Supplementary methods for: *Sea-level driven glacial-age refugia and post-glacial mixing on subtropical coasts, a palaeohabitat and genetic study*

Dolby *et al.*

Table of Contents

Genetic methods

Age of genetic patterns

Comparison of our microsatellite to mtDNA sequence data indicates that the microsatellite data are informative for timescales relevant to the glacial-interglacial processes of interest. North-south mitochondrial clades of *Gillichthys mirabilis* on the Pacific coast (Figure S2A) diverged at 0.63 Mya (95 % CI 0.24–1.08 Mya) [1,2]. The same mtDNA markers were used on *Quietula y-cauda* and reveal similar mtDNA patterns, likely reflecting a similar age of diversification as inferred in *G. mirabilis.* Microsatellite loci often mutate faster on average and reflect a range of mutation rates [3], and the microsatellite loci used here exhibit high degrees of polymorphism (average number of alleles per locus ranged 9.8–18 for individuals sampled across 1000 km). Given the inferred higher mutation rates of microsatellites, our microsatellite data reflect a range of evolutionary processes and events younger than the 0.63 Mya mtDNA divergence age. Thus, a subset of the microsatellite data would correspond to the glacial-interglacial timescale of interest. We therefore use the Discriminant Function Analysis (DFA) to obtain a refuge-associated partition of the data to examine LGM-present processes explicitly and to complement the full microsatellite data analysed in STRUCTURE.

Diversity metrics, mismatch distributions, F_{ST}

There are competing expectations regarding patterns of traditional diversity metrics in refuge-recolonisation scenarios. Refugia are usually centres of high genetic

diversity and recolonised sites are bottlenecked and exhibit lower diversity [4]. However, recolonised sites that are admixed from two genetically distinct source populations (e.g., refuges) can instead lead to high diversity measures in those populations [5]. Consistent with this latter scenario, populations inferred here to be recolonised show similar measures of allelic richness [6] and gene diversity as refugial source (Vizcaíno and N. Conception) populations (Figure S5). In mean allelic richness, there is a very slight trend decreasing toward the north in *F. parvipinnis*, with a similar pattern for mean gene diversity in *Q. y-cauda*, however the northernmost population, Morro Bay, may also have been bottlenecked (see DFA training N. Conception Refugium by proxy). In addition, ranges of these taxa are extensive to the south of the study area (Pta Eugenia), potentially providing an intermittent source of alleles from the south, which is beyond the scope of this study.

Gillichthys mirabilis has sufficiently distinct northern and southern mitochondrial clades (Figure S2A), and adequate populations and individuals sampled such that mismatch distributions may reflect the admixed or non-admixed nature of populations (for example see [7]). Broadly, the mismatch distributions reveal unimodal distributions for refuge populations, and bimodal or multimodal distributions for several intervening (inferred as recolonised) populations (Figure S6). This pattern suggests that, for *G. mirabilis*, refuges are stable through time (single modes) and intervening sites experience contributions from genetically distinct sources (bimodal or multi-modal patterns). The inferred recolonised populations that show unimodal distributions (DEV, USB, MGU) are within the Southern California Bight, north of the offshore islands where eddy mixing may homogenize genetic signatures during the pelagic larvae phase. This is also the location where STRUCTURE results begin showing notable admixture (Figure 4A).

Pairwise F_{ST} measures using microsatellite data for all three species indicate *G. mirabilis* may be more dispersive than *F. parvipinnis* (Table S4). Sample limitations for *Q. y-cauda* render inferences difficult due to low statistical power.

Overall, factors such as sample size and local founder events confound traditional population genetic metrics, which is why in this study we relied primarily on STRUCTURE and a novel DFA approach to evaluate genetic structure. Based on such results, DFA may be a tool for population-level inference when traditional metrics are problematic due to mixing of multiple sources, founder effects, and sample sizes.

Approximate Bayesian Computation (ABC) simulations

ABC Motivation and Parameterization An expanding wave front during northern range expansion can enrich northerly populations in different alleles than the source (southern) refugium, possibly resulting in an appearance of two end-members with admixture of intervening populations, which we observe (Figure 3). To assess this possibility, we hypothesis-tested two simplified phylogeographic scenarios using DIYABC v2.1.0 [8]. Populations for each species were grouped separately into three groups: a southern refugial group, a middle 'admixed' group, and a northern 'refugial' group. This grouping allowed us to simplify demographic assumptions and increase statistical power to explicitly test whether a two-refugium or one-refugium scenario would be more likely to produce our observed genetic data. Scenario 1 had two refugia (north and south, as inferred from our habitat models) with bi-directional recolonization (admixture) of intervening habitats by the two source refugia. Scenario 2 had a single southern-refugium

(Vizcaíno) with stepping stone recolonization of northern sites (Figure S9). Scenario 2 lacks admixture because there is no second genetic source to contribute.

We increased the default 40-bp allele range to 50 bp for *G. mirabilis* and 130 bp for *Q. y-cauda*, and 90 bp for *F. parvipinnis* after excluding two markers with larger ranges. We ran two million simulations per species assuming constant population sizes. The admixture in scenario 1 assumes equal source contributions. We assume a 2-year generation time and conservatively based event ages on our habitat models where colonization of the 'admixed' population occurred ~8 kya and populations merged ~20 kya at the LGM. The assumed merging of populations 20 kya is dictated by *scenario 2* in which isolation of the refuge coincides with sea-level lowstand ~20 kya. Unless otherwise noted, we used default parameters.

ABC Results We calculated posterior values under the two scenarios using 2 million simulated datasets and inference from both a direct or logistic approach. Direct approach used 50–500 closest datasets (sample interval: 50), and the logistic approach used 4,000–20,000 closest datasets (sample interval: 4000). All posterior values sampled favored scenario 1, which was the two-refuge scenario (Table S6). Though the range of predictive posterior errors we present for the three species appears high (0.14–0.23), these values are within the range documented elsewhere for the Monte Carlo estimation of predictive posterior error in DIYABC [9]. Moreover, we expect somewhat high error values given the simplifying assumptions we made about the phylogeographic scenarios we tested, and a lack of knowledge about effective population size through time, relative admixture rates, the assumed absence of migration, and unknown generation times for these species.

ABC–testing Isolation By distance We tested another alternative hypothesis (data not shown in manuscript) for *Quietula y-cauda* to try to test support for our inferred two-refugium + admixture scenario against isolation by distance (IBD). We couldn't explicitly model IBD in DIYABC, so we assumed that for IBD (and not admixture) to have produced the genetic patterns we observe would require that the populations persisted over this timescale *in situ*. So, we allowed the 3 groups to evolve separately without admixture since 20 kya (10,000 generations), which essentially assumes all populations sheltered in place during the LGM. The two-refugium model was greatly favored (direct approach: 0.96; logistic approach: 0.999) over this independent 'evolved *in situ*' scenario. This result suggests that an IBD-only scenario (without extirpation and end-member refugia) is unlikely to produce this pattern.

<u>DFA training N. Conception Refugium by proxy</u>

Modern populations immediately north and south of the North Conception Refugium (NCR) were used as a training proxy in the DFA discriminant allele analysis. Morro Bay, immediately north of the NCR, was used for each species, as well as the first population immediately south of the NCR for each species (Devereaux, Goleta, Carpinteria populations for *G. mirabilis*, *Q. y-cauda*, and *F. parvipinnis*, respectively). Since they are immediately adjacent to the NCR (Morro Bay is ~30 km and the farthest site included to the south is ~100 km), and given the early post glacial formation of

habitat in the Santa Barbara Channel (Figure S7), we assume they were founded from the NCR prior to any southern admixture. Using these proxy populations provided a similar number of individuals relative to the southern refuge for the discriminant analysis (N:S training sample sizes were 19:14, 12:8, 26:18 for *G. mirabilis, Q. y-cauda, F. parvipinnis,* respectively)

This proxy was necessary, because although the NCR identified in our habitat models is predicted to support tidal estuarine habitat between 140 mbpsl until about 5 mbpsl, at present it does not have tidal habitat or support populations of these three fish species. Conversion of this habitat likely resulted from natural infilling from wave action and easily eroded Transverse Ranges [10,11], and anthropogenic processes of leveeing and damming that promote conversion to a closed lagoon state [12]. Historical maps indicate that at 1895 the Arroyo Grande/Pismo Creek system in the North Conception refugium was larger and more open to the ocean than today [13]. Flood control measures now separate Arroyo Grande and Pismo Creek, precluding tidal behaviour. We therefore used the two most geographically proximate populations of each species in the genetic DFA as the N. Conception training group.

DFA assumptions

Discriminant Function Analysis (DFA) assumes that independent variables are normally distributed. While the nature (0s, 1s, 2s) of allelic count data is likely to violate this normality assumption, we use DFA to identify alleles discriminating alleles between the two refuge sites. These alleles are then used in a separate exercise to analyse mixing along the coastline (Figure 4B). Thus, we are not using DFA as a test statistic to assess the adequacy of different classification schemes, which makes the violation of normality less consequential.

