 **Material and Methods**

46

47 **DNA EXTRACTION AND DOUBLE DIGEST RAD-SEQUENCING**

48 Leaf tissue was collected and preserved in silica gel from 42 *H. belmoreana* and 54 *H. forsteriana*
49 individuals sampled at Far Flats on LHI where both species co-occur. Genomic DNA was extracted
50 from leaf tissue using DNeasy Plant Mini kits (Qiagen). Agarose gel electrophoresis was used to
51 assess the quality of each DNA extract and DNA quantification was performed with a Qubit 2.0
52 fluorometer (Life technologies). DNA samples were then processed following a modified version of
53 the double digest RADseq protocol of Peterson *et al.* (2012). Digestion of the DNA template of
54 each sample (250-1000ng) was performed by combining the sample with 0.1 μ l *EcoRI*-HF (10 U;
55 NEB), 1 μ l *MspI* (10U; NEB), 5 μ l NEB buffer 4, and nuclease-free water to a total volume of 50 μ l.
56 Digestion reactions were incubated at 37°C for 3 hours. Digests were cleaned using the Agencourt
57 AMPure XP bead clean up (Beckman Coulter) and quantified using a Qubit 2.0 fluorometer.
58 Ligation of 6 base pairs (bp) - barcoded P1 (*EcoRI* overhang) and P2 (*MspI*) adaptors was
59 performed in individual reactions composed of ca. 400 ng of the digested DNA product, 2 μ l of
60 each adaptor (4 nM), 2 μ l T4 DNA ligase (4000 U; NEB), 4 μ l 10x Ligase buffer (NEB) and
61 nuclease-free water to a total of volume 40 μ l. Ligation reactions were incubated at 23°C for 30
62 min, followed by 65°C for 10 min, and then cooled to 18°C at a rate of 2°C / 90 seconds. Samples
63 were pooled into batches of 12 containing compatible sets of barcodes, and cleaned using
64 Agencourt AMPure XP beads. Each pool was size-selected between 344-408 bp using a Pippin Prep
65 electrophoresis system (Sage Science). Each size-selected pool was PCR amplified in 6-8 reactions
66 each composed of 6 μ l of DNA template, 2 μ l of each ddRAD primer, 0.2 μ l Phusion *Taq*
67 polymerase (NEB), 0.4 μ l dNTPs (10 mM; NEB), 4 μ l 5x Phusion HF buffer (NEB) and 5.4 μ l
68 nuclease-free water. PCR reactions were run on a Veriti 96-Well Fast Thermal Cycler (Applied
69 Biosystems) at 98°C for 30s, 12 cycles of 98°C for 15s, 65°C for 30s, 72°C for 30s, and final 72°C
70 for 7 min. PCR reactions for each pool were combined and cleaned using Agencourt AMPure XP
71 beads. Each cleaned pool was diluted to 4 nM and four pools with compatible barcodes were

72 combined to produce libraries of 48 uniquely barcoded samples. Two ddRAD sequence libraries
73 were prepared in this way. All libraries were sequenced to 100 bp, paired-end on an Illumina HiSeq
74 2500 (one lane per library) at the MRC Clinical Sciences Centre, Imperial College London. This
75 generated an average of 6,186,469 reads per sample. Genotyping of ddRAD data was performed
76 using the *STACKS* pipeline, building upon the catalogue generated in Papadopoulos *et al.* (2019b).
77 To expand the catalogue to encompass haplotypes present in both *Howea* species, samples were
78 assembled into loci using *USTACKS* (-m20, -M3) and these stacks were merged into the existing
79 catalogue allowing 3 mismatches between loci in different individuals. To genotype individuals,
80 loci were assembled with lower coverage in *USTACKS* (-m5, -M3) and these stacks were mapped to
81 the catalogue loci.

82

83 DEMOGRAPHIC MODELLING

84 For the demographic analysis, the Far Flats individuals were genotyped at 4,581 loci (23,109 SNPs)
85 with fewer than 20 missing individuals (minor allele frequency = 0.05). To account for missing
86 data, we projected the number of individuals down to 36/54 *H. forsteriana* and 23/42 *H.*
87 *belmoreana* and calculated the joint folded site-frequency-spectrum using *δaδi* (Gutenkunst *et al.*
88 2009). We then inferred the demographic history of *Howea* from the site-frequency-spectrum using
89 two methods: (i) the composite-likelihood framework implemented by *fastsimcoal2* (Excoffier *et al.*
90 2013) and (ii) the diffusion approximation approach implemented in a modified version of *δaδi*
91 (Gutenkunst *et al.* 2009; Tine *et al.* 2014). This modified version of *δaδi* accounts for variation in
92 the rate of gene flow across the genome by dividing the genome into two types of loci (in P and 1-P
93 proportions) with potentially different migration rates.

