Strong selection against hybrids maintains a narrow contact zone between morphologically cryptic lineages in a rainforest lizard: Supplementary Information

Sonal Singhal and Craig Moritz
singhal@berkeley.edu
craigm@berkeley.edu

Museum of Vertebrate Zoology
University of California, Berkeley
3101 Valley Life Sciences Building
Berkeley, California 94720-3160

Department of Integrative Biology
University of California, Berkeley
3101 Valley Life Sciences Building
Berkeley, California 94720-3140

1 Additional Information on Simulations

To investigate the behavior of clines and disequilibrium statistics under a number of scenarios, we simulated secondary contact between two isolated lineages using the forward-time program simuPOP (Peng and Kimmal 2005). Here, we report additional rationale for how we designed the simulations and why we chose the parameter space we did (Fig. S5, Table S4). First, we simulated a hybrid zone as a one-dimensional chain of demes, even though species exist in two-dimensional space. This simplification was appropriate because (1) cline theory is applied to linear transects (Endler 1977) and (2) one-dimensional systems generally approximate the behavior of two-dimensional systems (citation?). Second, we simulated sixty populations, as we found (1) it minimized edge effects, (2) unlike if we used fewer demes, it allowed us to look at changes in clines over longer periods of time, and (3) any greater number of demes significantly slowed down simulation speed. Third, dispersal, as used here, does not translate directly to any physical measure as it is given in terms of deme number. Thus, these simulations should apply to a wide range of organisms with varied dispersal capabilities. Fourth, fitness was characterized as a multiplicative selection model, which is most appropriate for a complex trait like fitness that likely has a polygenic model of inheritance. Further, although this model's assumption that all loci contribute evenly to fitness is biologically unrealistic, it is a necessary simplification. We simulated under the full possible range of fitness values, from no selection to nearly complete selection against hybrids. Finally, to model assortative mating, we used a "group-based model", in which individuals mate preferentially with those who are of similar ancestry (Felsenstein 1981). Here, this effectively becomes a multilocus model of assortative mating. Most other hybrid zone models use a one-locus or two-locus assortative mating model (Servedio 2004); although this model has support from biological systems (Kronforst et al. 2006; Shaw and Lesnick 2009), it is likely that both the phenotypes of female preference and male traits generally have a polygenic basis. In this case, an ancestry-based model is appropriate. As with selection, we looked at the outcomes of the model under the full range of assortative mating, from random mating to nearly complete assortative mating.
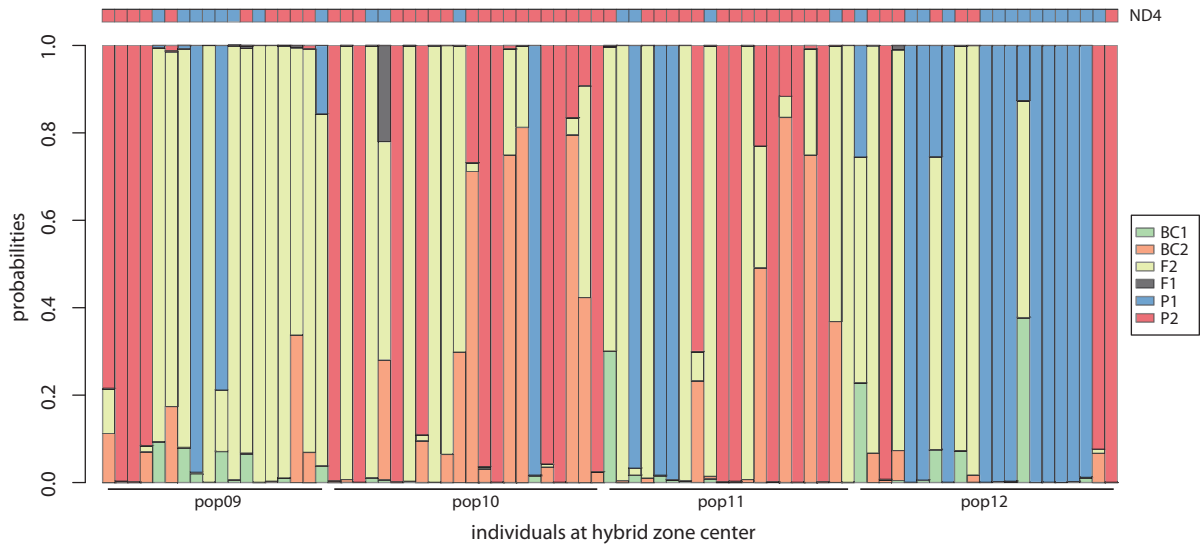
2 Figures & Tables

Figure 1: NewHybrids classification of hybrid class of individuals located in the hybrid zone center.