Discriminant Function Analysis also assumes equal variance among independent variables (alleles). We found that per-allele variance of total observations ranged in *F. parvipinnis* from 0.01 to 0.25 (mean = 0.082, median = 0.06), for example. Another issue of concern in this analysis is multicollinearity, in which variables are correlated. In this study, the multicollinearity of our variables is dependent on, and limited by, basic biological processes, such a random versus non-random mating and low recombination rates relative to the microsatellite loci studied, such that linkage disequilibrium may colinearize otherwise independent alleles. Similarly, the assumption of random sampling is satisfied to the extent possible given that individuals are components of interbreeding populations, and in that regard are not truly independent of other individuals. We took care to sample estuaries thoroughly, and individuals from different locations within estuaries were mixed, which may reduce batch effects from any individual seine haul.

Sampling and marker development

Individuals were collected via seining and preserved in 100 % ethanol in the field (permit numbers DGOPA 14253.101005.6950 CASCP No. 2679). DNA extractions were performed using Qiagen DNeasy Blood and Tissue Kit according to manufacturer's directions for muscle tissue. Microsatellite loci were developed using sequencing on the Roche 454 platform of one individual per species and processed with MSATCOMMANDER [14] to generate primers; tetra-, tri-, and di-nucleotide repeats were favoured, respectively. Markers were screened using a subsample of individuals across populations and repeat

number of some homozygotes were verified by standard PCR and Sanger sequencing methods using 1 µL of each microsatellite primers (10 mM) in separate reactions. Microsatellite genotyping plates were run on six to twelve individuals per estuary (where available) according to Ellingson [1] and genotyped in GENEIOUS v5.6 (http://www.geneious.com, [15]). Small sample sizes were recovered from some estuaries (Table S1). After discarding loci of substandard quality and individuals with significant missing data (not genotyped for > 2 loci), the number of loci, total number of alleles, and sample sizes are as follows: *G. mirabilis* (16, 80, 100), *Q. y-cauda* (17, 148, 44), and *F. parvipinnis* (20, 199, 79).

Worth noting, the DFA and STRUCTURE analyses are fundamentally different approaches to analysing genetic data. We analysed the STRUCTURE output from one run of *F. parvipinnis* ($K = 2$) and identified alleles that had an estimated per cluster allele frequency greater than 0.7, which yielded 15 alleles. Comparing the identity of 15 alleles to the identities of discriminant alleles significant through DFA (N = 39) yielded a match of 47 %. In summary, these approaches draw on partially independent components of the overall genotypic dataset, analyses them through different statistical/probabilistic methods, and produce very similar results.

PCR protocols and tree reconstruction

Microsatellite PCR reactions used one hybrid primer combination: 2 µl Reverse primer (100 µM), 4 µl Forward M13 hybrid primer (2.5 µM), 4 µl M13 dye labelled primer (2.5 µM), 90 µl H_2O for a total of 100 µl. Thermocycler protocol is: 1) 95 °C 15 minutes, 2) 94 °C 30 sec, 3) 55 °C 1 min 30 sec, 4) 72 °C 1 min, 5) repeat steps 2–4 24x, 6) 94 °C 30 sec, 7) 50 °C 1 min 30 sec, 8) 72 °C 1 min, 9) repeat steps 68 24x, 10) 60 °C 30 min, 11) end. PCR products are diluted to 5 % (2 µl PCR product to 38 µl H_2O) for genotyping reaction with 10 µl of a 1:50 LIZ: Hi-Di mix (95 °C for 5 minutes).

Mitochondrial Control Region (mtCR) and cytochrome B (Cyt B) were amplified and sequenced for *G. mirabilis* and *Q. y-cauda* using A and M, AJG15 and H5 primer sets [16,17]. Primers K and N from [17] were used to amplify and sequence mtCR for *Fundulus parvipinnis*. Amplification and sequencing protocols are available in detail [1]. Trees were constructed in MRBAYES v3.1.2 [18] on the CIPRES Science Gateway [19]. Sequences was partitioned by gene and a rate partitioning scheme was applied to mtCR region in *Q. y-cauda* following [2] and eliminating the fastest rate column of four due to concern over homoplasy (Figure S3A). Three runs of 12 million generations were completed with 4 chains per run under default model settings and a burn-in fraction of 25 % trees discarded. While unresolved in our Bayesian analysis, *F. parvipinnis* structure was recovered in a Neighbour-Joining tree reconstruction method previously [20] and showed north-south geographically structured clades.

The following programs were used for file conversions: CONVERT, GENODIVE, and PGDSPIDER [21-23]. Observed mtDNA mismatch distributions and pairwise F_{ST} were calculated in Arlequin v3 [24]; gene diversity and allelic richness were calculated in FSTAT v1.2 [25]. STRUCTURE v2.3 [26] was used to run K= 2–5 (3 replicates each) that were analysed in STRUCTURE HARVESTER [27]. The following graphics R packages were used: LATTICE, ADE4, PLYR, RESHAPE2, GGPLOT2 [28-31]. All other statics were performed in JMP® v11 (SAS Institute Inc., Cary, NC, 1989-2007).

Habitat modelling methods

Detailed modelling methodology

Parameterization To predict estuarine habitat area, we defined two criteria necessary to form estuarine habitat. First, we used GOOGLE EARTH to calculate modern bathymetric slopes amongst the 18 estuaries in this study (Table S3). For five relatively large and heterogeneous estuaries, we captured a range of within-estuary slopes at the centre, sides, stream entry, mouth, as applicable. We calculated a single slope from each of 13 relatively small estuaries. The 'run' used for slope calculations varied with estuary size from 200–5,000 m. Slopes ranged between 0 % and 1.3 % (mean = 0.45, median = 0.39). Our second criterion was a sea-level requirement. Using a composite sea-level curve [32], sea-level lowstand was determined as 130–140 mbpsl. The midpoint depth value was used to date each bin (e.g., 135 mbpsl).

Implementation Using the raster calculator tool in ARCMAP v10 (ESRI, Redlands, CA), we queried an SRTM30_PLUS [33] Digital Elevation Model (DEM) (WGS_1984_UTM_Zone_11N) for areas matching the slope analysis range (0.0–1.3 %) and 10-metre depth range (e.g., 130 – 140 mbpsl). We iterated this process for 0–140 mbpsl to yield a sequence of depth-specific layers using the following equation (Eq 1):

Eq 1 ("Elevation" < x) & ("Elevation" ≥ y) & ("Slope" ≤ 1.3)
Example ("Elevation" < -130) & ("Elevation" ≥ -140) & ("Slope" ≤ 1.3)

where x is the upper and y is the lower limit of each depth bin, respectively. For the present (0 kya) bin we used 0 ± 5 mbpsl. We converted areas matching our query (value = 1) to a sequence of feature layers in which simplified polygons bounded areas that met habitat requirements. To obtain per-depth area estimates for individual coastal regions we also created a feature layer for each coastal location. With the *"Select Features by Location"* tool, we selected the habitat area polygons within each coastal region using the *'Target layer(s) features are within (Clementini) the source layer'* setting. On these selected features we used the "Statistics" feature to provide the following statistical attributes: number of polygons, minimum polygon area, maximum polygon area, total polygon area, mean polygon area, and standard deviation of polygon area. We added an additional attribute, which normalised the summed polygon area by the coastal feature area to account for different coastal area sizes (analogous to habitat density within a given coastal area). These statistical attributes were calculated per depth bin within each coastal region: 14 depth bins, 9 coastal regions, 7 statistical attributes per bin-region produced 882 observations. Of note, the Mercator projection used here could bias polygon areas by a maximum of 7 % of width over the latitude range studied (larger in the northern regions and smaller in the south) relative to an equal area projection. As it is however, the northern polygon areas are already smaller than southern polygons (i.e. Vizcaíno), and would be unlikely to alter interpretations herein.

Statistical assessment To better determine whether the three fish species studied here would likely inhabit the lowstand-associated polygon habitat, we used the modern (0 kya, 0 ± 5 mbpsl) depth bin and species occurrences from this study to determine which

277 polygon statistical attribute(s) predict species occurrences. We performed Discriminant
278 Function Analysis (DFA) using JMP on the seven statistical attributes of 8 coastal
279 locations grouped by habitat presence (N = 6) or absence (N = 2). Vizcaíno was excluded
280 from the DFA analysis after a Robust Fit Outliers analysis (using Huber and Quartile
281 methods with the default K = 4) revealed anomalous coastal area size, which biased the
282 statistical attributes. A stepwise variable selection process (SSP) in the DFA produced
283 two statistically significant predictive variables: Maximum polygon area (maximum size
284 of a single polygon) and total habitat area. We then entered these variables into a
285 Generalized Linear Model (GLM) with binomial distribution (variable states were 'yes'
286 or 'no') to determine which coastal region(s) were likely to have supported refuge
287 populations within the 130–140 m (~20 kya) depth bin. If the GLM was significant, it
288 was re-run using Firth's Biased Adjustment estimates and False Discovery Rate. We
289 performed this iteratively for different refuge scenarios. Unlike typical GLM analyses,
290 this was not used to exclude variables from the refuge scenario model, but rather test
291 whether the refuge scenario was statistically significant using the two variables identified
292 *a priori* to be predictive. Key refuge scenarios are listed in Table S6 with significance
293 scores. Only one refuge scenario was statistically significant (Vizcaíno + North
294 Conception). Vizcaíno was run as a refuge model individually with each additional
295 population not listed in Table S6, none of which were significant in the GLM.
296
297 Climatic, oceanographic factors
298 For the tidal estuarine habitat of focus here, we modelled the major physical
299 geomorphological parameters required for estuary formation. In traditional Ecological
300 Niche Modelling geomorphology is considered constant and temperature and
301 precipitation indices are usually the foremost predictors of palaeohabitat distributions for
302 both terrestrial and intertidal species [34,35]. However, we argue that temperature and
303 precipitation are less important for estuarine habitat than the fundamental geomorphic
304 processes that physically form the estuaries in the first place. The application of
305 geomorphic primacy in this study is further supported by the relatively small change in
306 temperature from the LGM to present, because tidal systems generally have a range of
307 salinities within the system due to marine and freshwater (river) inputs, and because
308 spring and summer estuary temperatures are often controlled by cloud cover which is in
309 turn controlled by upwelling. Upwelling driven factors are partially independent of other
310 glacial and typical seasonal temperature controls. The physical shape, size, and ecology
311 of tidal estuaries can also greatly affect temperature, but are rarely well studied in modern
312 systems. Such detailed reconstructions would be very difficult for palaeoestuaries
313 because palaeorecords are difficult to recovery (via coring or seismic imaging) and
314 estuaries migrate over time. Thus, local non-geomorphic variables are difficult to assess,
315 and regional temperature patterns are probably secondary and difficult to recover on a
316 biologically meaningful temporal or spatial scale.
317 Finally, there are additional oceanographic features that we did not take into
318 account [36]. Specifically, Point Conception marks the northern extent of eddy formation
319 in the southern California Bight (Pt. Conception to Dana Point) [37,38]. The resulting
320 increased retention and mixing of water in this region may have an impact on larval
321 dispersal through a homogenizing effect in southern California specifically [39]. This
322 homogenization may help explain why the north-south cline observed in Figure 4A