94

95 For *fastsimcoal2*, we estimated parameters 60 times for each of 10 demographic models to
96 determine the combination of parameters with the highest likelihood. These models are shown on
97 Figure 1a and assume either no population growth, or population growth: model 1 - speciation
98 without gene flow; model 2 - speciation with recent gene flow following secondary contact; model
99 3 - speciation with initial gene flow; model 4 - speciation with constant gene flow; and model 5 -
100 speciation with two distinct periods of migration where gene flow may vary. Model fit was assessed
101 using the Akaike Information Criterion (AIC). Non-parametric bootstrapping (100 simulated
102 datasets under the best model with 10 rounds of parameter estimation for each simulation) was used
103 to estimate 95% confidence intervals for each parameter for the best model.

104

105 For *δaδi*, we compared the fit of the same 10 models as above plus another eight, that is, models 2-5
106 with and without population growth, but also including heterogeneous rates of migration across the

107 genome. We ran two rounds of simulated annealing (one hot and one cold) followed by a final
108 round of Broyden–Fletcher–Goldfarb–Shanno optimisation. For each of the 18 models, we
109 performed a minimum of 30 runs to ensure thorough estimation of the maximum likelihood and
110 used AIC to perform model selection. For the best fitting model, we then ran 30 bootstrap replicates
111 (ensuring that each replicate had at least 10 runs) using the built-in *δaδi* procedure to get confidence
112 intervals around the parameter estimates.

113

114 To calibrate the demographic models in *fastsimcoal2* and *δaδi*, we estimated the substitution rate in
115 *Howea*. We first built a phylogenetic tree using genome wide data for all available palm species.
116 Transcriptome assemblies were obtained for *H. forsteriana* (Dunning et al 2016), *Elaeis guineensis*
117 and *E. oleifera* (African and American oil palms; Singh et al. 2013). Short read transcript data for
118 *Phoenix dactylifera* (date palm; Al-Mssallem et al. 2013) were obtained from the Short Read
119 Archive (Accession Number SRR341952) and a transcriptome assembly was produced using
120 *Trinity* (Grabherr et al. 2011) with default parameters and *min_kmer_cov* set to 2. *TransDecoder*
121 was used to predict open reading frames (ORF) with a minimum length of 100 amino acids, and the
122 longest ORF was selected. Reciprocal blast hits for all four palm species were established by
123 collating reciprocal best blast results from a pair of species with the remaining two species; this
124 rendered an initial set of orthologous alignments for the four species. An *M-coffee* (Wallace et al.
125 2006) pipeline was used to score these alignments. The nucleotide sequences were translated into
126 protein using *t-coffee*, then the protein sequences were aligned with *MAFFT* (Katoh & Toh 2008),
127 *Muscle* (Edgar 2004), *t-coffee* (Notredame et al. 2000) and *k-align* (Lassmann & Sonnhammer
128 2005) and translated back to DNA. Low quality positions with scores lower than 8 were trimmed.
129 Protein alignments were used to guide gap placement in the nucleotide alignments. Finally,
130 *maxalign* was used to check for any misaligned sequences in the alignments (Gouveia-Oliveira et
131 al. 2007). Only alignments where all the orthologous sequences were properly aligned and deemed
132 as high quality were retained and concatenated into a single file. *CODEML* (Yang 2007) was used
133 to calculate 4-fold degenerate sites, which are considered not to be subject to selection. Finally,
134 *MCMCtree* (Yang 2007) was used to build a tree. To calibrate this tree, we used secondary
135 calibrations from the most complete phylogenetic tree of the palm family (Baker & Couvreur 2013).
136 Independent substitution rates were inferred for each branch in the tree, and HKY85 (the most
137 complex substitution model available in *MCMCtree*) was selected. The MCMC chain was run to
138 gather 20,000 samples after convergence had been achieved. The first 10,000 iterations were
139 discarded as a burn-in. Assuming a generation time of 10 years for *Howea* palms (Lord Howe
140 Island Nursery, pers. comm.), we estimated the rate of substitution of the branch leading to *H.*
141 *forsteriana* to be 1.3×10^{-8} mutations per site per generation.

