

1 **Supplementary information for ‘Bridgehead effect and multiple introductions shape the**
2 **global invasion history of a termite’**

3

4 Alexander J. Blumenfeld^{1,†,*}, Pierre-André Eyer^{1,†}, Claudia Husseneder², Jianchu Mo³, Laura
5 N.L. Johnson^{1,4}, Changlu Wang⁵, J. Kenneth Grace⁶, Thomas Chouvenc⁷, Shichen Wang⁸,
6 Edward L. Vargo¹

7

8 ¹Department of Entomology, 2143 TAMU, Texas A&M University, College Station, Texas,
9 77843-2143, USA

10 ²Department of Entomology, Louisiana State University Agricultural Center, Baton Rouge, LA
11 70803, USA

12 ³Ministry of Agriculture Key Lab of Molecular Biology of Crop Pathogens and Insect Pests,
13 Institute of Insect Sciences, College of Agricultural and Biotechnology, Zhejiang University,
14 Zhejiang, PR China

15 ⁴Department of Veterinary Sciences, University of Wyoming, Laramie, Wyoming, 82070, USA

16 ⁵Department of Entomology, Rutgers, The State University of New Jersey, New Brunswick, NJ
17 08901, USA

18 ⁶Department of Plant and Environmental Protection Sciences, University of Hawaii at Manoa,
19 Honolulu, HI, USA

20 ⁷Entomology and Nematology Department, Fort Lauderdale Research and Education Center,
21 Institute of Food and Agricultural Sciences, University of Florida, Fort Lauderdale, FL, USA

22 ⁸Texas A&M AgriLife Genomics and Bioinformatics Service, College Station, TX, USA

23 † Alexander J. Blumenfeld and Pierre-André Eyer should be considered joint first authors

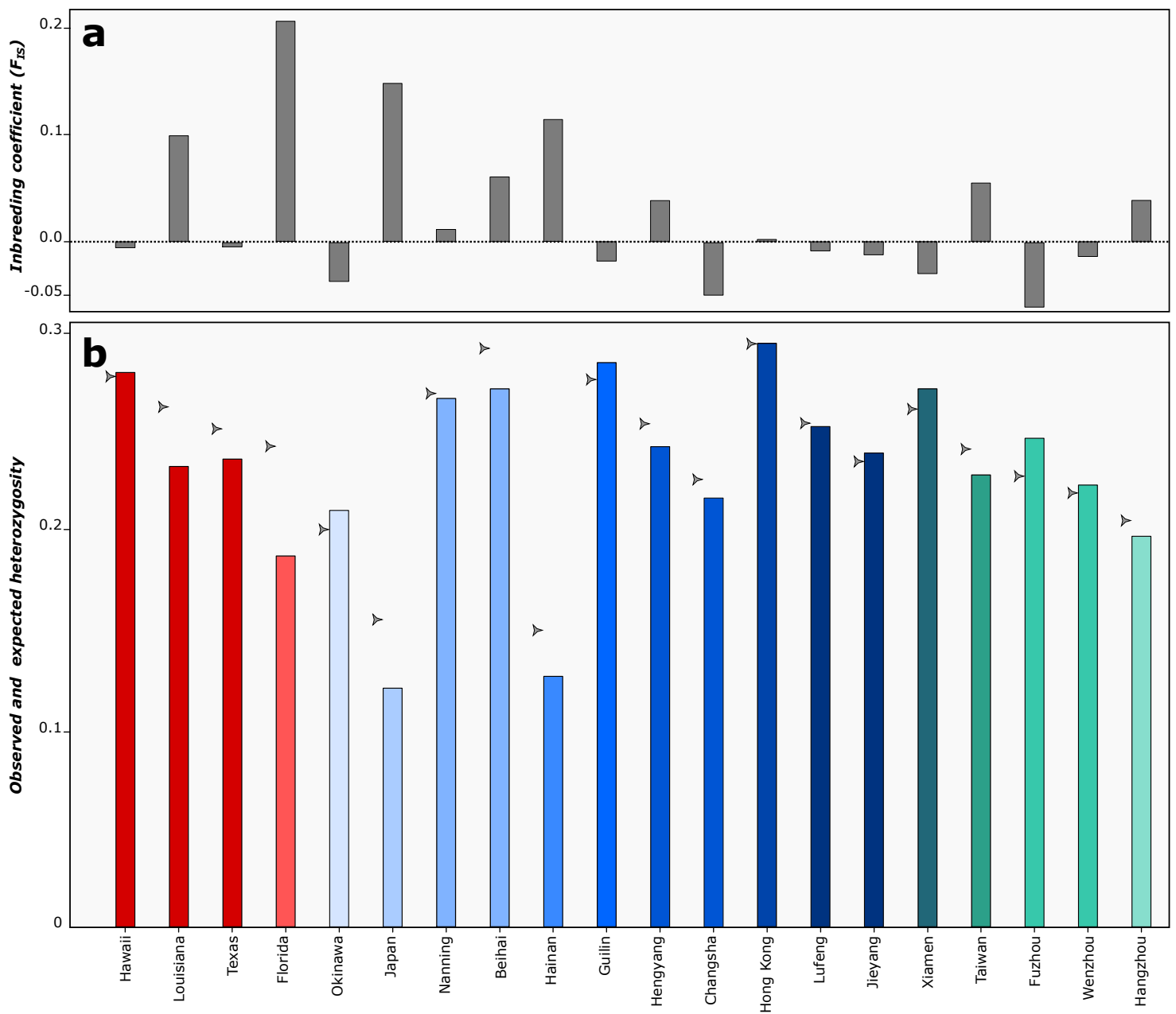
24 *Corresponding author:
25 Alexander J. Blumenfeld
26 Department of Entomology,
27 Texas A&M University,
28 College Station, 77843, Texas, USA
29 e-mail: alex93@tamu.edu