begins near the southern end of eddy mixing, and the northern (Morro Bay-Mugu) populations of all three taxa appear mixed on this scale.

Coastal process and reworking of sediments

As is captured in our analysis there is loss of estuary habitat potential through the Holocene. This is a product of coastal process and sediment transport that reshapes the coast and fills estuaries with sediment in the absence of rising sea-levels that form accommodation space and create estuaries, as was the case from 20–10kya. Coastlines retreated in the Holocene due to wave action in some regions [10,40], while riverine embayments filled with sediment from long-shore transport and down-stream sediment supply (Jacobs et al. 2011). Even within the Holocene these processes vary across time, with coastal orientation, and with regional lithology [10,41]. Lithology is important because easily eroded formations will supply more sediment to be reworked. We acknowledge that these factors would have influenced the coastal and underwater topography over the timescale of this study. However, these effects are difficult to parameterize, and we therefore used modern topo-bathymetric data to approximate this aspect of our models.

More formal coastal process/coastal evolution models could inform about details of the nature and quantity of estuarine habitat through time. More detailed process models could include changing wave attack with sea-level, and climatic influence on coastal and stream-flood transported sediment through time. However, there would be considerable complexity in applying such a detailed formal coastal process model that covered estuary formation, and there are also data limitations that preclude simple parameterisation of such a model if it were to be comprehensive over the last 20 kyrs. It is difficult to assess the accuracy of these short-term processes. Such model development and refinements are desirable, but are of a second order and well beyond the scope of the first order work presented here.

Uplift

Significant coastal uplift could, in theory, affect the depth-time correlations inferred from the sea-level curve. However, uplift rates along the coast are typically less than a millimetre per year and unlikely to influence the results of this work when extrapolated over the LGM to present (20 kyrs). As a sensitivity test, we used a 0.7 mm/yr uplift rate extrapolated over 20 kyr, which still produced qualitatively and quantitatively similar results, including the existence of the Conception and Vizcaíno Refugia. Estuaries along this coast are typically on the downthrown block in locally tectonically active areas and are therefore experiencing minimal or no uplift. For example, Pts. Buchon, Loma, and Banda are on uplifting blocks with rates of 0.24–0.09, 0.14–0.16, 0.22–0.25 mm/yr, providing upper limits on uplift rates for the adjacent estuaries of Morro Bay, San Diego Bay, and Banda, respectively [42,43]. At these rates the effects of uplift on our habitat modelling are negligible.

Exceptions to low uplift rates (i.e. 2 mm/yr) are observed locally in the Santa Barbara Channel and could affect our results by biasing the habitat origination ages in this region towards younger estimates [44,45]. Given higher uplift rates in this region we cannot exclude the possibility of habitat in this region at lowstand 20 kya. Such habitat

could then be viewed as an extension of the adjacent North Conception Refugium, and would not greatly alter our biological interpretations.

Supplementary references

1. Ellingson, R. 2012 Phylogenetics and phylogeography of North Pacific bay gobies: adaptive convergence, relictual endemism, and climate-driven population structure. 1–107.

2. Ellingson, R. A., Swift, C. C., Findley, L. T. & Jacobs, D. K. 2014 Convergent evolution of ecomorphological adaptations in geographically isolated Bay gobies (Teleostei: Gobionellidae) of the temperate North Pacific. *Mol. Phylogenet. and Evol.* **70**, 464–477. (doi:10.1016/j.ympev.2013.10.009)

3. Wan, Q.-H., Wu, H., Fujihara, T. & Fang, S.-G. 2004 Which genetic marker for which conservation genetics issue? *Electrophoresis* **25**, 2165–2176. (doi:10.1002/elps.200305922)

4. Hewitt, G. 2000 The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907–913. (doi:10.1038/35016000)

5. Petit, R. J. et al. 2003 Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* **300**, 1563–1565. (doi:10.1126/science.1083264)

6. Nei, M. 1973 Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences* **70**, 3321–3323.

7. Dawson, M. N., Louie, K. D., Barlow, M., Jacobs, D. K. & Swift, C. C. 2002 Comparative phylogeography of sympatric sister species, *Clevelandia ios* and *Eucyclogobius newberryi* (Teleostei, Gobiidae), across the California Transition Zone. *Mol Ecol* **11**, 1065–1075.

8. Cornuet, J.-M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., Marin, J.-M. & Estoup, A. 2014 DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics* **30**, 1187–1189. (doi:10.1093/bioinformatics/btt763)

9. Robert, C. P., Cornuet, J.-M., Marin, J.-M. & Pillai, N. S. 2011 Lack of confidence in approximate Bayesian computation model choice. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 15112–15117. (doi:10.1073/pnas.1102900108)

10. Masters, P. M. 2006 Holocene sand beaches of southern California: ENSO forcing and coastal processes on millennial scales. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **232**, 73–95. (doi:10.1016/j.palaeo.2005.08.010)

11. Upson, J. E. 1949 Late Pleistocene and Recent changes of sea level along the coast of Santa Barbara County, California. *Am. J. Sci.* **247**, 94–115. (doi:10.2475/ajs.247.2.94)

12. Jacobs, D., Stein, E. D. & Longcore, T. 2011 Classification of California estuaries based on natural closure patterns: Templates for restoration and management. *Southern California Coastal Water Research Project* **619**, 1–50. (doi:10.2307/3768203)

13. Gannet, H., Goode, R. U. & Fletcher, L. C. 1895 California. Arroyo Grande quadrangle (15'), 1987 (1925). *U.S.C. G. Survey*. 35.00000,–120.50000; 35.25000,–120.50000; 35.25000,–120.75000; 35.00000,–120.75000; 35.00000,–120.50000.

14. Faircloth, B. C. 2008 msatcommander: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Resources* **8**, 92–94. (doi:10.1111/j.1471-8286.2007.01884.x)

15. Kearse, M. et al. 2012 Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649. (doi:10.1093/bioinformatics/bts199)

16. Akihito et al. 2000 Evolutionary aspects of gobioid fishes based upon a phylogenetic analysis of mitochondrial cytochrome B genes. *Gene* **259**, 5–15.

17. Lee, W. J., Conroy, J., Howell, W. H. & Kocher, T. D. 1995 Structure and evolution of Teleost mitochondrial control regions. *J Mol Evol* **41**, 54–66.

18. Ronquist, F. & Huelsenbeck, J. P. 2003 MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574. (doi:10.1093/bioinformatics/btg180)

19. Miller, M. A., Pfeiffer, W. & Schwartz, T. 2010 Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *2010 Gateway Computing Environments Workshop (GCE)*, 1–8. (doi:10.1109/GCE.2010.5676129)

20. Bernardi, G. & Talley, D. 2000 Genetic evidence for limited dispersal in the coastal California killifish, *Fundulus parvipinnis*. *Journal of Experimental Marine Biology and Ecology* **255**, 187–199.

21. Glaubitz, J. C. 2004 convert: A user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. *Mol Ecol Notes* **4**, 309–310. (doi:10.1111/j.1471-8286.2004.00597.x)

22. Meirmans, P. G. & Van Tienderen, P. H. 2004 GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Mol Ecol Notes* **4**, 792–794. (doi:10.1111/j.1471-8286.2004.00770.x)

23. Lischer, H. E. L. & Excoffier, L. 2012 PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* **28**, 298–299. (doi:10.1093/bioinformatics/btr642)

24. Excoffier, L., Laval, G. & Schneider, S. 2005 Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* **1**, 47–50.

25. Goudet, J. 1995 FSTAT (version 1.2): a computer program to calculate F-statistics. *The Journal of Heredity* **86**, 485–486.