142

143

144 **Results and discussion**

145

146 The demographic analyses provide strong support that *Howea* palms on LHI diverged in the face of
147 ongoing gene flow. We have categorised the demographic models into four groups: (i) no
148 population growth; (ii) population growth; (iii) no growth but heterogeneous migration across the
149 genome; (iv) growth with heterogeneous migration across the genome. Using *fastsimcoal2*, where
150 heterogeneous migration across the genome does not apply, model 5 (divergence with two periods
151 where gene flow may vary) was most likely when no growth was modelled (Table 1). When
152 population growth was permitted, model 5 was within 10 AIC from models 2 and 3, and therefore
153 these three models are undistinguishable (Burnham & Anderson 2002) (Table 1). Using *δadi*, model
154 5 was always the most likely model. However, when growth and heterogeneous migration across
155 the genome were permitted, model 5 was indistinguishable from model 3 (speciation with initial
156 gene flow; Table 1). The more complex growth models had smaller AIC than the no growth
157 counterparts (Table 1), pointing to a period of exponential growth following species divergence. –
158 However, for both *δadi* and *fastsimcoal2* the confidence intervals included the point estimates for
159 all parameters only for the simplest scenario (i.e., without population size changes or heterogeneous
160 rate of genome flow across the genome, Table S1). Other scenarios had greater levels of uncertainty
161 indicating that our data were not sufficient to constrain such complex models and the point
162 estimates for these models are likely to be unreliable. As the two demographic methods produced
163 comparable results for the no growth scenarios, we focus our discussion on these simpler cases. In
164 models including two distinct periods of migration, we did not constrain the earlier migration rate to
165 be higher than the more recent migration rate. Nevertheless, in all of these models the earlier rate of
166 gene flow is estimated to be higher than more recent rates, supporting a reduction in gene flow
167 during speciation. A similar pattern was found during the divergence of *Senecio* on Mt. Etna
168 (Filatov *et al.* 2016).

169

170 Figure 1b shows the detailed scenario for model 5 without growth. Confidence intervals for all
171 parameters for *δadi* were much wider than for *fastsimcoal2* (Table S1), so we mainly discuss results
172 from the latter. To assess the fit of the data to the model, we calculated the likelihood ratio G-
173 statistic (Composite likelihood ratio = 470.2) and compared this to the null distribution of simulated
174 values (Excoffier *et al.* 2013). The observed value is in the tail of the simulated distribution (above
175 the 99.9 percentile; range = 327.4 – 485.9). This is expected as these demographic models are a

176 simplification of the real history of the species, whereas the null distribution is based on data
177 simulated under the simple model (Excoffier *et al.* 2013).

178 Migration was initially two orders of magnitude higher from the smaller *H. forsteriana* population
179 into the larger *H. belmoreana* lineage (proportion of migrants received per generation = 4.00×10^{-4}
180 vs 4.71×10^{-6} ; effective migrants per generation, $N_m = 13.01$ vs 0.27). This initial period was
181 followed by a reduction in gene flow (proportion of migrants received per generation = 1.6×10^{-7} vs
182 3.3×10^{-7} ; $N_m = 0.01$ vs 0.02). This is consistent with *H. forsteriana* being derived from a
183 *belmoreana*-like ancestor that colonised a new habitat in which the *H. belmoreana* genetic
184 background was selected against. The initially high N_m (mean = 6.64) is in the top 6% of values
185 found in other examples speciation with gene flow (range = 0.00 - 25.22; Pinho & Hey 2010;
186 Filatov *et al.* 2016) and the proportion of migrants is similar to that found in the sympatric
187 Nicaraguan cater lake cichlids (7.48×10^{-5} - 8.51×10^{-5} ; Kautt *et al.* 2016). These migration estimates
188 fall below those expected under population genetic definitions of sympatric speciation ($m = 0.5$;
189 Fitzpatrick *et al.* 2008). However, it is important to note that our migration estimates are averages
190 over long periods of time, forced by a model that has an abrupt transition from one population to
191 two populations. If we had a model that allowed a progressive reduction in gene flow, we may have
192 seen values close to 0.5 at the start and then a rapid reduction as the initial barriers were built.
193 Unfortunately, we do not have the data to fit such a model. Of course, it would also have strong
194 heterogeneity across the genome. Based on a generation time of 10 years, we found that the species
195 initially diverged roughly 2.7 million years ago, which is older than previously estimated by
196 phylogenetic analysis of two genes (1.92 +/- 0.52 million years ago; Savolainen *et al.* 2006).
197 Allowing for different rates of migration across the genome (using $\delta a \delta i$) resulted in a more recent
198 divergence time than other models at 1.41 Mya, but this fell outside the bootstrap confidence
199 intervals. Our estimates predate the proposed mid-Pleistocene age of the current calcareous deposits
200 of the island (Brooke 2003; Woodroffe *et al.* 2006; Papadopulos *et al.* 2013a). It is possible that
201 colonisation was on calcareous soils pre-dating those currently on the island, and which would have
202 eroded since then. Alternatively, colonisation of mid-Pleistocene calcareous sites may have taken
203 place after divergence with the initial selection pressures stemming from other sources, such as
204 water availability or salinity (Papadopulos *et al.* 2013b). The timing of the shift to a lower migration
205 rate differs substantially between the methods; *fastsimcoal2* points to a large reduction in migration
206 ~40,000 years after initial divergence, whereas a $\delta a \delta i$ indicates smaller change that took place much
207 more recently (100,000 year after divergence). Estimates of current population sizes (*H. forsteriana*
208 $N_e = 32,510$; *H. belmoreana* $N_e = 57,181$) are similar to the estimated census population size (all
209 LHI *Howea* ~100,000, with 2.7 times as many *H. belmoreana* as *H. forsteriana*) (Savolainen *et al.*
210 2006; Hipperson *et al.* 2016). The *fastsimcoal2* estimated ancestral N_e (424,288) is within the