30 **Supplementary Information**

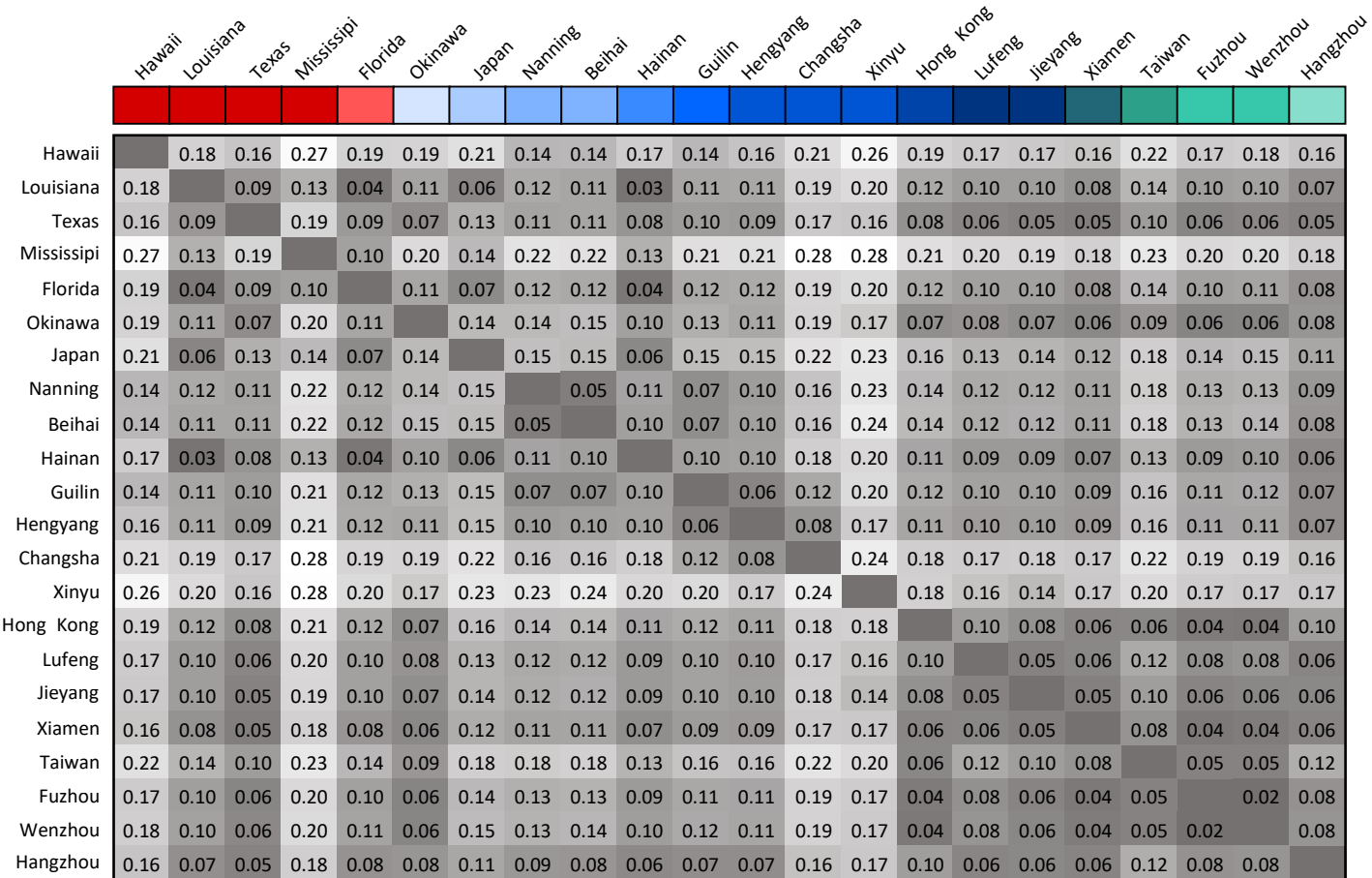
31 **Supplementary Figures 1 – 13**

32 **Supplementary Tables 1 – 12**

33 **Supplementary Methods:** A detailed description of the step-by-step ABC RF analysis is
34 covered, including: 1) priors used for the analyses, 2) graphical representation and random
35 forest votes for each scenario within each step, 3) an overall linear discriminant analysis (LDA)
36 of the simulated datasets for every scenario for each step and 4) parameter estimates for the
37 final invasion model. Additionally, output from the functions *find.clusters* and *optim.a.score*
38 from our DAPC analysis is provided.



Supplementary Figure 1: Genetic diversity statistics. **a** Inbreeding coefficients (F_{IS}) for each population. **b** Mean level of observed heterozygosity for each population. Arrows indicate the expected level of heterozygosity for each population. Each population is colored according to its genetic assignment using fastSTRUCTURE (for $K = 15$).



Supplementary Figure 2: Genetic differentiation. Matrix of F_{ST} values for each pair of populations. Each population is colored according to its genetic assignment using STRUCTURE (for $K = 15$).

39 ABC RF analysis

40 **Supplementary Table 1:** Priors used for each step of the ABC analysis.

Step(s)	Parameter	Lower bound	Upper bound
All	N_i^1	100	10000
	$N_i b^2$	1	100
	db_i^3	0	50
	ra_i^4	0.001	0.999
1,2	$t1^5$	50	300
	$t2$	300	10000
3	$t1$	50	300
	$t2$	50	300
	$t3$	300	10000
4,5	$t1$	50	150
	$t2$	70	300
	$t3$	300	10000
6	$t1$	50	150
	$t2$	50	150
	$t3$	70	300
	$t4$	300	10000

41 ¹Effective population size.

42 ²Effective population size of colonizing force (*i.e.*, initial introduced population).

43 ³Duration of bottleneck event (years).

44 ⁴Admixture coefficient.

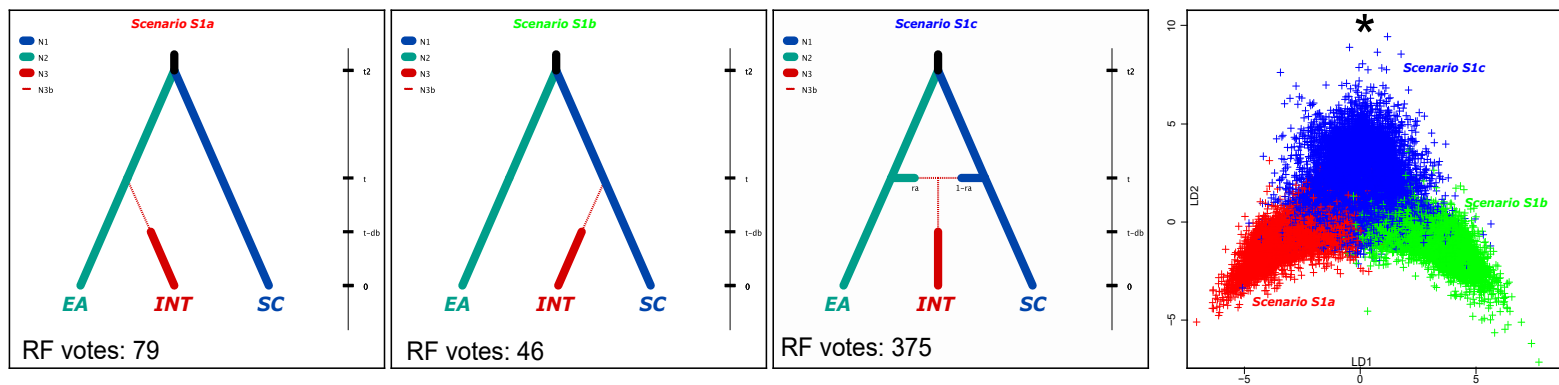
45 ⁵Time in the past that an event occurred (years).

46 *Step 1*

47 The first step aimed at identifying which part(s) of the native range have played a role in the
 48 introduction of *C. formosanus*. The US introduced range was pooled as a single population
 49 (hereafter 'INT'), while the native range and Japan was split into the two regions identified by
 50 the PCA analysis – southcentral China (hereafter called 'SC'; Beihai, Changsha, Guilin,
 51 Hainan, Hengyang and Nanning) and eastern Asia (hereafter called 'EA'; Fuzhou, Hangzhou,
 52 Wenzhou, Hong Kong, Jieyang, Lufeng, Taiwan, Xiamen, Okinawa and mainland Japan). The
 53 Mississippi and Xinyu samples were not included in any group as only one sample was
 54 available for each location (for this and all subsequent steps). Three scenarios were simulated
 55 and tested during this first step. The emergence of the introduced population from a single
 56 introduction event followed by a bottleneck out of EA, SC or a combination and admixture or
 57 both regions was compared. Both EA and SC were found to contribute to the introduced
 58 population and were therefore kept for further analyses. In the LDA plot on the next page,
 59 although the observed dataset (represented by the black star) is separate from most of the
 60 simulations (colored pluses), it is closest to the simulated datasets for the S1c scenario.

61 **Supplementary Table 2:** Prior error rates, RF votes and the posterior probability (for the
 62 winning scenario) for the three scenarios tested in step 1.