26. Pritchard, J. K., Stephens, M. & Donnelly, P. 2000 Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.

27. Earl, D. A. & vonHoldt, B. M. 2011 STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genet Resour* **4**, 359–361. (doi:10.1007/s12686-011-9548-7)

28. Wickham, H. 2007 Reshaping data with the reshape package. *Journal of Statistical Software* **21**, 1–20.

29. Wickham, H. 2011 The split-apply-combine strategy for data analysis. *Journal of Statistical Software* **40**, 1–29.

30. Sarkar, D. 2008 *Lattice: multivariate data visualization with R*. Springer Science & Business Media.

31. Chessel, D., Dufour, A. B. & Thioulouse, J. 2004 The ade4 package-I-One-table methods. *R news* **4**, 5–10.

32. Chaytor, J. D., Goldfinger, C., Meiner, M. A., Huftile, G. J., Romsos, C. G. & Legg, M. R. 2008 Measuring vertical tectonic motion at the intersection of the Santa Cruz-Catalina Ridge and Northern Channel Islands platform, California Continental Borderland, using submerged paleoshorelines. *Geol Soc America Bull* **120**, 1053–1071. (doi:10.1130/B26316.1)

33. Becker, J. J. et al. 2009 Global bathymetry and elevation data at 30 arc seconds resolution: SRTM30_plus. *Marine Geodesy* **32**, 355–371. (doi:10.1080/01490410903297766)

34. Waltari, E. & Hickerson, M. J. 2013 Late Pleistocene species distribution modelling of North Atlantic intertidal invertebrates. *J. Biogeogr.* **40**, 249–260. (doi:10.1111/j.1365-2699.2012.02782.x)

35. Syphard, A. D. & Franklin, J. 2009 Differences in spatial predictions among species distribution modeling methods vary with species traits and environmental predictors. *Ecography* **32**, 907–918. (doi:10.1111/j.1600-0587.2009.05883.x)

36. Wares, J. P., Gaines, S. D. & Cunningham, C. W. 2001 A comparative study of asymmetric migration events across a marine biogeographic boundary. *Evolution* **55**, 295–306.

37. Bernstein, R. L., Breaker, L. & Whritner, R. 1977 California Current Eddy Formation: Ship, Air, and Satellite Results. *Science* **195**, 353–359. (doi:10.1126/science.195.4276.353)

38. Seapy, R. R. & Littler, M. M. 1980 Biogeography of rocky intertidal macroinvertebrates of the Southern California Islands. In *The California islands* (ed D. M. Powers), pp. 307–323.

39. Bucklin, A. 1991 Population genetic responses of the planktonic copepod *Metridia pacifica* to a coastal eddy in the California Current. *J. Geophys. Res.* **96**, 14977–14808.

40. Masters, P. M. 2003 Archaeological proxies for sediment flux to Holocene littoral cells of southern California. *OCEANS 2003 Proceedings* (doi:10.2307/1544815?ref=no-x-route:a90c68308cf77ba0c24096bfc1852714)

41. Hogarth, L. J., Babcock, J., Driscoll, N. W., Dantec, N. L., Haas, J. K., Inman, D. L. & Masters, P. M. 2007 Long-term tectonic control on Holocene shelf sedimentation offshore La Jolla, California. *Geol* **35**, 275–5. (doi:10.1130/G23234A.1)

42. Muhs, D. R., Rockwell, T. K. & Kennedy, G. L. 1992 Late Quaternary uplift rates of marine terraces on the Pacific coast of North America, southern Oregon to Baja California Sur. *Quaternary International* **15-16**, 121–133. (doi:10.1016/1040-6182(92)90041-Y)

43. Lettis, W. R. & Hanson, K. L. 1992 Quaternary tectonic influences on coastal morphology, south-central California. *Quaternary International* **15/16**, 135–148.

44. Gurrola, L. D., Keller, E. A. & Chen, J. H. 2014 Tectonic geomorphology of marine terraces: Santa Barbara fold belt, California. *Geol Soc America Bull* **126**, 219–233. (doi:10.1130/B30211.1)

45. Niemi, N. A., Oskin, M. & Rockwell, T. K. 2008 Southern California Earthquake Center Geologic Vertical Motion Database. *Geochem.-Geophys.-Geosyst.* **9**, 1–14. (doi:10.1029/2008GC002017)

Supplementary tables and figures for: *Sea-level driven glacial-age refugia and post-glacial mixing on subtropical coasts, a palaeohabitat and genetic study*

Dolby *et al*.

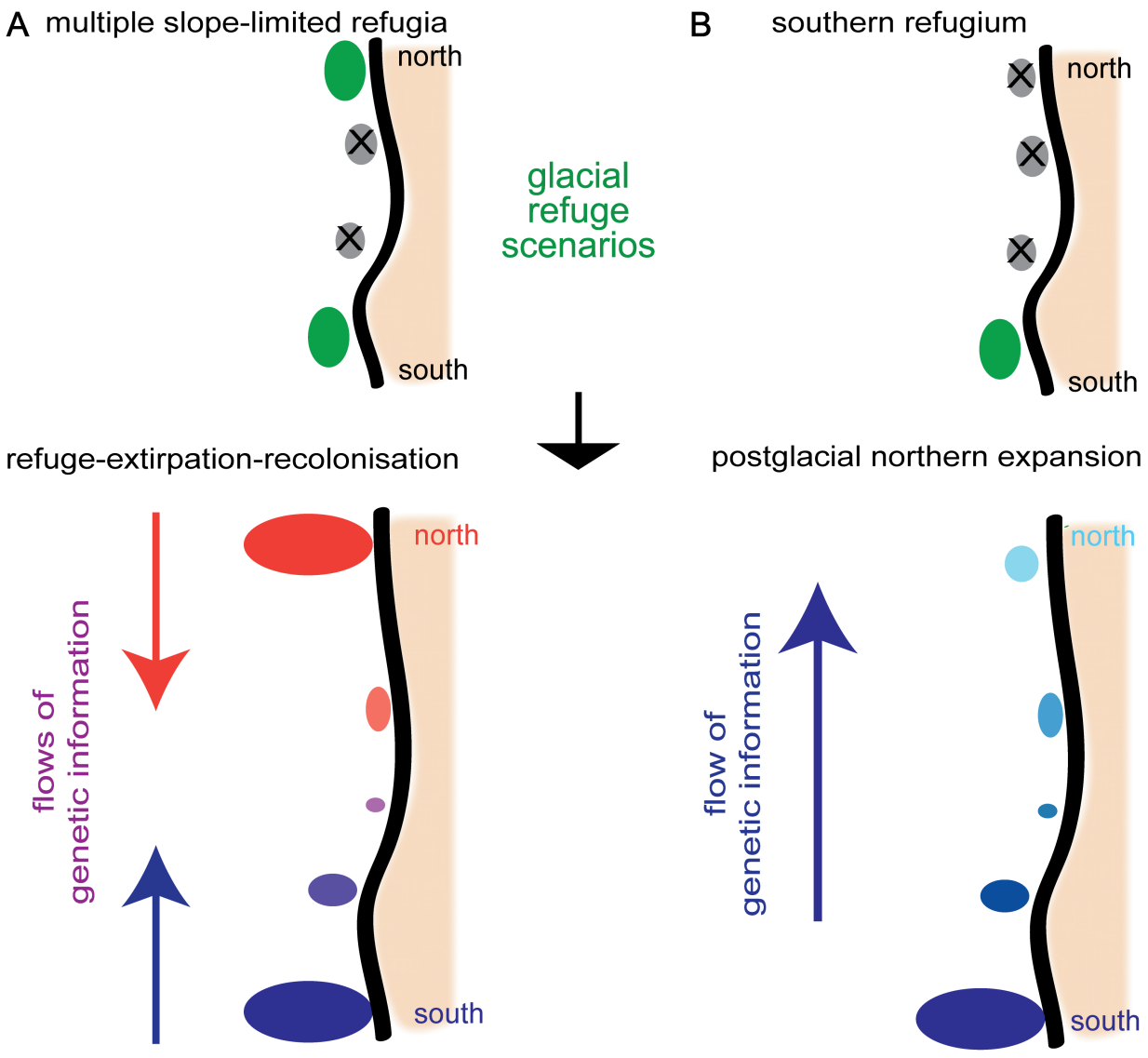

Figure S1. Conceptual schematic. Presented are two refuge-recolonisation scenarios. Colours represent genetic relatedness, where more similar colours are more genetically similar. A) Illustration of our hypothesis where several estuarine populations reduce to two (upper panel), which diverge (different colours, lower panel), and admix (blending of red and blue to form purple) as they bi-directionally recolonize. B) This is the conventional model where individuals follow isotherms. Here, southern refuge(s) (upper panel) retain all the genetic diversity of the range (blue), and isolation by distance northern range expansion (lower panel) renders populations a series of genetic subsampling (blue gradient) from the south as individuals post-glacially move northward.