211 bounds of possibility, but is likely to be an overestimate as variation in coalescence time due to
212 selection in the ancestral species may cause *fastsimcoal2* to explain excess variance by inferring
213 very large N_e . The *δadi* ancestral N_e was an order of magnitude lower (12,570), and it is therefore
214 unclear which estimate is best. It seems likely that small initial colonising group came from a large
215 mainland population and then grew rapidly, but this not possible to determine with our analyses.
216 The inconsistent results for the more complex models show that our analyses are limited by the
217 data. Refinements could be made using whole genome re-sequencing data, although the large
218 genome size and complexity makes this challenging. Additionally, more detailed data would allow
219 the inclusion of other parameters in more complex frameworks that have emerged recently - such as
220 that of Roux *et al* (2016), which models heterogeneity in coalescence times due to selection.
221 Currently, reproductive isolation between the species is strong, although not complete given that
222 occasional fertile hybrids are formed (Babik *et al.* 2009). Our results are consistent with the idea
223 that initial local adaptation and post-zygotic isolation were supplemented by the rapid completion of
224 pre-zygotic isolation through flowering time differences (Papadopoulos *et al.* 2013b, 2014).
225 Furthermore, given the supporting evidence these results allow us to rule out continuous absence of
226 gene flow, even though LHI was larger at the time of speciation (Papadopoulos *et al.* 2011, 2013a;
227 Linklater *et al.* 2018). Unlike recent genomic reanalyses of classic cases of speciation in sympatry,
228 our results support the proposition that *Howea* palms must have diverged with continuous gene
229 flow. We note, however, that genomic data by themselves may only permit the rejection of the
230 simplest form of allopatry (Yang *et al.* 2017). In this sense, our demographic analyses should be
231 seen in concert with other lines of evidence such as the geological history of LHI, lack of
232 population structure on LHI, and the finding of candidate reproductive isolation genes (Dunning *et al.*
233 *al.* 2016). Furthermore, our analyses indicate that divergence may have predated the origin of the
234 calcarenite soils on LHI, and therefore identifying candidate ‘speciation genes’ with functions
235 related to drought and salt tolerance may be more important than extremes of pH.

236

237

238 REFERENCES

239

240 Al-Mssallem, I.S., Hu, S., Zhang, X., Lin, Q., Liu, W., Tan, J., *et al.* (2013). Genome sequence of
241 the date palm *Phoenix dactylifera* L. *Nat. Commun.* 4: 2274. Nature Publishing Group.

242 Babik, W., Butlin, R.K., Baker, W.J., Papadopoulos, A.S.T., Boulesteix, M., Anstett, M.C., *et al.*
243 (2009). How sympatric is speciation in the *Howea* palms of Lord Howe Island? *Mol. Ecol.* 18:
244 3629–3638.