Scenario	Prior error rate, %	RF votes	Posterior probability
S1a: Introduction from EA alone	2.95	79	-
S1b: Introduction from SC alone	2.85	46	-
S1c: Introduction from EA + SC	1.27	375	0.55



Supplementary Figure 3: Step 1. The three scenarios compared in step 1, along with the LDA plot of the simulated (colored crosses) and observed (black star) datasets.

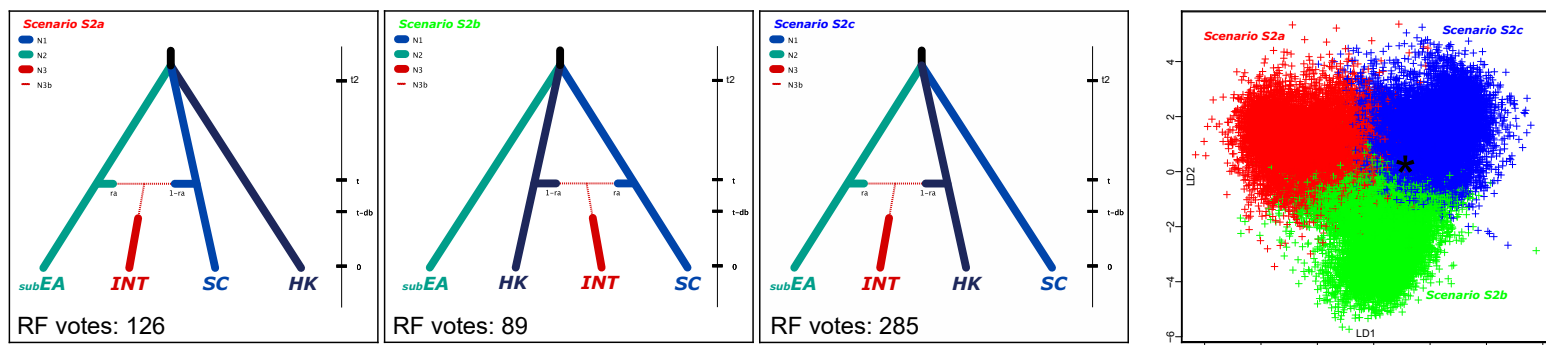
63 *Step 2*

64 The second and third steps aimed at determining whether the Hong Kong region (hereafter
 65 called 'HK'; Hong Kong, Jieyang, Lufeng, Taiwan, Xiamen and Okinawa) is the sole source of
 66 the introduced population of *C. formosanus*. Therefore, HK was separated from the other
 67 eastern Chinese localities (hereafter called sub-eastern Asia '*subEA*'; Fuzhou, Hangzhou,
 68 Wenzhou and mainland Japan). This separation is based on the substructure present within
 69 the EA cluster in the PCA, fastSTRUCTURE and fineRADstructure, as well as the clustering
 70 of the introduced population out of the HK branch in the maximum likelihood tree. The second
 71 step tested for the emergence of the introduced population from admixture between only two
 72 of the three native populations (HK+SC vs. HK+*subEA* vs. *subEA*+SC). The two-population
 73 admixture scenario between HK and *subEA* was found most probable and will be compared
 74 to a three-population admixture scenario in the third step. In the LDA plot on the next page,
 75 the observed dataset is closest to the simulated datasets for the S2c scenario.

76 **Supplementary Table 3:** Prior error rates, RF votes and the posterior probability (for the
 77 winning scenario) for the three scenarios tested in step 2.

Scenario	Prior error rate, %	RF votes	Posterior probability
S2a: Introduction from <i>subEA</i> + SC	2.86	126	-
S2b: Introduction from HK + SC	2.32	89	-
S2c: Introduction from HK+<i>subEA</i>	2.36	285	0.57

78



Supplementary Figure 4: Step 2. The three scenarios compared in step 2, along with the LDA plot of the simulated (colored crosses) and observed (black star) datasets.

79 *Sub-step 2*

80 Sub-steps 2A and 2B sought to determine whether either of the Japanese populations were
 81 driving step 2's result and required separation from the Hong Kong region group (for Okinawa)
 82 or sub-eastern Asian group (for mainland Japan). Okinawa and mainland Japan were split
 83 between the two eastern Asian regions due to their distinct clustering in fastSTRUCTURE,
 84 fineRADstructure and the PCA. Therefore, Okinawa (**Supplementary Table 4**) and mainland
 85 Japan (**Supplementary Table 5**) were separated from their groups and compared against
 86 their remaining groupmates under the best model from step 2. In both sub-steps, neither
 87 Okinawa nor mainland Japan alone explained the invasion history better than their group,
 88 justifying their inclusion in the groups defined in step 2 and not their separation into distinct
 89 groups. In both LDA plots on the next page, the observed datasets are closest to the simulated
 90 datasets for the winning scenarios (Sub2A-a and Sub2B-a).

91 **Supplementary Table 4:** Prior error rates, RF votes and the posterior probability (for the
 92 winning scenario) for the two scenarios tested in sub-step 2A (Okinawa separation).