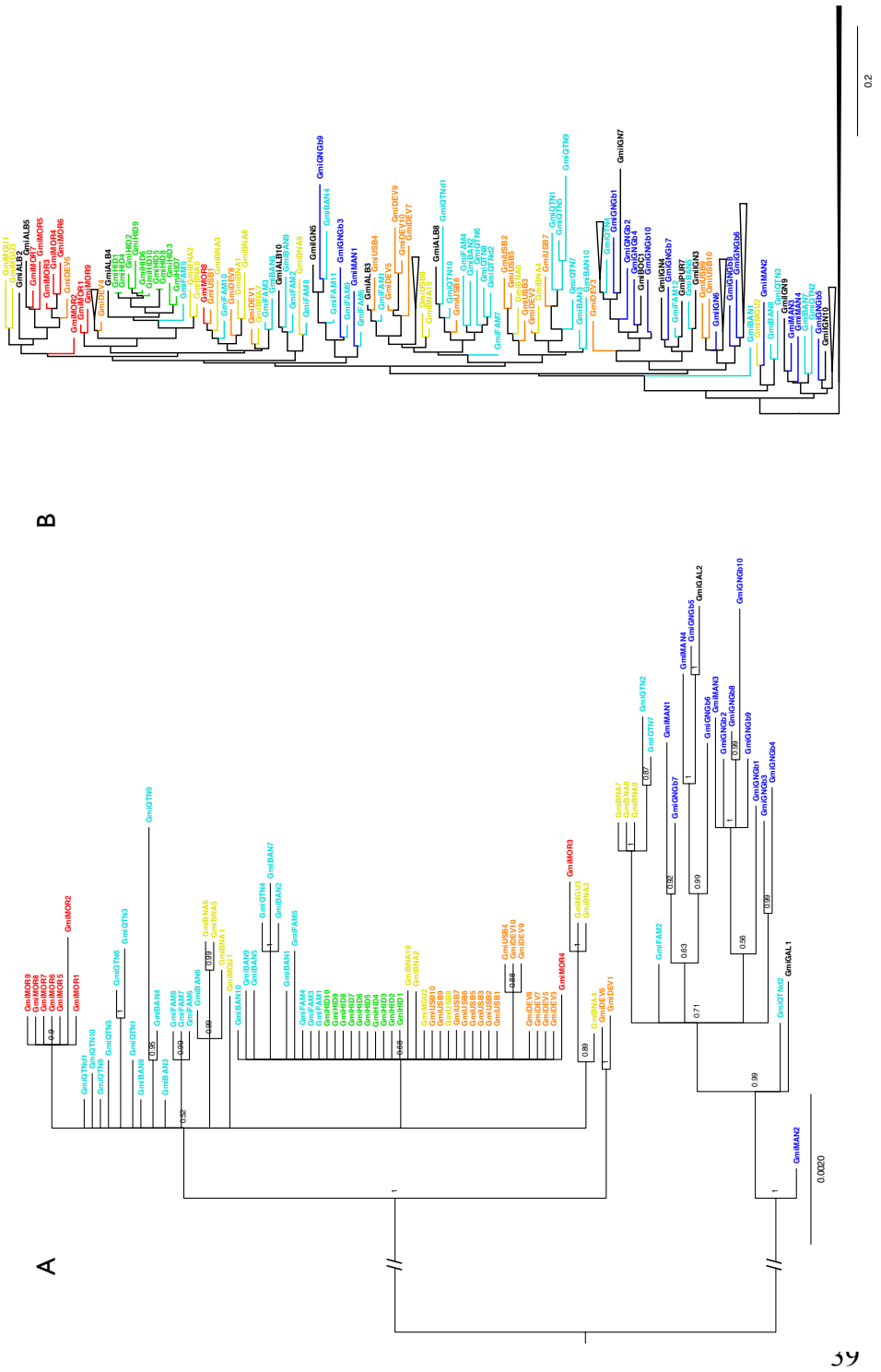

Figure S2. New tree reconstructions for *Gillichthys mirabilis*. A) Construction in MRBAYES using mtDNA (mitochondrial control region and Cyt B, 1,831 bp). Node posterior support is shown. B) Neighbour-joining tree made in POPULATIONS with the 16 microsatellite loci used in this study. Collapsed branches are samples outside the geographic region of the study. Parallel bars indicate shortened branch lengths for viewing. Individuals are colour-coded by geographic region, consistent with the scale in Table S1, with a red (north) to blue (south) gradient.

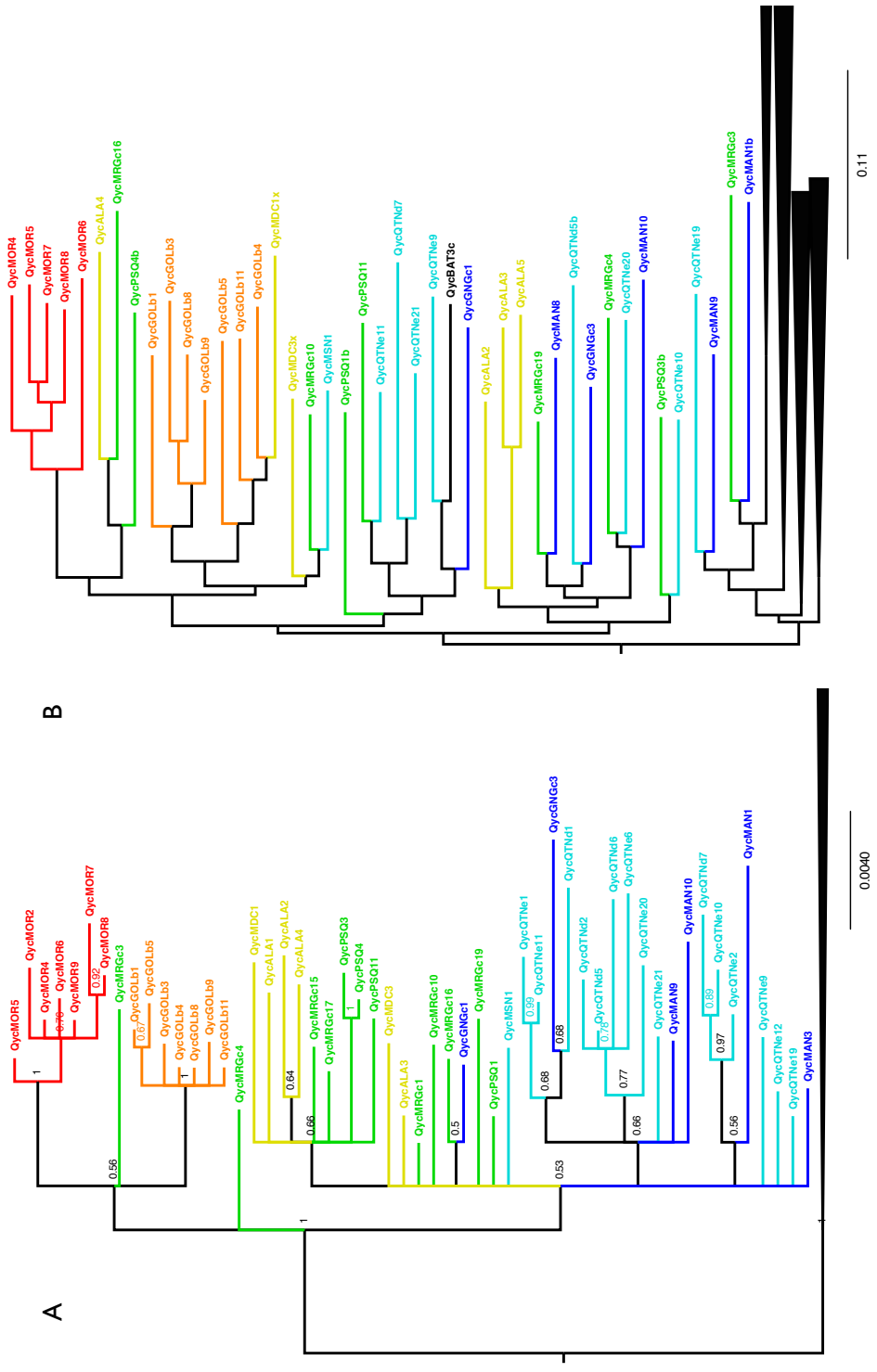

Figure S3. New tree reconstructions for *Quietula y-cauda*. A) Bayesian tree reconstructed in MRBAYES using mtDNA (mitochondrial control region and Cyt B, 1,668 bp). Node posterior support is shown. B) Neighbour-joining tree made in POPULATIONS with the 17 microsatellite loci used in study. Collapsed branches are samples outside the geographic region of the study. Parallel bars indicate shortened branch lengths for viewing. Individuals are colour-coded by geographic region, consistent with the scale in Table S1, with a red (north) to blue (south) gradient.