245 Baker, W.J. & Couvreur, T.L.P. (2013). Global biogeography and diversification of palms sheds

246 light on the evolution of tropical lineages. I. Historical biogeography. *J. Biogeogr.* 40: 274–
247 285.

248 Brooke, B. (2003). Quaternary calcarenite stratigraphy on Lord Howe Island, southwestern Pacific
249 Ocean and the record of coastal carbonate deposition. *Quat. Sci. Rev.* 22: 859–880.

250 Burnham, K.P. & Anderson, D.R. (2002). *Model selection and multimodel inference: a practical*
251 *information-theoretic approach*. Springer, New York.

252 Butlin, R.K., Galindo, J. & Grahame, J.W. (2008). Sympatric, parapatric or allopatric: the most
253 important way to classify speciation? *Philos. Trans. R. Soc. B Biol. Sci.* 363: 2997–3007.

254 Coyne, J.A. (2011). Speciation in a small space. *Proc. Natl. Acad. Sci.* 108: 12975–12976.

255 Cruickshank, T.E. & Hahn, M.W. (2014). Reanalysis suggests that genomic islands of speciation
256 are due to reduced diversity, not reduced gene flow. *Mol. Ecol.* 23: 3133–3157.

257 Dunning, L.T., H. Hipperson, Baker, W.J., Butlin, R.K., Devaux, C., Hutton, I., *et al.* (2016).
258 Ecological speciation in sympatric palms: 1. Gene expression, selection and pleiotropy. *J.*
259 *Evol. Biol.* 29: 1472–1487.

260 Edgar, R.C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high
261 throughput. *Nucleic Acids Res.* 32: 1792–1797.

262 Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V.C. & Foll, M. (2013). Robust
263 Demographic Inference from Genomic and SNP Data. *PLoS Genet.* 9.

264 Filatov, D.A., Osborne, O.G. & Papadopoulos, A.S.T. (2016). Demographic history of speciation in a
265 *Senecio* altitudinal hybrid zone on Mt. Etna. *Mol. Ecol.* 25: 2467–2481.

266 Fitzpatrick, B.M., Fordyce, J.A. & Gavrillets, S. (2008). What, if anything, is sympatric speciation?
267 *J. Evol. Biol.* 21: 1452–1459.

268 Gavrillets, S. (2004). *Fitness Landscapes and the Origin of Species*. Princeton University Press,
269 Princeton, NJ.

270 Gouveia-Oliveira, R., Sackett, P.W. & Pedersen, A.G. (2007). MaxAlign: maximizing usable data
271 in an alignment. *BMC Bioinformatics* 8: 312.

272 Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., *et al.* (2011). Full-
273 length transcriptome assembly from RNA-Seq data without a reference genome. *Nat.*
274 *Biotechnol.* 29: 644–652.

275 Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H. & Bustamante, C.D. (2009). Inferring the
276 joint demographic history of multiple populations from multidimensional SNP frequency data.
277 *PLoS Genet.* 5.

278 Hipperson, H., Dunning, L.T., Baker, W.J., Butlin, R.K., Hutton, I., Papadopoulos, A.S.T., *et al.*
279 (2016). Ecological speciation in sympatric palms: 2. Pre- and post-zygotic isolation. *J. Evol.*
280 *Biol.* 29: 2143–2156.

281 Katoh, K. & Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment
282 program. *Brief. Bioinform.* 9: 286–298.

283 Kautt, A.F., Machado-Schiaffino, G. & Meyer, A. (2016). Multispecies Outcomes of Sympatric
284 Speciation after Admixture with the Source Population in Two Radiations of Nicaraguan
285 Crater Lake Cichlids. *PLoS Genet.* 12: 1–33.

286 Lassmann, T. & Sonnhammer, E.L.L. (2005). Kalign--an accurate and fast multiple sequence
287 alignment algorithm. *BMC Bioinformatics* 6: 298.

288 Linklater, M., Hamylton, S.M., Brooke, B.P., Nichol, S.L., Jordan, A.R. & Woodroffe, C.D. (2018).
289 Development of a Seamless , High-Resolution Bathymetric Model to Compare Reef
290 Morphology around the Subtropical Island Shelves of Lord Howe Island and Balls Pyramid ,
291 Southwest Pacific Ocean. *Geosciences* 8: 11.

292 Mallet, J., Meyer, A., Nosil, P. & Feder, J.L. (2009). Space, sympatry and speciation. *J. Evol. Biol.*
293 22: 2332–2341.