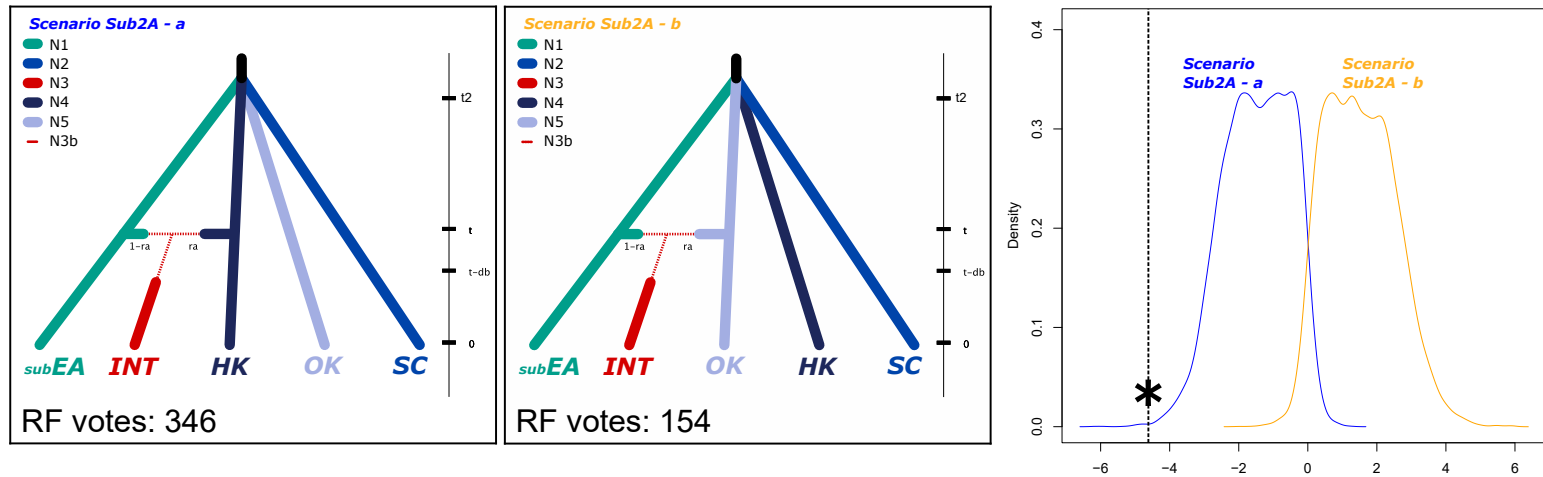
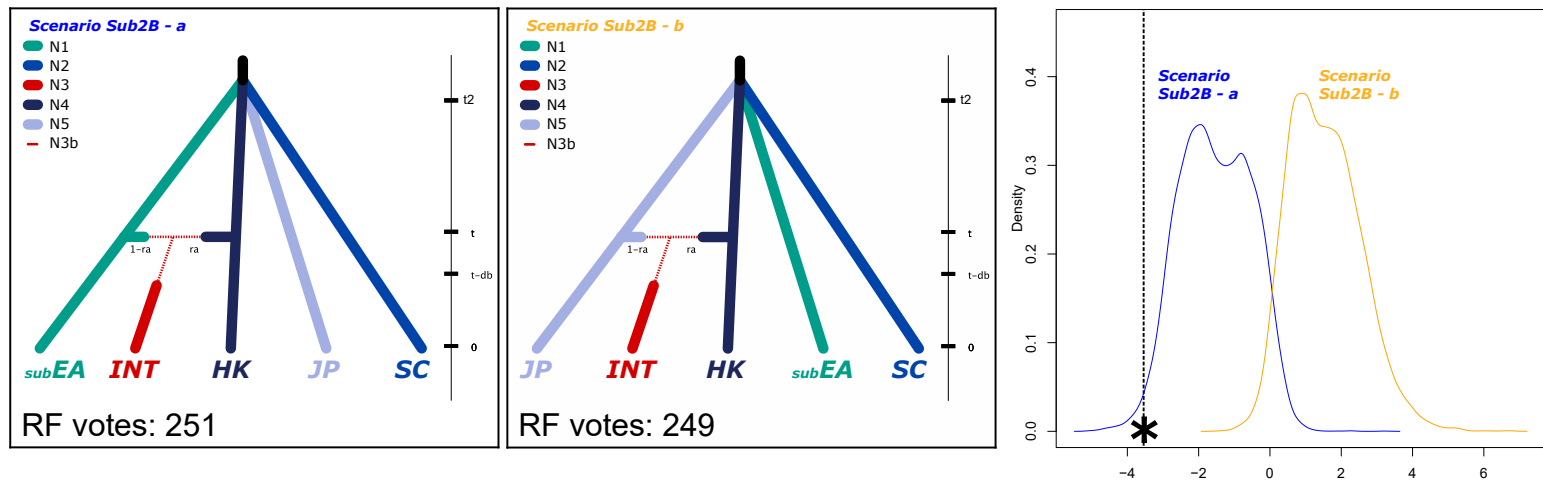
Scenario	Prior error rate, %	RF votes	Posterior probability
Sub2A-a: Introduction from <i>subEA</i> + HK (without Okinawa)	3.27	346	0.58
Sub2A-b: Introduction from <i>subEA</i> + Okinawa	2.99	154	-

93

94 **Supplementary Table 5:** Prior error rates, RF votes and the posterior probability (for the
 95 winning scenario) for the two scenarios tested in sub-step 2B (mainland Japan separation).

Scenario	Prior error rate, %	RF votes	Posterior probability
Sub2B-a: Introduction from <i>subEA</i> (without mainland Japan) + HK	2.89	251	0.64
Sub2B-b: Introduction from mainland Japan + HK	3.83	249	-

96

a**b**

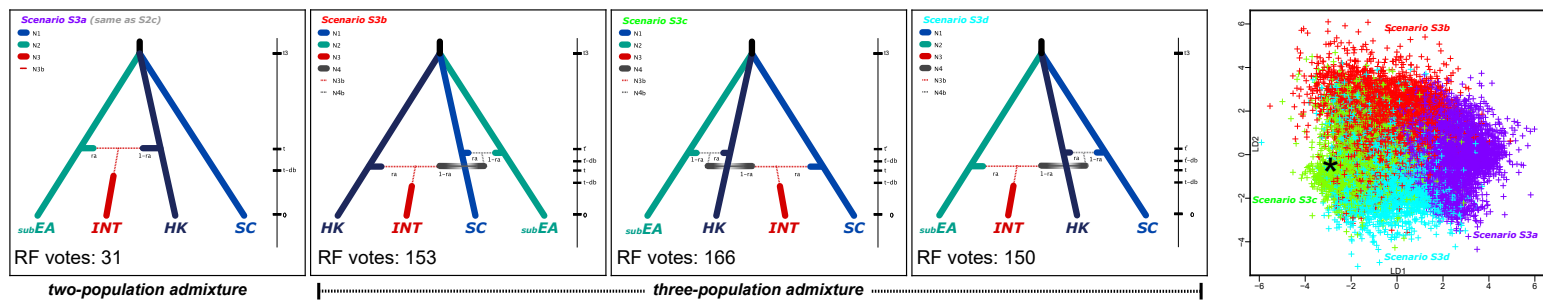
Supplementary Figure 5: Sub-step 2. **a** The two scenarios compared in sub-step 2A, along with the LDA plot of the simulated (blue and gold lines) and observed (black star) datasets. **b** The two scenarios compared in sub-step 2B, along with the LDA plot of the simulated (blue and gold lines) and observed (black star) datasets.

97 *Step 3*

98 The third step then compared the best two-population admixture scenario with scenarios
 99 simulating the origin of the introduced population from admixture between all three populations.
 100 As DIYABC does not allow for the simultaneous admixture of three populations, we divided
 101 the simulations into three scenarios interchanging which two-population admixture comes first
 102 [1) (HK+SC)+*subEA*, 2) (HK+*subEA*)+SC, 3) (*subEA*+SC)+HK]. These three similar scenarios
 103 limit the number of RF votes each can obtain individually, therefore we calculated the number
 104 of RF votes for an overall three-population admixture as the sum of all three scenarios
 105 (cumulative RF votes: 469 votes). This step suggested that the introduced range contained
 106 elements of all three native clusters (consistent with Step 1), therefore all were retained for
 107 further analyses. In the LDA plot on the next page, the observed dataset is closest to the
 108 simulated datasets for the three-population admixture scenarios (S3b, S3c and S3d).

109 **Supplementary Table 6:** Prior error rates, RF votes and the posterior probability (for the
 110 winning scenario) for the four scenarios tested in step 3.

Scenario	Prior error rate, %	RF votes	Posterior probability
S3a: Introduction from best two-population admixture HK + <i>subEA</i>	17.82	31	-
S3b: Introduction from <i>subEA</i> + SC first, then HK	45.76	153	-
S3c: Introduction from HK + <i>subEA</i> first, then SC	42.99	166	0.55
S3d: Introduction from HK + SC first, then <i>subEA</i>	45.06	150	-