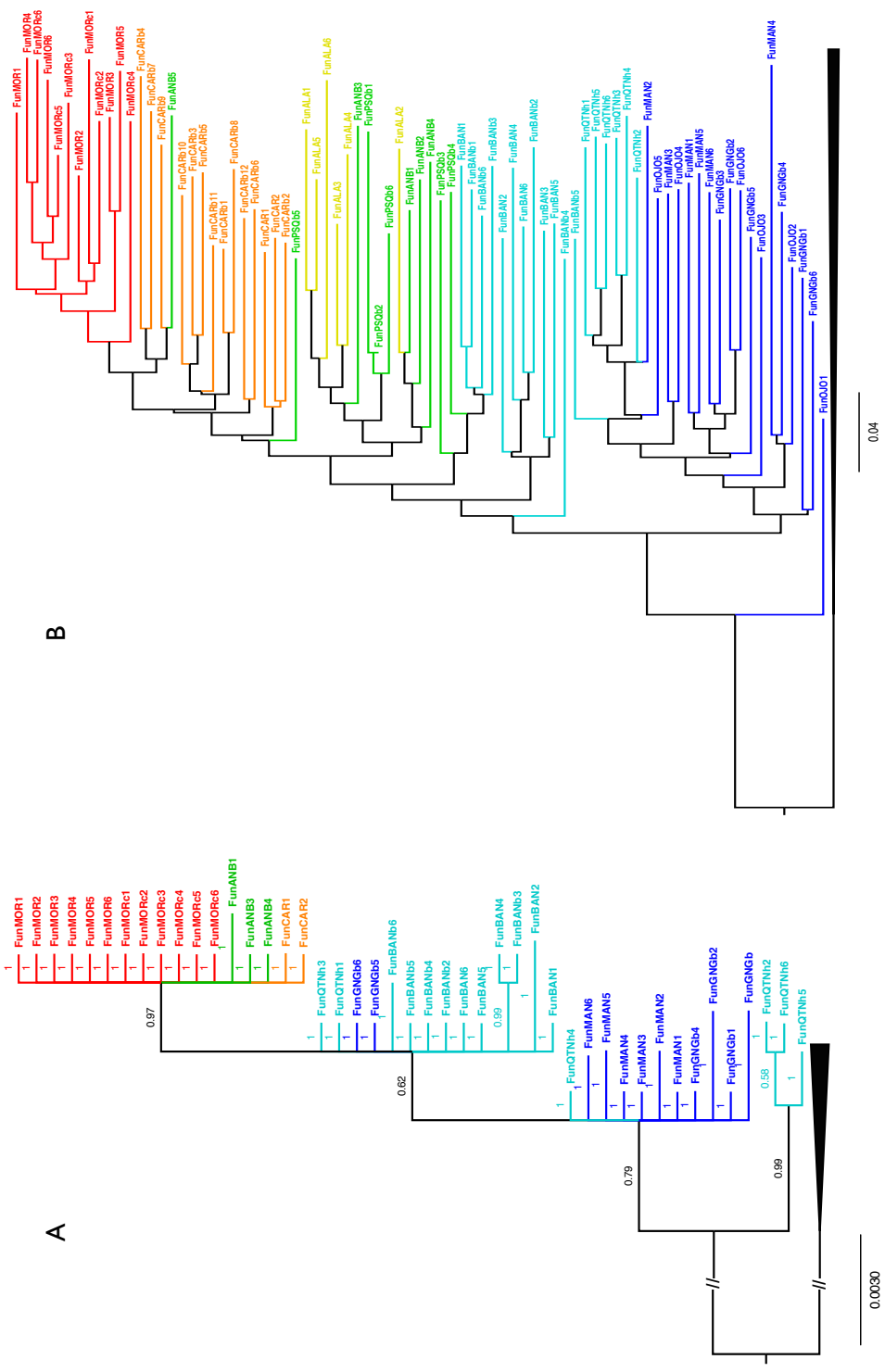

Figure S4. New tree reconstructions for *Fundulus parvipinnis*. A) Bayesian tree reconstructed in MRBAYES using mtDNA (mitochondrial control region, 883 bp). Branch posterior support is shown. B) Neighbour-joining tree made in POPULATIONS with the 20 microsatellite loci used in study. Collapsed branches are samples outside the geographic region of the study. Parallel bars indicate shortened branch lengths for viewing. Individuals are colour-coded by geographic region, consistent with the scale in Table S1, with a red (north) to blue (south) gradient.

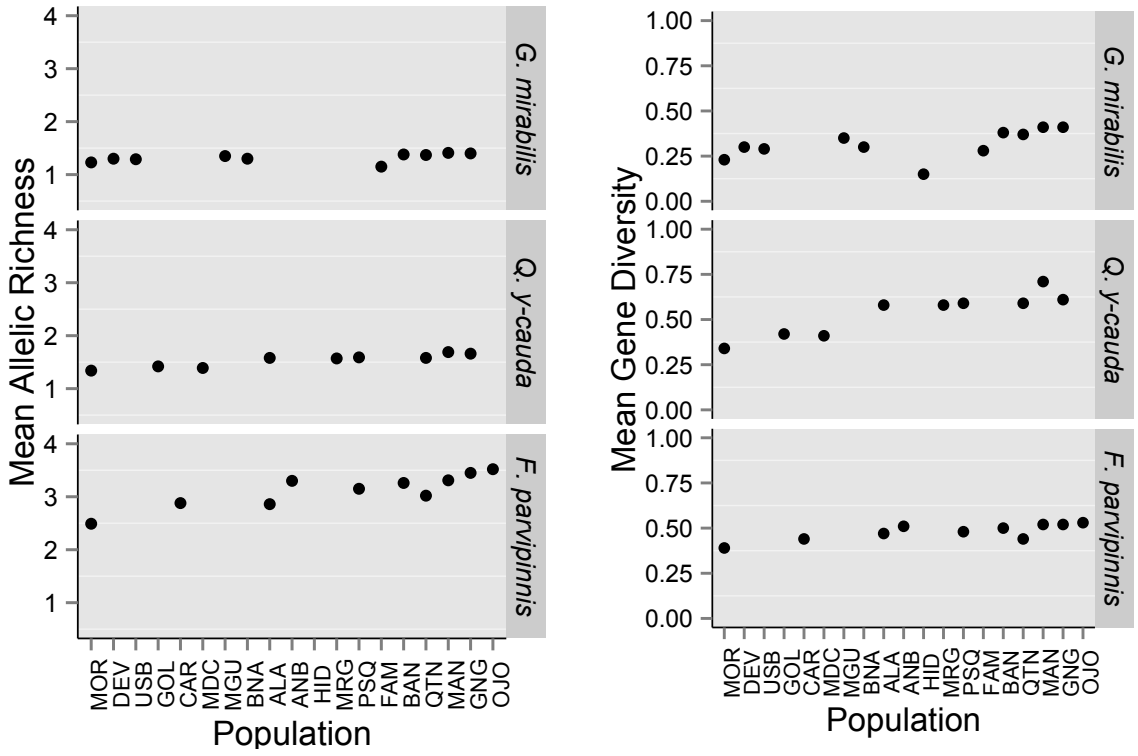

Figure S5. Diversity indices. Mean allelic richness (A) and mean gene diversity (B) for species (listed at right, top to bottom); populations oriented north (left) to south (right) on the x-axis. Sites thought to be admixed are not higher nor lower in diversity than refuge source populations (MOR, MAN, GNG, OJO). Note that the x-axis is not absolute geographic distance because population sites are not equidistantly spaced along the coast.

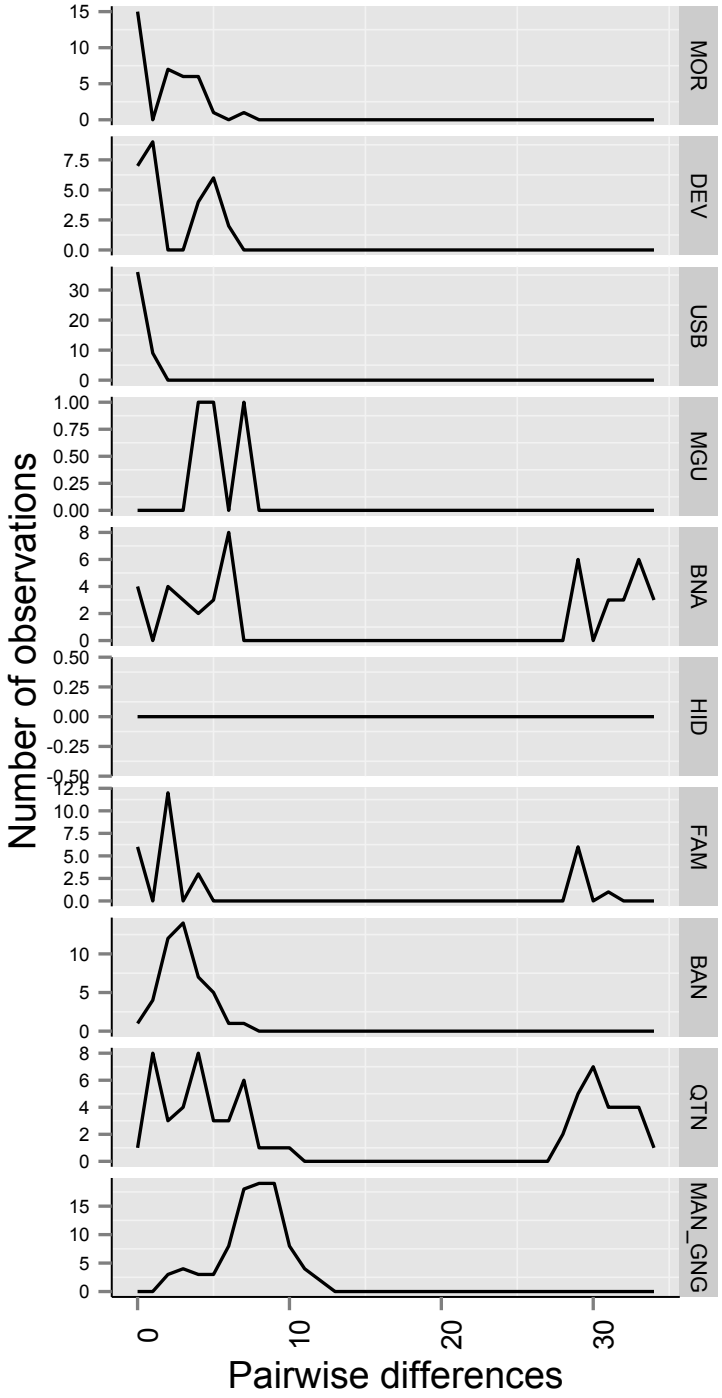

Figure S6. Mismatch distribution. Pairwise differences per population for observed haplotypes in *Gillichthys mirabilis*. Sites ordered north to south. Bi- and multi-modal distributions are observed in sites inferred to be admixed, and unimodal distributions are observed in expected refuges (MOR, MAN_GNG). One site (HID) is monotypic and likely a founder bottleneck.