294 Martin, C.H., Cutler, J.S., Friel, J.P., Touokong, C.D., Coop, G. & Wainwright, P.C. (2015).
295 Complex histories of repeated gene flow in Cameroon crater lake cichlids cast doubt on one of
296 the clearest examples of sympatric speciation. *Evolution (N. Y.)*. 69: 1406–1422.

297 Notredame, C., Higgins, D.G. & Heringa, J. (2000). T-Coffee: A novel method for fast and accurate
298 multiple sequence alignment. *J. Mol. Biol.* 302: 205–17.

299 Papadopulos, A.S.T., Baker, W.J., Crayn, D., Butlin, R.K., Kynast, R.G., Hutton, I., *et al.* (2011).
300 Speciation with gene flow on Lord Howe Island. *Proc. Natl. Acad. Sci. U. S. A.* 108: 13188–
301 13193.

302 Papadopulos, A.S.T., Baker, W.J. & Savolainen, V. (2013a). Sympatric Speciation of Island Plants:
303 The Natural Laboratory of Lord Howe Island. In: *Speciation: Natural Processes, Genetics &*
304 *Biodiversity* (P. Michalak, ed). Nova Science.

305 Papadopulos, A.S.T., Kaye, M., Devaux, C., Hipperson, H., Lighten, J., Dunning, L.T., *et al.*
306 (2014). Evaluation of genetic isolation within an island flora reveals unusually widespread
307 local adaptation and supports sympatric speciation. *Philos. Trans. R. Soc. B Biol. Sci.* 369:
308 20130342.

309 Papadopulos, A.S.T., Price, Z., Devaux, C., Hipperson, H., Smadja, C.M., Hutton, I., *et al.* (2013b).
310 A comparative analysis of the mechanisms underlying speciation on Lord Howe Island. *J.*
311 *Evol. Biol.* 26: 733–745.

312 Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S. & Hoekstra, H.E. (2012). Double digest
313 RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and
314 non-model species. *PLoS One* 7: e37135. Public Library of Science.

315 Pinho, C. & Hey, J. (2010). Divergence with Gene Flow: Models and Data. *Annu. Rev. Ecol. Evol.*

316 *Syst.* 41: 215–230.

317 Renaut, S., Grassa, C.J., Yeaman, S., Moyers, B.T., Lai, Z., Kane, N.C., *et al.* (2013). Genomic
318 islands of divergence are not affected by geography of speciation in sunflowers. *Nat. Commun.*
319 4: 1827.

320 Roux, C., Fraïsse, C., Romiguier, J., Anciaux, Y., Galtier, N. & Bierne, N. (2016). Shedding Light
321 on the Grey Zone of Speciation along a Continuum of Genomic Divergence. *PLoS Biol.* 14: 1–
322 22.

323 Savolainen, V., Anstett, M.-C., Lexer, C., Hutton, I., Clarkson, J.J., Norup, M. V, *et al.* (2006).
324 Sympatric speciation in palms on an oceanic island. *Nature* 441: 210–213.

325 Singh, R., Ong-Abdullah, M., Low, E.-T.L., Manaf, M.A.A., Rosli, R., Nookiah, R., *et al.* (2013).
326 Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds.
327 *Nature* 500: 335–9. Nature Publishing Group.

328 Tine, M., Kuhl, H., Gagnaire, P.A., Louro, B., Desmarais, E., Martins, R.S.T., *et al.* (2014).
329 European sea bass genome and its variation provide insights into adaptation to euryhalinity and
330 speciation. *Nat. Commun.* 5: 5770.

331 Wallace, I.M., O’Sullivan, O., Higgins, D.G. & Notredame, C. (2006). M-Coffee: combining
332 multiple sequence alignment methods with T-Coffee. *Nucleic Acids Res.* 34: 1692–9.

333 Woodroffe, C.D., Kennedy, D.M., Brooke, B.P. & Dickson, M.E. (2006). Geomorphological
334 evolution of Lord Howe Island and carbonate production at the latitudinal limit to reef growth.
335 *J. Coast. Res.* 22: 188–201. Coastal Education and Research Foundation.

336 Yang, M., He, Z. & Shi, S. (2017). Can genomic data alone tell us whether speciation happened
337 with gene flow ? *Mol. Ecol.* 26: 2845–2849.

338 Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* 24:
339 1586–1591.