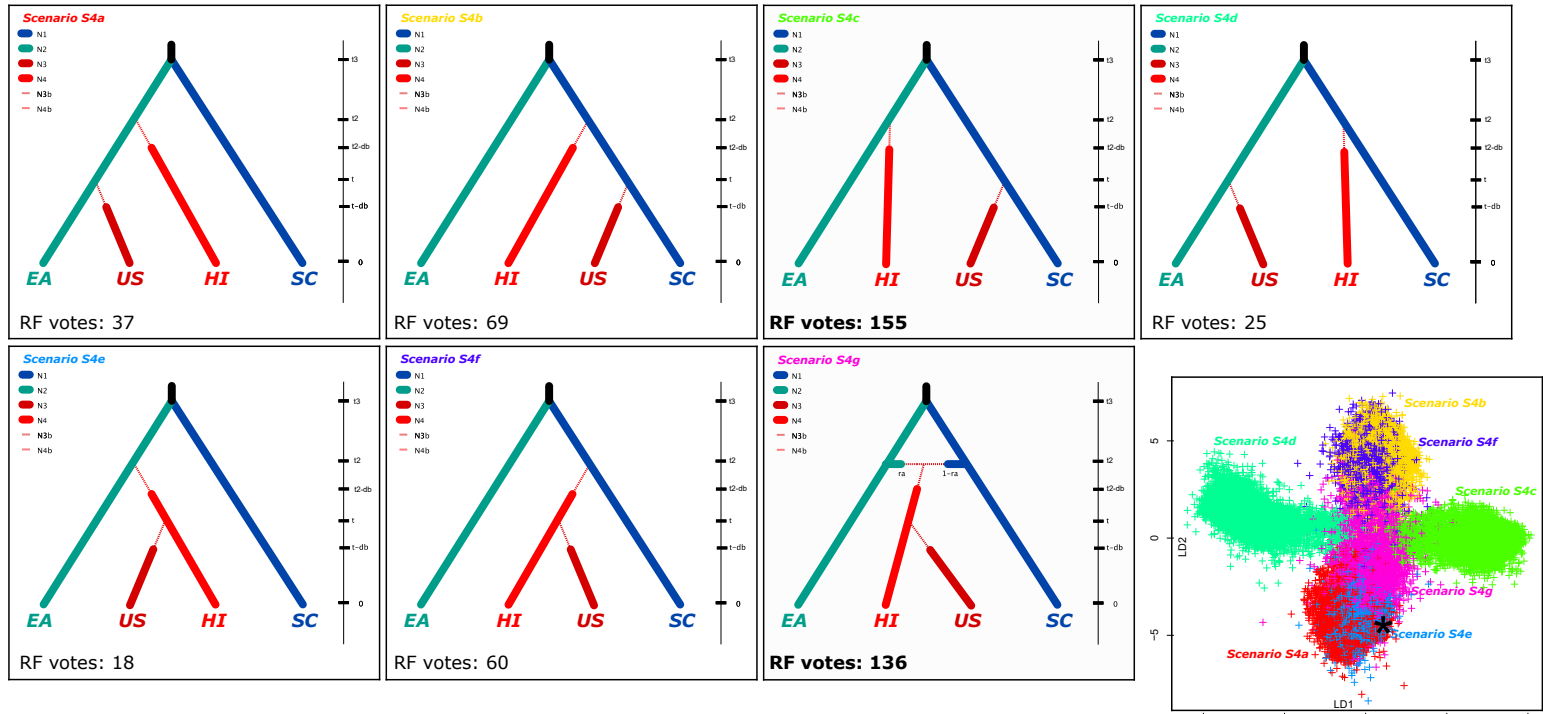
Supplementary Figure 6: Step 3. The four scenarios compared in step 3, along with the LDA plot of the simulated (colored crosses) and observed (black star) datasets.

112 *Step 4*

113 The fourth step aimed at identifying whether the US mainland descended from a Hawaiian
 114 bridgehead or from a separate invasion out of the native range. This step also tested whether
 115 the Hawaii (hereafter called 'HI') and the US mainland (hereafter called 'US') arose from the
 116 same part of Asia (either EA or SC), from a different part, or from admixture between the two
 117 (EA+SC). The HI population was therefore separated from the US population, and HK was
 118 reintegrated into EA to reduce computational effort. Two scenarios obtained a similarly high
 119 number of votes. One scenario (S4c) depicted HI originating from EA, with US originating
 120 independently from SC. The second scenario (S4g) depicted US arising from an already
 121 introduced, admixed HI population through bridgehead. Therefore, we retained both scenarios
 122 for further analyses. In the LDA plot on the next page, the observed dataset is closest to the
 123 simulated datasets for scenarios S4a, S4e and S4g.

124 **Supplementary Table 7:** Prior error rates, RF votes and the posterior probability (for the
 125 winning scenario) for the seven scenarios tested in step 4.

Scenario	Prior error rate, %	RF votes	Posterior probability
S4a: US and HI arose independently from EA	8.67	37	-
S4b: US and HI arose independently from SC	8.54	69	-
S4c: US arose from SC, HI arose from EA	0.3	155	0.71
S4d: US arose from EA, HI arose from SC	0.13	25	-
S4e: US arose from HI bridgehead, HI arose from EA	6.25	18	-
S4f: US arose from HI bridgehead, HI arose from SC	5.32	60	-
S4g: US arose from HI bridgehead, HI arose from admixture between EA and SC	1.23	136	-



Supplementary Figure 7: Step 4. The seven scenarios compared in step 4, along with the LDA plot of the simulated (colored crosses) and observed (black star) datasets.