Name	Latitude	Longitude	N	location on transect
Lake Morris, QLD	-16.824109	145.640146	12	38 km N
Gillies Lookout, QLD	-17.171538	145.68705	17	3.8 km N
Gadgarra Forest, QLD	-17.273448	145.663125	18	2.8 km S
S Johnstone River, QLD	-17.623924	145.594004	12	45 km S
GilliesTransect_1	-17.204725	145.679049	11	0 m
GilliesTransect_2	-17.205197	145.679384	20	52 m
GilliesTransect_3	-17.209406	145.683135	19	515 m
GilliesTransect_4	-17.210666	145.683765	4	654 m
GilliesTransect_5	-17.212456	145.683809	22	852 m
GilliesTransect_6	-17.214125	145.684747	12	1036 m
GilliesTransect_7	-17.214935	145.685569	18	1041 m
GilliesTransect_8	-17.214589	145.686652	22	1086 m
GilliesTransect_9	-17.214737	145.687169	20	1102 m
GilliesTransect_10	-17.215187	145.688181	21	1151 m
GilliesTransect_11	-17.216708	145.687258	13	1320 m
GilliesTransect_12	-17.21745	145.687347	11	1402 m
GilliesTransect_13	-17.21813	145.687718	15	1477 m
GilliesTransect_14	-17.219121	145.688736	18	1586 m
GilliesTransect_15	-17.219673	145.690143	24	1646 m
GilliesTransect_16	-17.220009	145.690423	14	1683 m
GilliesTransect_17	-17.219659	145.695023	28	1701 m
LakeBarrine_1	-17.252015	145.630107	1	
LakeBarrine_2	-17.254058	145.631531	1	
LakeBarrine_3	-17.245581	145.63276	1	
LakeBarrine_4	-17.255035	145.636904	1	
LakeBarrine_5	-17.254427	145.639231	1	
LakeBarrine_6	-17.245229	145.639496	1	
LakeBarrine_7	-17.246107	145.639687	1	
LakeBarrine_8	-17.252836	145.64056	1	
LakeBarrine_9	-17.25114	145.640929	1	
LakeBarrine_10	-17.248131	145.641093	1	
LakeBarrine_11	-17.248276	145.641091	1	
LakeBarrine_12	-17.248403	145.641109	1	
LakeBarrine_13	-17.247808	145.641368	1	
LakeBarrine_14	-17.247954	145.641639	1	
LakeBarrine_15	-17.249258	145.64194	1	
LakeBarrine_16	-17.252039	145.629656	2	
LakeBarrine_17	-17.251924	145.630033	2	
LakeBarrine_18	-17.24535	145.633345	2	
LakeBarrine_19	-17.245101	145.636723	2	
LakeBarrine_20	-17.245139	145.636958	2	
LakeBarrine_21	-17.254006	145.639808	2	
LakeBarrine_22	-17.248529	145.64109	2	
LakeBarrine_23	-17.248131	145.641093	2	
LakeBarrine_24	-17.249009	145.6306	3	
LakeBarrine_25	-17.255033	145.632709	4	
LakeBarrine_26	-17.251095	145.628995	5	
LakeBarrine_27	-17.245125	145.638857	5	
LakeBarrine_28	-17.244263	145.638346	10	