340

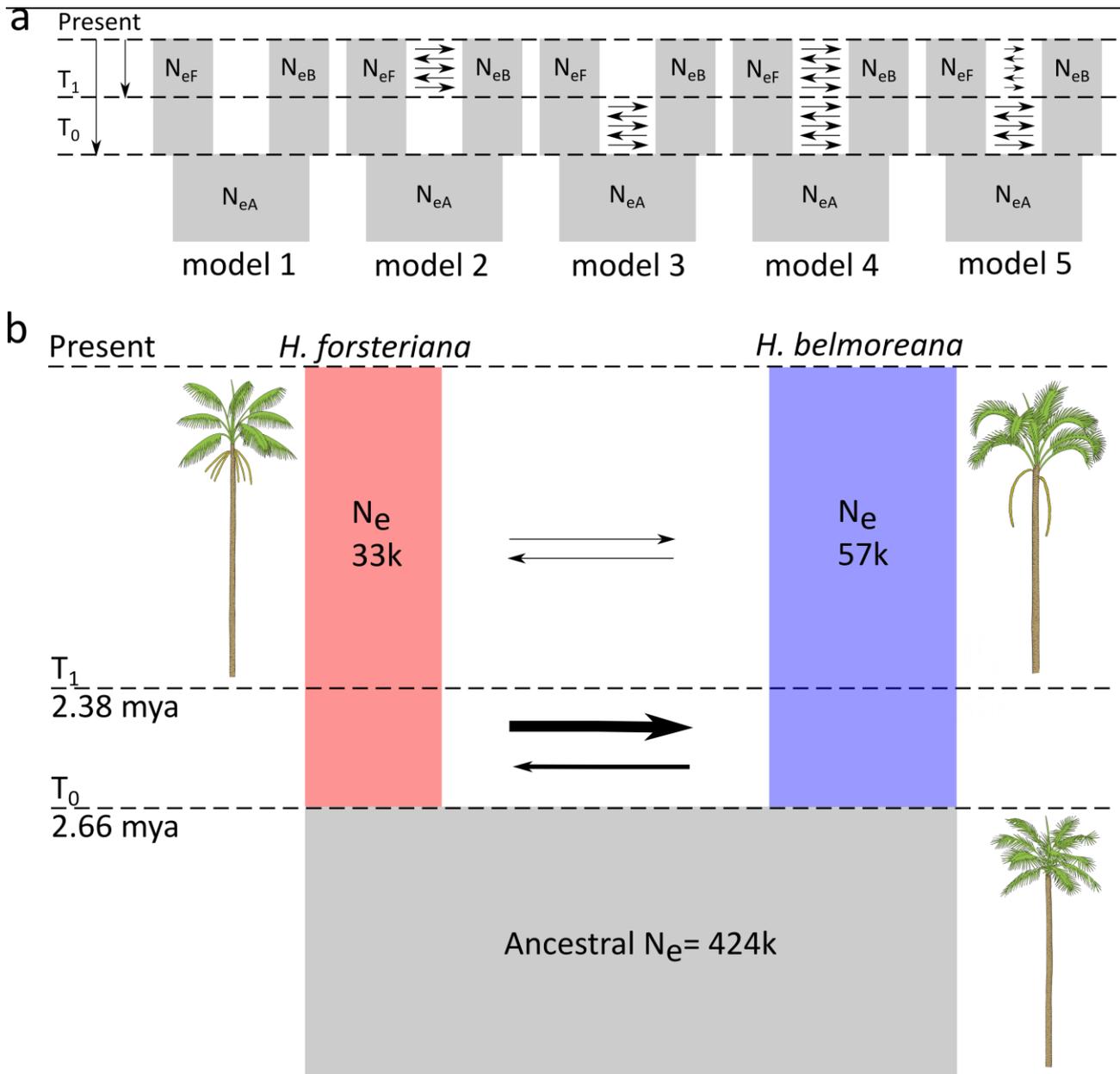
341

342

Table 1. Summary statistics for model selection. The most likely models within 10 AIC are highlighted in bold.