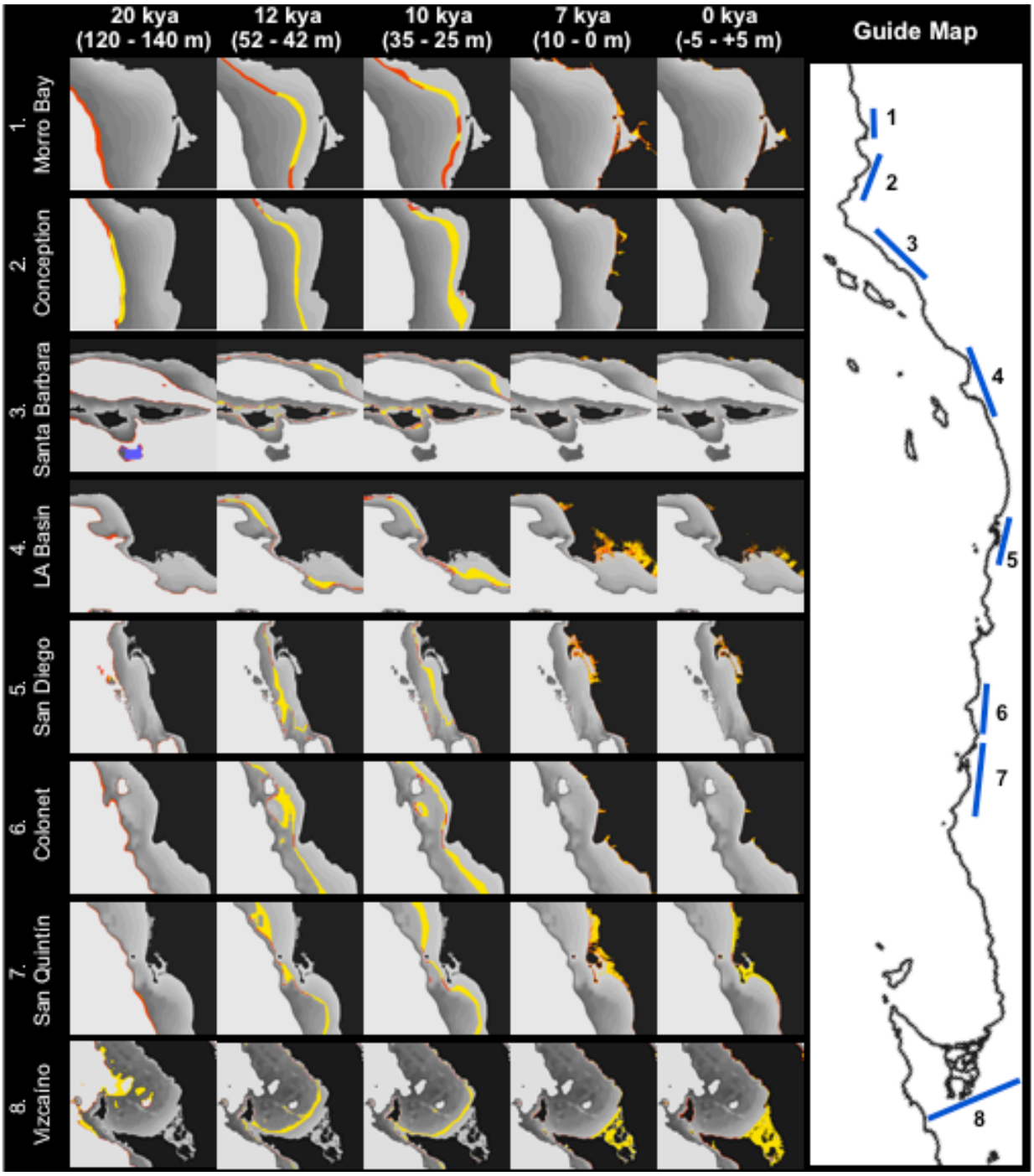

Figure S7. Time-series habitat maps. Locations of inhabitable (yellow) and uninhabitable (red) area for regions (rows) along the coast (see guide map) for specific time points (columns). Time points and depth as meters below present sea level are listed for each column with 0 kya extending to 5 meters above present sea level (+5). Purple denotes areas that meet slope but not the minimum upland drainage area requirement to form estuarine habitat. Coastal regions are not of equal size.

Age (kya)	19.5	18.5	17	15	14	13.5	13	11.5	11	10	9.5	8.5	7.5	6.5	0
mbpsl	140 - 130	130 - 120	120 - 110	110 - 100	100 - 90	90 - 80	80 - 70	70 - 60	60 - 50	50 - 40	40 - 30	30 - 20	20 - 10	10 - 0	0 +/- 5
Morro Bay	0	0	5	14	22	24	21	17	20	13	15	14	21	14	10
Conception Refuge	25	31	13	12	28	12	38	6	66	63	62	83	36	14	7
Santa Barbara Channel	0	0	10	25	59	79	101	141	101	113	131	183	134	39	46
LA Basin	0	1	3	2	3	9	43	31	75	87	97	162	142	46	68
San Diego	3	2	6	12	38	70	54	59	95	95	131	96	154	82	105
Punta Banda	0	0	24	33	45	36	40	44	41	37	68	143	68	82	45
Colonet	7	6	27	29	41	131	204	95	124	86	66	59	41	43	48
San Quintín	2	15	50	60	100	85	104	137	44	119	62	105	74	160	179
Vizcaíno Refuge	519	821	1028	1806	2816	2803	2189	1124	772	827	729	817	739	1495	1665

Figure S8. Habitat area per depth-time. Listed are habitat areas (km^2) for each time-depth bin in each coastal regions (left), which are ordered by latitude. Cells are coloured by habitat abundance from low (red) to high (blue): 0–5 km^2, red; 5–15 km^2, orange; 15–30 km^2, yellow; 30–60 km^2, green; 60–150 km^2, teal; >150 km^2, blue. These values are the total summed polygon area per coastal region that meet slope requirements. Coastal regions are not of equal area. They represent areas of contiguous habitat formation might occur.

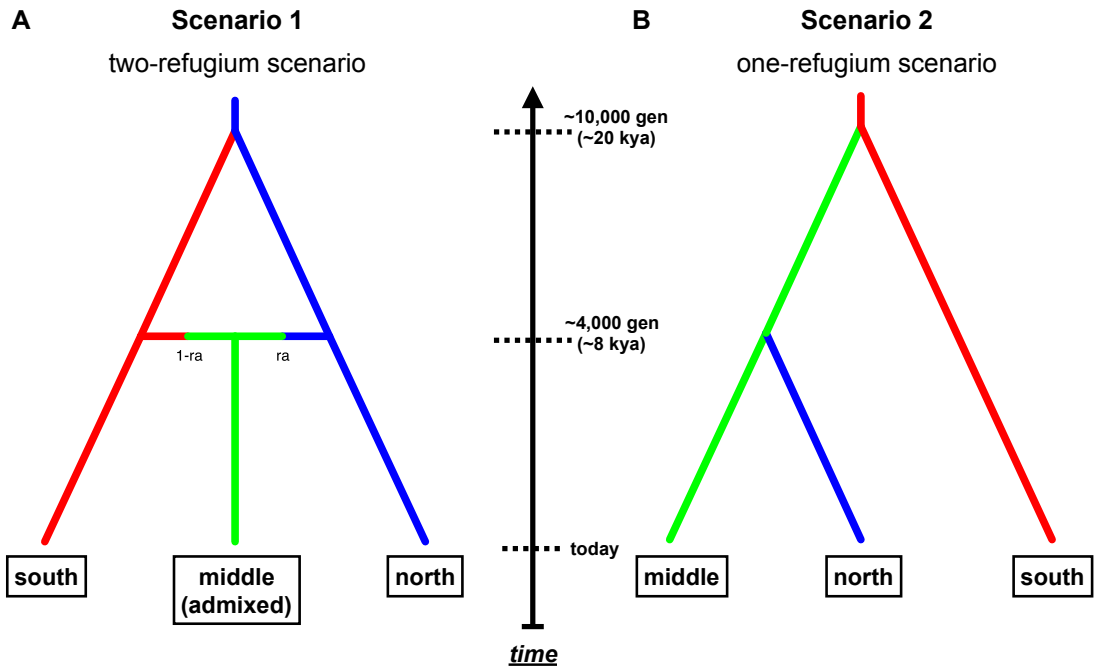

Figure S9. Approximate Bayesian Computation (ABC) scenarios. These are the models used to test statistical support for a two-refugium (Scenario 1) versus one-refugium (Scenario 2) scenario. In Scenario 1 there are two refugia: north (blue) and south (red), and these admix ~8 kya (4 thousand generations ago) to form the middle populations, which are grouped in the middle 'admixed' group. Scenario 2 models an alternative hypothesis of one-refugium (southern, red) that subsequently colonizes northward to form the middle and northern groups. Populations were grouped the same for these scenarios and tested in each of the three species.