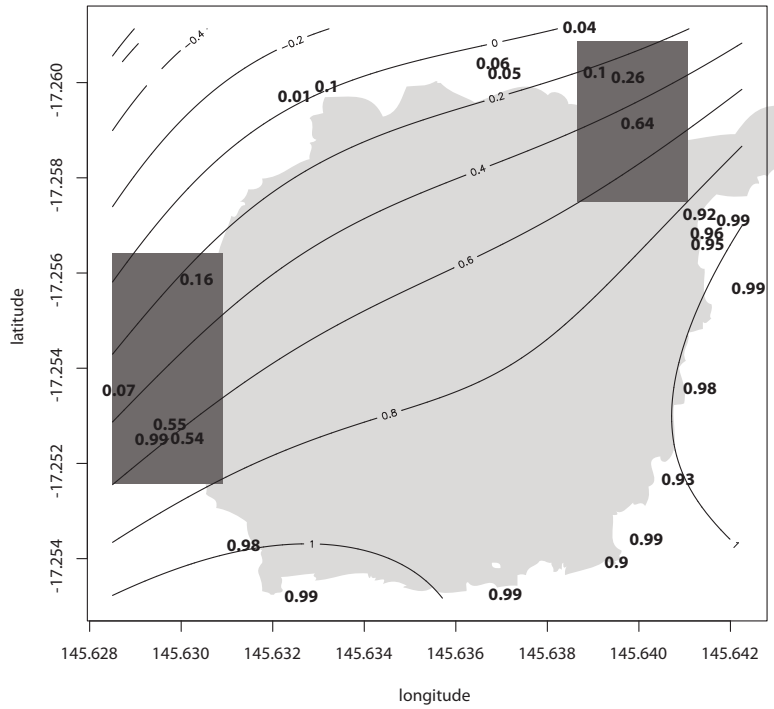Table 1: Sampling points for this study.