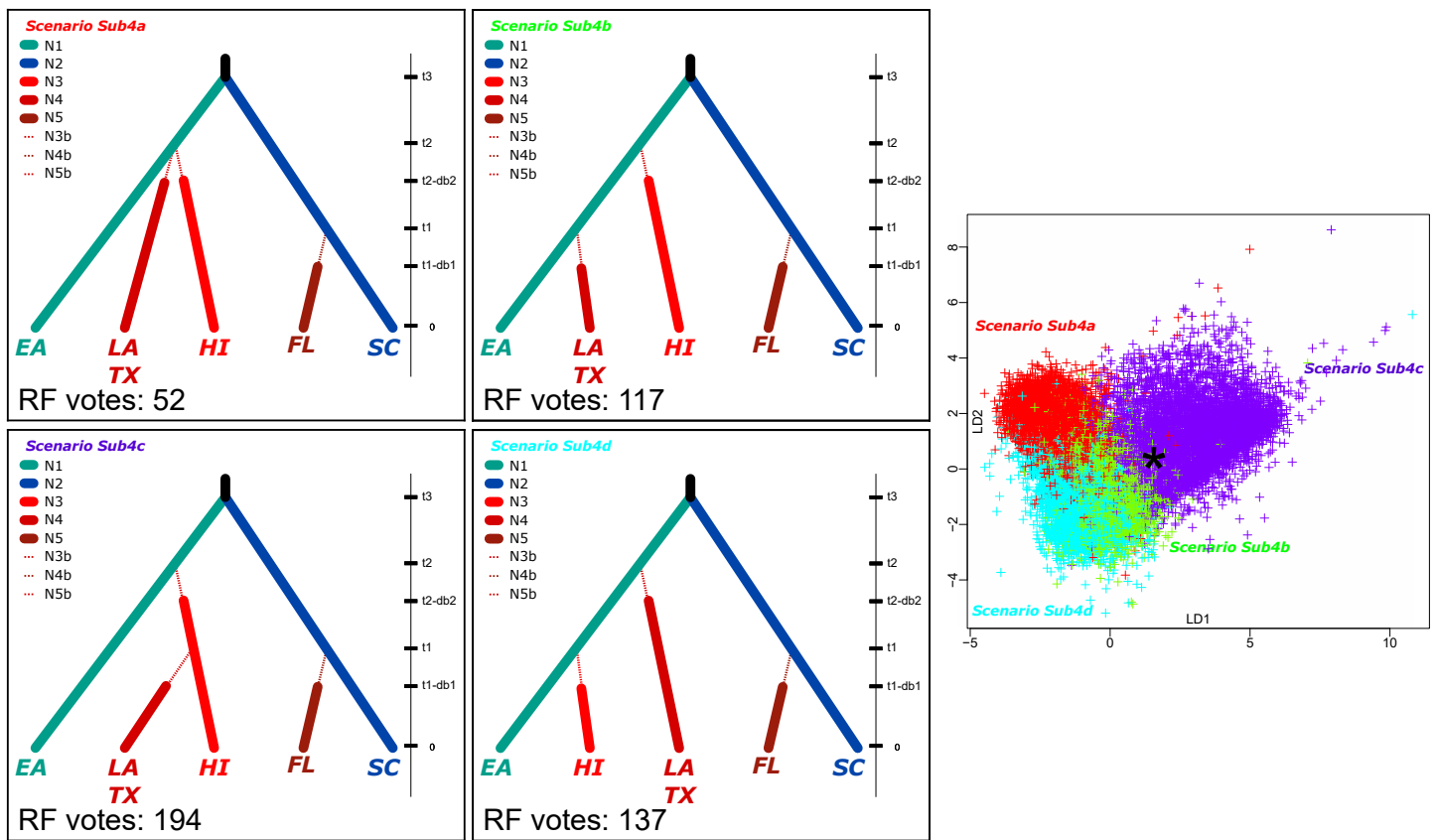
127 *Sub-step 4*

128 Sub-step 4 sought to confirm a Hawaiian bridgehead to the US mainland. Two scenarios in
 129 step 4 obtained a similar number of votes, with the difference between the two most likely
 130 scenarios seemingly driven by the US mainland. Therefore, US was split into Louisiana/Texas
 131 (hereafter called 'LA+TX') and Florida (hereafter called 'FL'). This separation is justified based
 132 on the distinct cluster found among some Florida samples at the best value of K (15) in
 133 fastSTRUCTURE and in the DAPC. Additionally, as HI and LA+TX clustered the same at the
 134 best K and HI appears more related to EA samples (step 4, scenario c), both groups were
 135 descended from EA in all scenarios below, with FL as a constant descended from SC
 136 (reflecting the aforementioned discrepancy). The third scenario (Sub4c) depicting a HI
 137 bridgehead to LA+TX was found most likely, confirming at least one bridgehead from HI to the
 138 US mainland has occurred. In the LDA plot on the next page, the observed dataset is closest
 139 to the simulated datasets for scenario Sub4c.

140 **Supplementary Table 8:** Prior error rates, RF votes and the posterior probability (for the
 141 winning scenario) for the four scenarios tested in sub-step 4.

Scenario	Prior error rate, %	RF votes	Posterior probability
Sub4a: HI and LA+TX arose independently and simultaneously from EA, FL arose from SC	11.37	52	-
Sub4b: HI and LA+TX arose independently from EA (HI first), FL arose from SC	14.59	117	-
Sub4c: HI arose from EA, LA+TX arose from HI bridgehead, FL arose from SC	2.61	194	0.32
Sub4d: HI and LA+TX arose independently from EA (LA+TX first), FL arose from SC	15.29	137	-

142
 143



Supplementary Figure 8: Sub-step 4. The four scenarios compared in sub-step 4, along with the LDA plot of the simulated (colored crosses) and observed (black star) datasets.

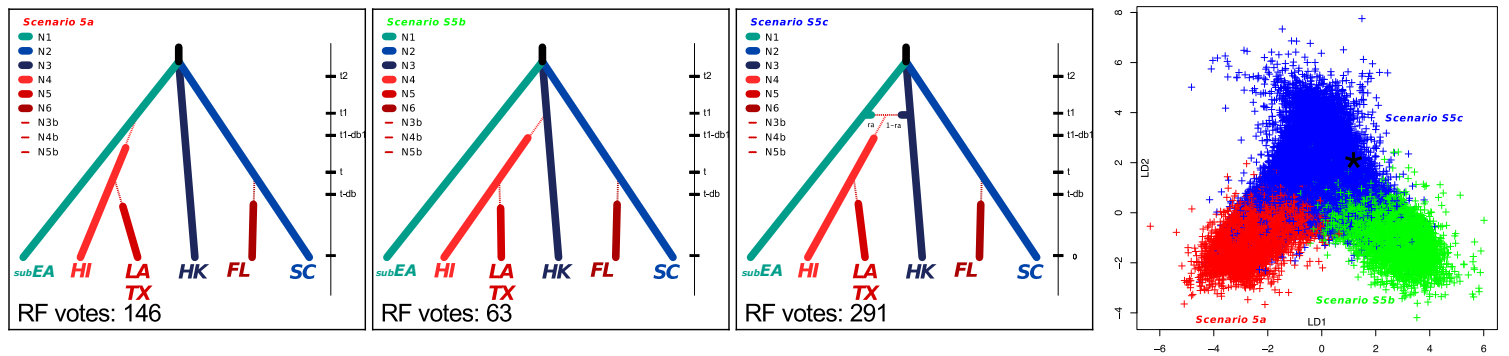
144 *Step 5*

145 Similar to step 3 above, the fifth step further investigated whether HK and *subEA* played distinct
 146 roles in the introduction of *C. formosanus* to HI. This step used the most likely scenario
 147 obtained from sub-step 4 (HI bridgehead to LA+TX) and again split the overall EA region into
 148 HK and *subEA* (and kept FL as a constant descended from SC). This step confirmed that both
 149 sub-regions of eastern Asia played a role in the invasion of HI. In the LDA plot on the next
 150 page, the observed dataset is closest to the simulated datasets for scenario S5c.

151 **Supplementary Table 9:** Prior error rates, RF votes and the posterior probability (for the
 152 winning scenario) for the three scenarios tested in step 5.

Scenario	Prior error rate, %	RF votes	Posterior probability
S5a: HI arose from <i>subEA</i> , LA+TX arose from HI bridgehead, FL arose from SC	7.44	146	-
S5b: HI arose from HK, LA+TX from HI bridgehead, FL arose from SC	6.83	63	-
S5c: HI arose from admixture between <i>subEA</i> and HK, LA+TX arose from HI bridgehead, FL arose from SC	3.62	291	0.38

153



Supplementary Figure 9: Step 5. The three scenarios compared in step 5, along with the LDA plot of the simulated (colored crosses) and observed (black star) datasets.

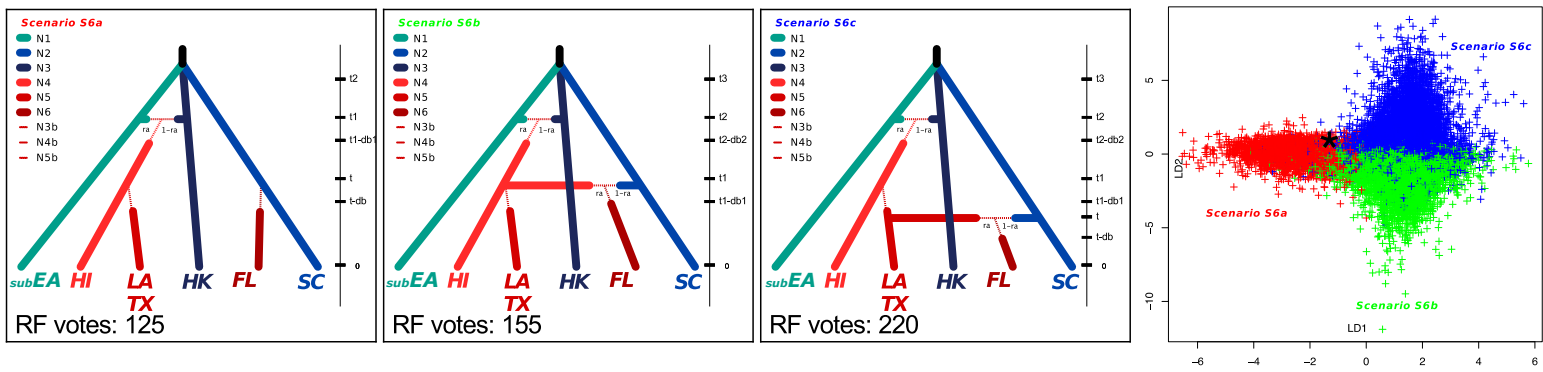
154 *Step 6*

155 This sixth step investigated whether the FL introduction event is connected with the invasion
 156 of HI and LA+TX, or separate. Therefore, we compared the best scenario from step 5 (S5c)
 157 with bridgehead introduction events from either HI or LA+TX that admixed with a separate
 158 introduction event from SC. The SC influence in Florida (as opposed to LA+TX) is supported
 159 by the distinct clustering of some Florida samples at best K in fastSTRUCTURE, as well as
 160 results from step 4 above. The sixth step found it most likely that FL is linked with the invasion
 161 of HI and LA+TX, with FL resulting from admixture between LA+TX and SC. In the LDA plot
 162 on the next page, the observed dataset is near the midpoint of all three scenarios.