Model	Method	Max. Ln (likelihood)	No. of parameter s	AIC	Delta AIC
1 (speciation without gene flow)	<i>fastsimcoal2</i>	-175730	4	351467	3092
2 (speciation with recent gene flow following secondary contact)	<i>fastsimcoal2</i>	-174992	7	349997	1622
3 (speciation with initial gene flow)	<i>fastsimcoal2</i>	-175522	7	350109	1734
4 (speciation with constant gene flow)	<i>fastsimcoal2</i>	-175049	6	351059	2683
5 (species divergence with two periods where gene flow may vary)	<i>fastsimcoal2</i>	-174955	9	349928	1552
1 + growth	<i>fastsimcoal2</i>	-174898	6	349808	1433
2 + growth	<i>fastsimcoal2</i>	-174179	9	348375	0
3 + growth	<i>fastsimcoal2</i>	-174181	9	348377	2
4 + growth	<i>fastsimcoal2</i>	-174265	8	348548	173
5 + growth	<i>fastsimcoal2</i>	-174180	11	348382	7
1	<i>δaδi</i>	-2602.12	4	5210	2705
2	<i>δaδi</i>	-2051.04	7	4114	1609
3	<i>δaδi</i>	-2055.46	7	4123	1617
4	<i>δaδi</i>	-2067.86	6	4146	1640
5	<i>δaδi</i>	-2012.61	9	4041	1536
1 + growth	<i>δaδi</i>	-1747.95	6	3506	1000
2 + growth	<i>δaδi</i>	-1384.78	9	2786	280
3 + growth	<i>δaδi</i>	-1427.55	9	2871	366
4 + growth	<i>δaδi</i>	-1431.38	8	2877	371
5 + growth	<i>δaδi</i>	-1377.74	11	2775	270
2 + heterogeneous M	<i>δaδi</i>	-1741.57	10	3501	996
3 + heterogeneous M	<i>δaδi</i>	-1786.64	10	3591	1086
4 + heterogeneous M	<i>δaδi</i>	-1785.31	9	3587	1081
5 + heterogeneous M	<i>δaδi</i>	-1730.07	14	3486	981
2 + growth + heterogeneous M	<i>δaδi</i>	-1255.63	12	2533	28
3 + growth + heterogeneous M	<i>δaδi</i>	-1241.78	12	2506	0
4 + growth + heterogeneous M	<i>δaδi</i>	-1256.03	11	2532	26

5 + growth + heterogeneous M $\delta a d i$ -1239.00 16 2508 2



343

344 **Fig. 1.** Coalescence analyses of demography in *Howea*. (a) Five models were tested, either
 345 assuming constant population sizes, allowing for exponential population growth through time, or
 346 heterogeneous migration across the genome (18 scenarios in total). Model 1: speciation without
 347 gene flow; 2: speciation with recent gene flow following secondary contact; 3: speciation with
 348 initial gene flow; 4: speciation with constant gene flow; and 5: species divergence with two periods
 349 where gene flow may vary. (b) Parameter estimates for the best fitting model with no growth
 350 estimated in *fastsimcoal2* (model 5).

351

352

Table S1. Inferred parameters for the selected model 5

Parameter	<i>fastsimcoal2</i>			<i>δaδi</i>		
	Point estimation	Lower CI	Upper CI	Point estimation	Lower CI	Upper CI
Ancestral Ne	424288	125435	547787	12570	11589	19889
<i>H. forsteriana</i> Ne	32510	31192	34006	31806	29809	33687
<i>H. belmoreana</i> Ne	57181	54802	59951	49054	47272	52285
T ₀	2.66	2.56	4.52	2.71	0	40.34
T ₁	2.38	2.10	2.96	0.11	0	2.84
MIG _{F→B} (T ₀ -T ₁)	4.00 x 10 ⁻⁴	5.87 x 10 ⁻⁵	4.21 x 10 ⁻⁴	5.67 x 10 ⁻⁷	2.97 x 10 ⁻⁹⁹	1.45 x 10 ⁻⁵
MIG _{B→F} (T ₀ -T ₁)	4.71 x 10 ⁻⁶	1.32 x 10 ⁻⁵	5.94 x 10 ⁻⁵	7.06 x 10 ⁻¹¹	2.01 x 10 ⁻¹³⁸	1.27 x 10 ⁻⁵
MIG _{F→B} (T ₁ -Present)	1.57 x 10 ⁻⁷	1.57 x 10 ⁻⁷	2.27 x 10 ⁻⁷	1.49 x 10 ⁻⁷	3.50 x 10 ⁻⁷⁰	6.28 x 10 ⁻⁶
MIG _{B→F} (T ₁ -Present)	3.30 x 10 ⁻⁷	2.82 x 10 ⁻⁷	4.12 x 10 ⁻⁷	3.78 x 10 ⁻⁷	5.36 x 10 ⁻³¹	6.76 x 10 ⁻⁶

353

354

355