Figure 2: A map of Lake Barrine (shown in light gray), with individual hybrid indices shown in bold. Regions where hybrids are found are shown in dark grey. Isoclines represent projected hybrid index values along the lake, as based on a generalized least squares model (implemented in R package 'nmle' (Pinheiro et al. 2011)).
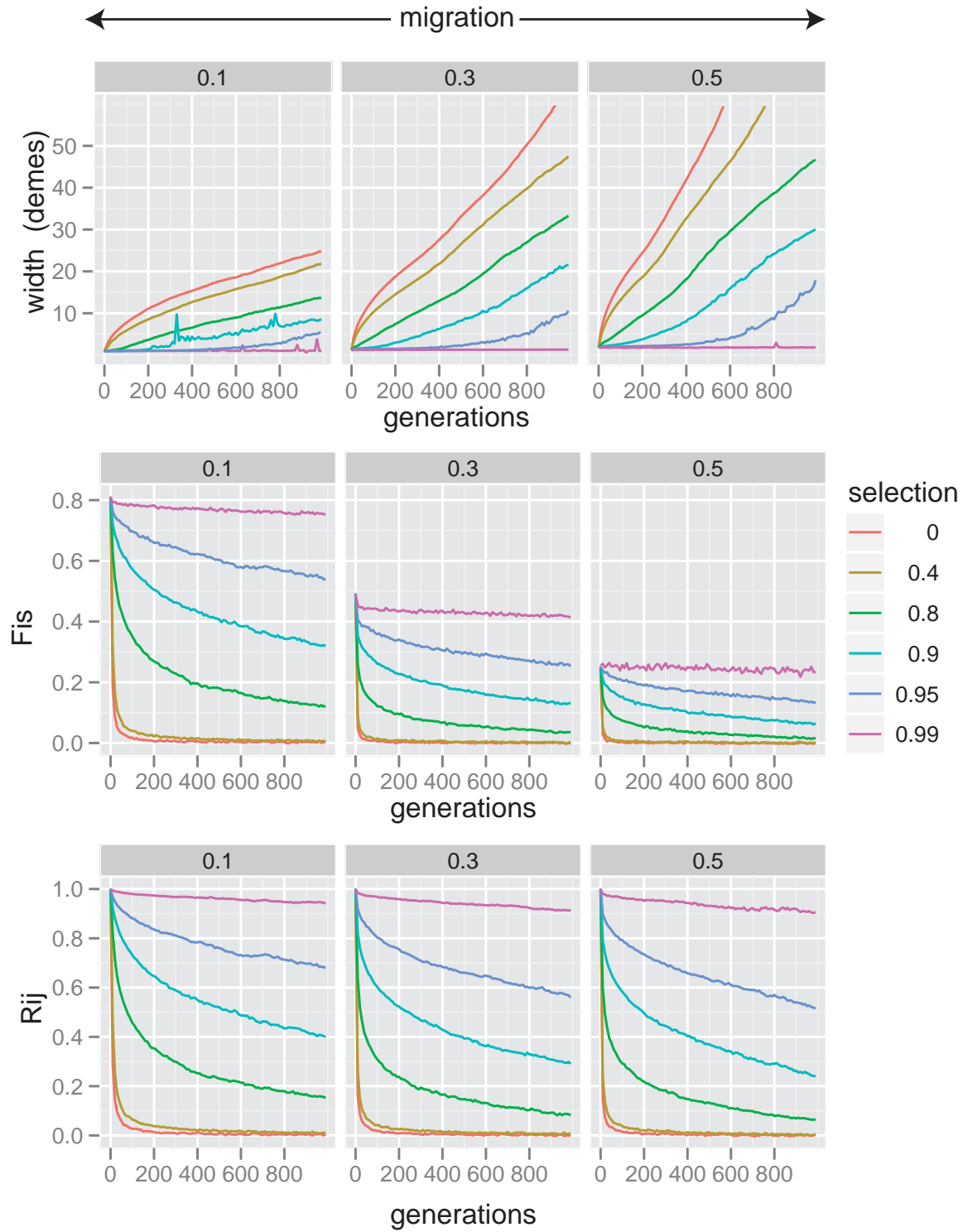
Figure 3: Cline width (in demes) at a neutral locus, F_{IS} at a neutral locus, and R_{ij} at a neutral locus for a range of values for selection against hybrids and migration rates. Shown for no assortative mating and with 10 loci under selection.
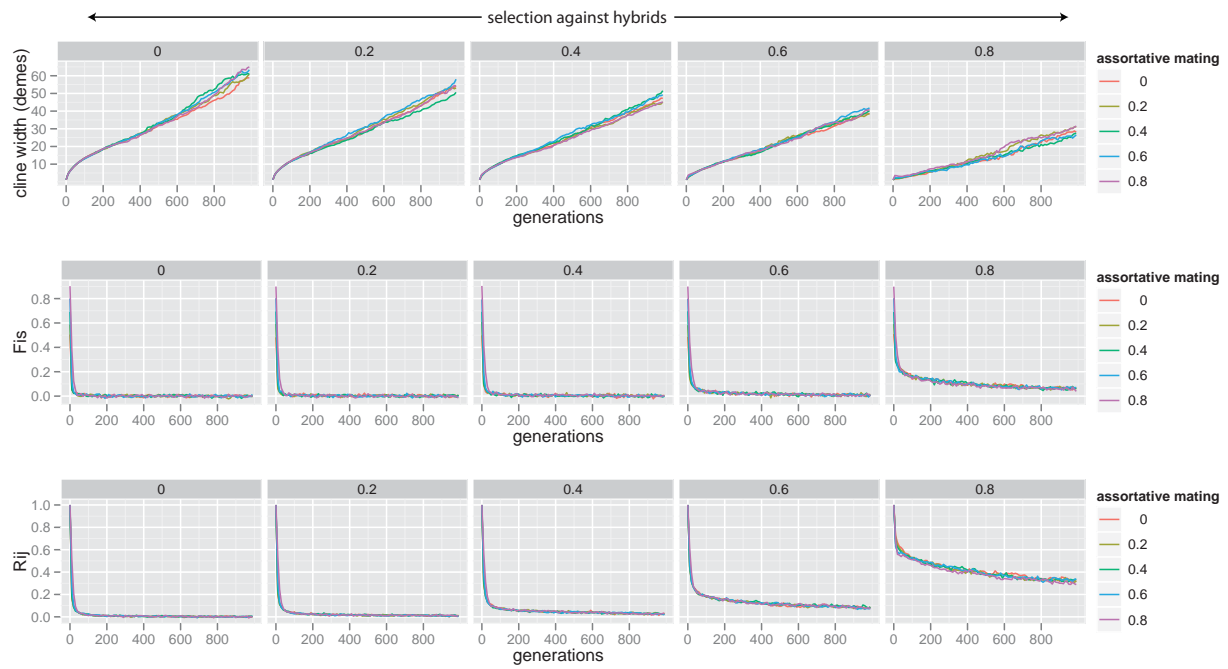
Figure 4: Cline width (in demes) at a neutral locus, F_{IS} at a neutral locus, and R_{ij} at a neutral locus for a range of values for selection against hybrids and strength of assortative mating. Shown for migration rate of 0.3 and with 10 loci under selection.
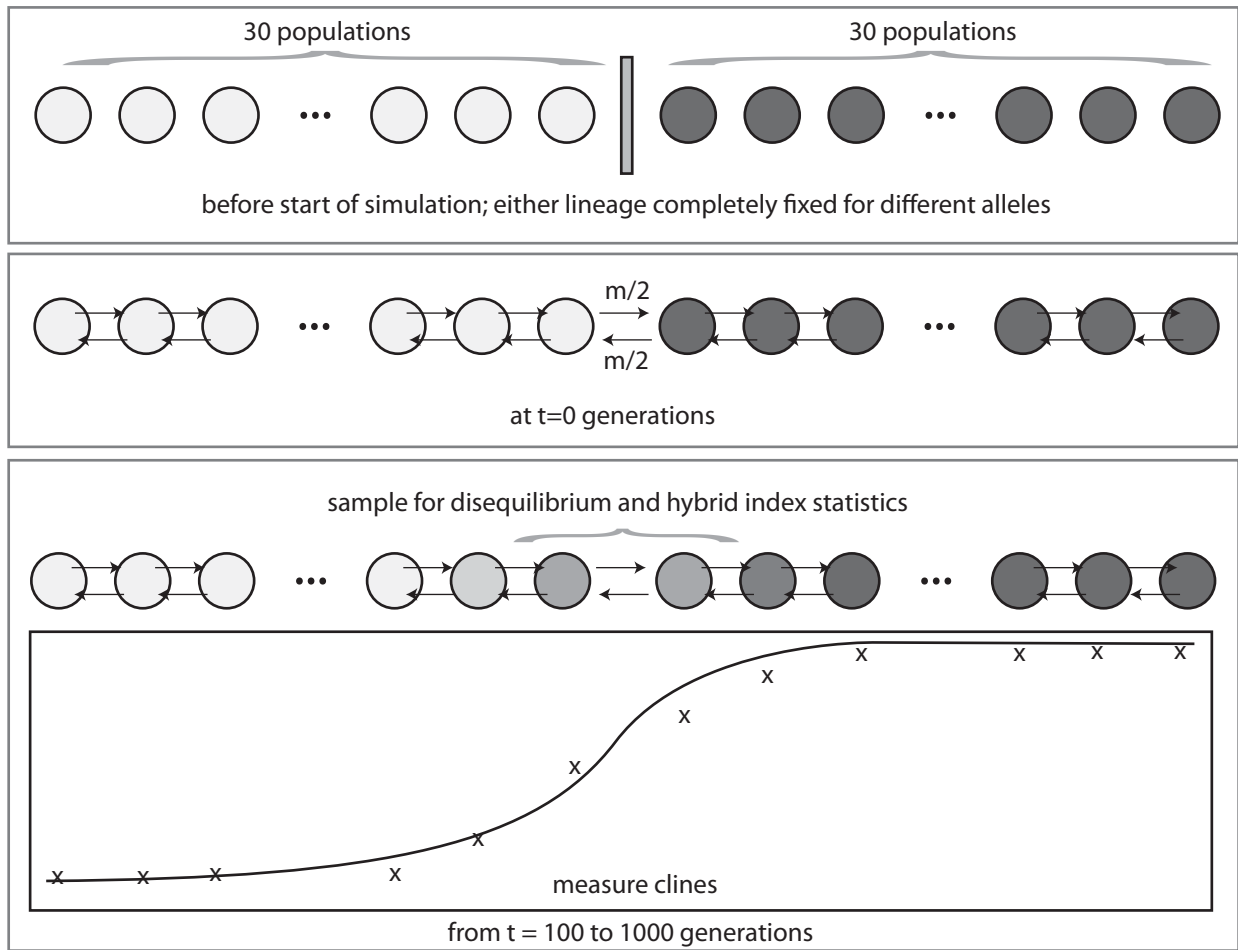
Figure 5: A cartoon schematic of how simulations were conducted.