163 **Supplementary Table 10:** Prior error rates, RF votes and the posterior probability (for the
 164 winning scenario) for the three scenarios tested in step 6.

Scenario	Prior error rate, %	RF votes	Posterior probability
S6a: HI arose from admixture between <i>subEA</i> and HK, LA+TX arose from HI bridgehead, FL arose from SC	11.12	125	-
S6b: HI arose from admixture between <i>subEA</i> and HK, LA+TX arose from HI bridgehead, FL arose from admixture between HI and SC	19.78	155	-
S6c: HI arose from admixture between <i>subEA</i> and HK, LA+TX arose from HI bridgehead, FL arose from admixture between LA+TX and SC	13.72	220	0.44

165



Supplementary Figure 10: Step 6. The three scenarios compared in step 6, along with the LDA plot of the simulated (colored crosses) and observed (black star) datasets.

166 **Supplementary Table 11:** Parameter estimates for the final invasion model (S6c).

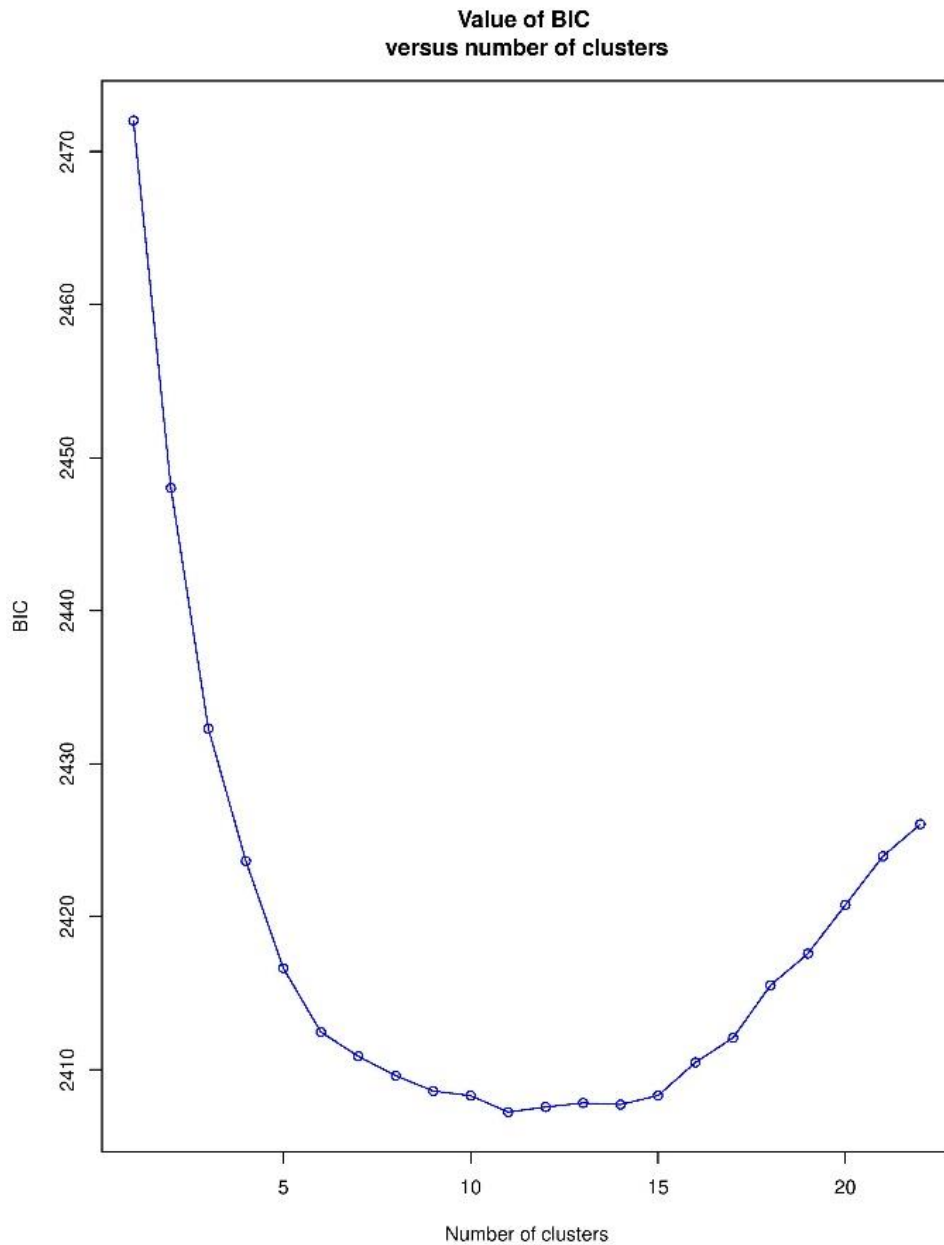
Population	Parameter	Estimate
-	t4	3700
Hong Kong region	N1	5868
Southcentral China	N2	7229
Sub-eastern Asia	N3	5114
Hawaii	N4	4962
	t3	138
	N4b	54
	db3	18
	ra2	0.52
Louisiana+Texas	N5	4671
	t2	98
	N5b	52
	db2	16
Florida	N6	4976
	t1	87
	N6b	52
	db1	18
	ra1	0.49

167
168

169 DAPC analysis

170 *Number of genetic clusters*

171 We utilized the Bayesian information criterion (BIC) to select the most likely number of genetic
172 clusters. After 15 clusters, the BIC begins to dramatically increase, supporting our selection of
173 15 clusters.