LocusID	Primer Sequence	PCR Temp.	SNP	Location/Type	R. Enz.	Cutting Pattern	Reference
ND4	ND4 5' CACCTATGACTACCAAAAGCTCATGTAGAAGC 3' LEU 5' CATTACTTTTACTTGGATTTGCACCA 3'	50	G(C|A)GC	CDS, syn.	HhaI	900 (530-370)	Arevalo et al, 1994
beta-globin (intron 2)	Bglo1CR 5'GCG AAC TGC ACT GYG ACA AG 3' Bglo2CR 5'GCT GCC AAG CGG GTG GTG A 3'	63	GCAG(C|T)	intron	ApeKI	670 I (640+30)	Dolman and Phillips 2004
LC5/LC6 triosephosphohote isomerase (intron 5)	LC5 5'TTC TAG CCT ATG AAC CAGTTT GG 3' LC6 5'CCT CAA CTT GTC ATG AAC TTC C 3'	50	15 bp indel	intron	--	--	Bell et al., 2010
LC17/LC18 40S Ribosomal protein S8 (intron 3)	LC17 5'CTC TTG GGC GTA AGA AAG GAG 3' LC18 5'CCG CTC ATC GTA TTT CTT CTG 3'	53	(A|G)C(A|G)C	intron	HhaI	630 I (400+230)	Bell et al., 2010
ABHD5 abhydrolase domain containing protein 5	F 5' ACCCACTTGTTCTTCTCCA 3' R 5' TGAGTAAGCAGCTGCCAAAA 3'	60	TTC(G|A)AA	CDS, syn.	BstBI	230 I (160+70)	this paper
AUTO autophagy-related protein 101	F 5' TGAGCAGGAAAGGCAAATCT 3' R 5' GTGCCAGTGTGTCCTTGATG 3'	62	G(G|A)CGCC	CDS, syn.	BanI	190(170+20)	this paper
NDST2 N-deacetylase/N-sulfotransferase 2	F 5' TCTTGGGGTTGTTTCCAGAC 3' R 5' CACTTGGCATTGTGAGCAGT 3'	60	CC(A|T)TGG	3'UTR, non-coding	NcoI	440 I (230+210)	this paper
LEMD2 LEM domain-containing protein 2	F 5' GTGCATTCAAGCAGACCAGA 3' R 5' GGCTAGCCACTCTCCACCAAG 3'	60	AAGCT(T|G)	3'UTR, non-coding	HindIII	240 I (140+100)	this paper
PCBD1 Pterin-4-alpha-carbinolamine dehydratase	F 5' TCTCTCTTGGCTGTGTGGAA 3' R 5' TAAATCATGTGCCCCCAAAT 3'	60	GAA(T|A)C	3'UTR, non-coding	HinfI	250 I (140+110)	this paper
RTN3 reticulon-3	F 5' AACCTGTTCCAACGCAATTC 3' R 5' TTGAGAAAGGGGAGTTGTGG 3'	60	GAGCT(C|A)	3'UTR, non-coding	Eco53KI	440 I (130+310)	this paper
SAR1 GTP-binding protein SAR1a	F 5' TAAATCACTTTGGCCCACCTC 3' R 5' TATCGCACAAATGCAAGAGC 3'	56	GTCTA(C|T)	3'UTR, non-coding	AccI	420 I (350+70)	this paper