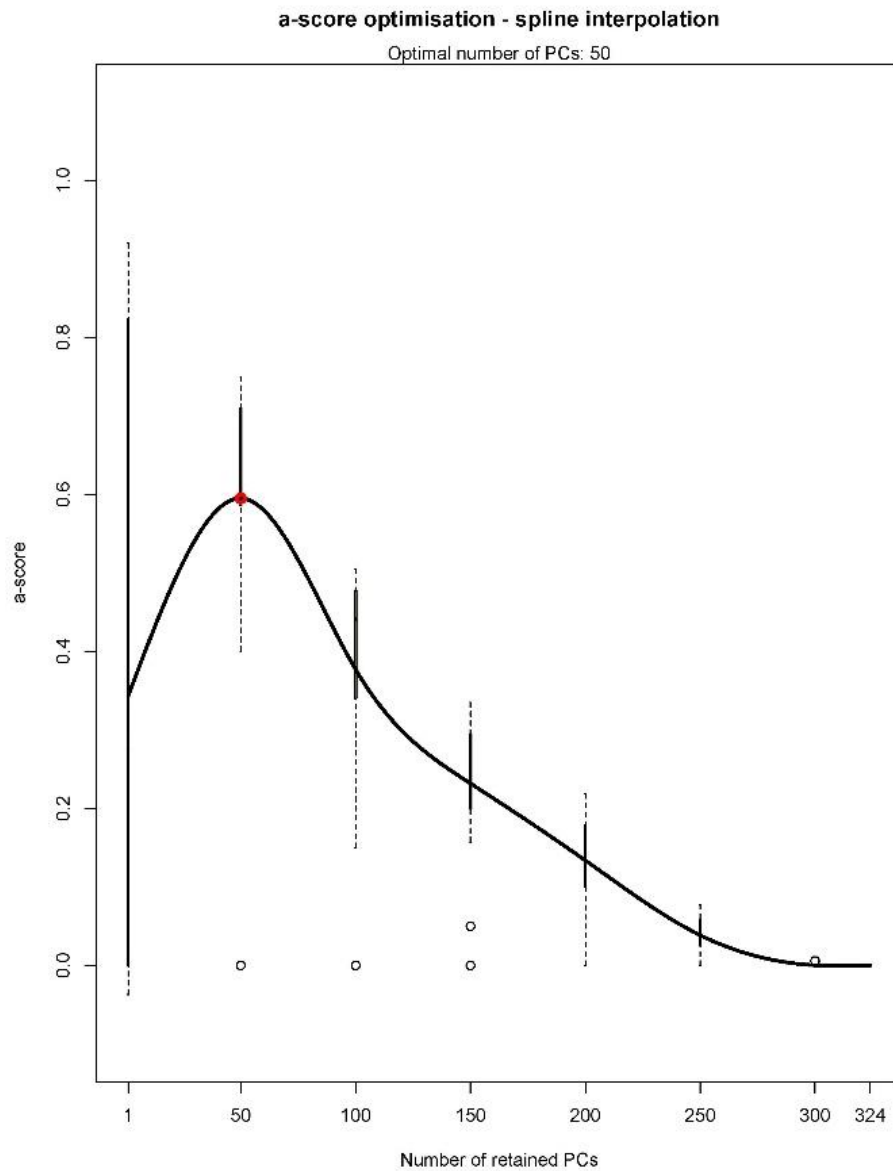
174 **Supplementary Figure 11: Genetic clusters.** The BIC value for each cluster. Note that the
175 BIC begins to dramatically increase after 15.

176

177 *Principal component optimization*

178 We utilized the a-score to optimize the number of principal components (PCs) to include in our

179 DAPC – 50 PCs maximized the a-score and were therefore selected.

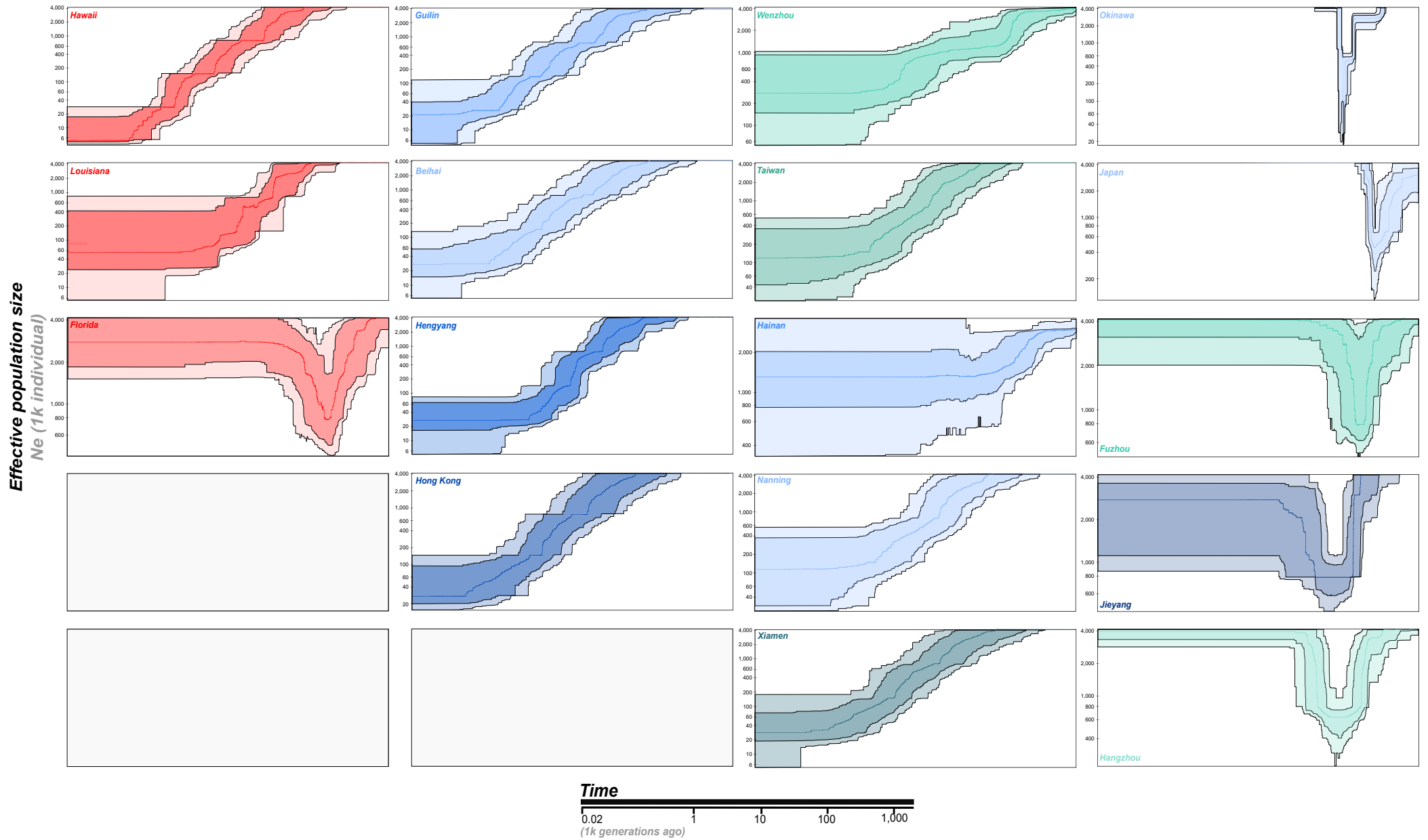


180 **Supplementary Figure 12: Principal component optimization.** The red dot represents the

181 number of retained PCs that maximize the a-score.

Supplementary Table 12: The number of *C. formosanus* workers sampled from each location.

Locality	N
United States	
Hawaii	14
Texas	4
Louisiana	16
Mississippi	1
Florida	46
Japan	
Okinawa	18
Mainland	20
China	
Nanning	20
Beihai	20
Hainan	14
Guilin	20
Hengyang	20
Changsha	3
Xinyu	1
Hong Kong	20
Lufeng	7
Jieyang	15
Xiamen	20
Fuzhou	20
Wenzhou	20
Hangzhou	20
Taiwan	20



Supplementary Figure 13: Demographic histories. Estimation of the variation of the effective population size through time for all invasive and native localities using Stairway Plot 2. Light and dark shaded areas for each plot correspond to the 95% and 75% confidence intervals for the effective population size, respectively. The dark line within the confidence intervals in each plot represents the estimate for the median effective population size.