Table 2: Loci used in this study, including their diagnostic SNPs and cutting patterns with listed restriction enzyme.

Parameter	Values
Number of loci under selection	2,5, 10
Assortative mating	0, 0.2, 0.4, 0.6, 0.8
Selection against hybrids	0, 0.2, 0.4, 0.6, 0.8, 0.9, 0.95, 0.99
Migration rates	0.1, 0.3, 0.5
Number of populations	60
Population size	1000
Recombination rates	0.5
Number of neutral loci	10

Table 3: Parameters for simulation for this study.

References

Endler, J., 1977. Geographic variation, speciation and clines. Princeton, Princeton, NJ.

Felsenstein, J., 1981. Skepticism towards Santa Rosalia, or why are there so few kinds of animals. Evolution 35:124–138.

Kronforst, M., L. Young, D. Kapan, C. McNeely, R. O'Neill, and L. Gilbert, 2006. Linkage of butterfly mate preference and wing color preference cue at the genomic location of *wingless*. Proc. Natl. Acad. Sci. 103:6575–6580.

Peng, B. and M. Kimmal, 2005. simuPOP: a forward-time population genetics simulation environment. Bioinformatics 21:3686–3687.

Pinheiro, J., D. Bates, S. DebRoy, D. Sarkar, and R Development Core Team, 2011. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-102.

Servedio, M., 2004. The evolution of premating isolation: local adaptation and natural and sexual selection against hybrids. Evolution 58:913–924.

Shaw, K. and S. Lesnick, 2009. Genomic linkage of male song and female acoustic preference QTL underlying a rapid species radiation. Proc. Natl. Acad. Sci. 106:9737–9742.