Table S1. Sample locales. Sites of collections used in this study (bold) and sites only used in Figures S1–S3 (not bold), corresponding 3-letter codes and coordinates in decimal degrees. Number of individuals per site listed; rough linear distance from the northernmost site in this study (Morro Bay) was calculated in GOOGLE EARTH using the path tool and following the general orientation of the coastline. These geographic distances are used in the regression analysis in Figure 4. Colours correspond to colour coding in Figures S1–S3.

Site location	code	latitude°	longitude°	distance from MOR (km)	SI Fig 1 color	Number of Individuals (N)		
						F. parvipinnis	*G. mirabilis*	*Q. y-cauda*
Albany race track	ALB	37.889333	-122.311683	-	black	-	-	-
Morro Bay	MOR	35.348517	-120.8336	0	red	12	9	5
Devereaux Slough	DEV	34.41735	-119.873983	176	orange	-	10	-
U. Santa Barbara	USB	34.409383	-119.845017	179	orange	-	10	-
Goleta Slough	GOL	34.417046	-119.839374	181	orange	-	-	7
Carpenteria	CAR	34.400167	-119.538667	211	orange	14	-	-
Mandalay Canal	MDC	34.136892	-119.183952	256	yellow	-	-	2
Point Mugu	MGU	34.11391	-119.0821	269	yellow	-	3	-
Ballona Lagoon	BNA	33.962764	-118.4458	334	yellow	-	10	-
Alamitos Bay	ALA	33.745519	-118.117547	391	yellow	6	-	5
Anaheim Bay	ANB	33.736302	-118.093844	394	green	5	-	-
Catalina Island	CAT	33.430928	-118.50608	448	green	-	-	1
Hidden Lagoon	HID	33.275532	-117.451668	474	green	-	10	-
Santa Margarita	MRG	33.234	-117.410833	480	green	-	-	5
Penasquitos	PSQ	32.9325	-117.258	517	green	6	-	4
Mission Bay	MSN	32.770833	-117.232333	538	cyan	-	-	1
Famosa Slough	FAM	32.751155	-117.228381	539	cyan	-	12	-
Punta Banda	BAN	31.765157	-116.617381	678	cyan	12	10	-
San Quintín	QTN	30.418794	-116.023086	872	cyan	6	12	8
Laguna Manuela	MAN	28.247533	-114.085517	1266	blue	6	4	4
Guerrero Negro	GNG	28.021722	-114.114667	1290	blue	6	10	2
Ojo de Liebre	OJO	27.78305	-114.3129	1323	blue	6	-	-
la Bocana	BOC	26.789283	-113.675733	-	black	-	-	-
Ignacio lagoon	IGN	26.818667	-113.1815	-	black	-	-	-
el Cuarente	CUA	26.556133	-113.0028	-	black	-	-	-
Batequi	BAT	26.42715	-112.776733	-	black	-	-	-
Purisima	PUR	26.06265	-112.282083	-	black	-	-	-
el Rosario	ROS	25.698083	-112.074717	-	black	-	-	-
el Tambor	TAM	24.831932	-112.055708	-	black	-	-	-
Punta Pajaro	PPJ	24.753467	-112.043317	-	black	-	-	-
Salinas	SAL	24.582114	-111.787706	-	black	-	-	-
Gallinitas	GAL	24.557442	-111.735303	-	black	-	-	-

Table S2. Microsatellite primers. Listed are primers developed for this project. *Gillichthys mirabilis* primers unlisted here are available in [1]. All forward primers in this study were labelled at the 5' end with the M13 complement: 5 'AGGGTTTTCCCAGTCACGACGTT '3

Species	Marker	Forward (5' - 3')	Reverse (5' - 3')
F. parvipinnis	**FMA02**	ATTTACGGCAACCACCTGC	AACCCTAGCTAACGCCTCC
F. parvipinnis	**FMA03**	TCCTGACCATCATAACAGATTTCG	CCTACCTGGCCAACAGC
F. parvipinnis	**FMA04**	GGAGGTAAACAGGGCACAG	CAGCATCCAGCAGCTTTCC
F. parvipinnis	**FMA05**	TCGAGTTGATCCAACAGATTGC	AGAGGCGGAAACATCCCTG
F. parvipinnis	**FMA07**	TCCAGTCTGAGCAAACTCC	ACGCAGGACACAGTTAGCC
F. parvipinnis	**FMA08**	GCCAACGTCAAGTCTCAAG	CTCGCCCATTGTATGCTGG
F. parvipinnis	**FMA09**	GAAGCAGGAATGGGTAGCG	AGTCAGTCCCAAACAGTCG
F. parvipinnis	**FMA10**	CACGCCTTTAACACGTCGG	CCTGGGAACGCCTTGGG
F. parvipinnis	**FMA13**	AACCCTGACCTGTATCGGC	CTGGCCTTTATCATGCTTTCC
F. parvipinnis	**FMA14**	TCATGCAAAGGTTAGTGTCGG	GAGGAGCTGGCCCAAGTAG
F. parvipinnis	**FMA15**	GCCTTGTACATAGAGCGTGG	GTGATCTTGTTGTGTACGGC
F. parvipinnis	**FMA16**	CCAGGAGAGACCATGGGAC	TTGACAGCTGGAGACAGGC
F. parvipinnis	**FMA18**	GTTCCCTGCAAGAACAGACG	CTCCAAGAGAATGTCGGGC
F. parvipinnis	**FMA19**	CGCTCCAGACAGCTAATGC	ATTCACGGTGCTACGGAGG
F. parvipinnis	**FMA21**	CCCACTCAACATACCAAGCTG	TCCATGCCAGTCATAGGCG
F. parvipinnis	**FMA23**	TCCTCCCGCTTTCATTCCG	GACTGCAGCCCAGATGTTG
F. parvipinnis	**FMA24**	CTCCAGCCACACTTTATGCG	CGGTGAATGTGCTCCAAGG
F. parvipinnis	**FMA25**	CAGAGCATCACAGAACCTCG	GTGGACTCTGATTTGCTGCC
F. parvipinnis	**FMA26**	CAGCCGCCAAATTAGAAAGC	TCCCATGCTGCAACTTGTTC
F. parvipinnis	**FMA29**	GCTACACTACCCACCTCTGG	GCATGCAGGCGCTCAACAAG
G. mirabilis	**GMA01**	GATTCCGATTCCAATGTTC	TTGCAACTTACAAGAAATTCAC
G. mirabilis	**GMA03**	TTGAAGACGTACAGCACCAC	CCAGTCAGAATGTGTTCCAC
G. mirabilis	**GMA08**	TAATGACGCAGTGTTTGATG	CTGTGTGCCTTGAAGGTG
G. mirabilis	**GMA14**	CATGAATTTAGCACCATCATC	TTCTTGTGGAGTCTCTTCAAAG
G. mirabilis	**GMA20**	GACTCTTTGTCCAGCATTTC	TGTTATTCAAGTGCCATCATC
Q. y-cauda	**QMA01**	CTGTGACTTTGGGCATTAG	AATGCCCTGGTTATCTGTC
Q. y-cauda	**QMA03**	CGACATTCACGACACAAATC	ACGAATTTGACCTGAGAGC
Q. y-cauda	**QMA04**	AATGAAACGGTGAAAGAAAC	TTCAGCTCCTTCAGTTTGAC
Q. y-cauda	**QMA05**	TTCTTTCTTGCCTTGTCC	CATGAAGGCACGAAAGAG
Q. y-cauda	**QMA06**	GACTGTTCCATGTTCCTGTG	TCAGAGCAGTTTAATCCAAAG
Q. y-cauda	**QMA07**	CTTCCTCCACTCTCTCACAG	AGCGACGTACTTCTGAAGAG
Q. y-cauda	**QMA08**	ACTGAAGCTCCAAGGACAC	TGATTGTGCTGTGACTCATG
Q. y-cauda	**QMA09**	AGTGCAGGCATACATACATG	TTTGATTTGATGTATGCACTG
Q. y-cauda	**QMA10**	GTGATTTATGCGTCCAGATG	TTCAGGGTCGTCTTTAAATC
Q. y-cauda	**QMA13**	AGGCTCAGGACTCTCATGTAC	CTTCTCCTCTACCGCTCAG
Q. y-cauda	**QMA17**	TATTTGTCATCGCCCTAATG	CAAATTAAAGCCAATTGTTG
Q. y-cauda	**QMA24**	CCCGCTCCGTCAACACTC	CAATGGTGAGCGCGTACATG
Q. y-cauda	**QMA25**	GACATGCTCCTCGTTTGACC	CACGCCCACATTTCAAGGAC
Q. y-cauda	**QMA26**	TTCGTCTGACTGTGCTGGTTG	CTCCTGCTCGGTTCATGCC
Q. y-cauda	**QMA27**	GACTGTTCCATGTTCCTGTGAG	ACCTACTTCGACTGACTGGC
Q. y-cauda	**QMA28**	ATCTGCAGTAACGTGGGCTC	AGTGTGCTCGTGACTTATGC
Q. y-cauda	**QMA30**	TTGACTGCGCTCTTACATGG	CACGGACTGTTCGACAATATTG

Table S3. Slope measurements. Sites where at least two of the three species co-occur were measured five times, others were measured once. Run lengths vary based on what portion of the estuary was being measured and overall size of the system.

Site	Slope (%)	Run Length (km)
Morro Bay	0.499	500-600
Morro Bay	0.906	500-600
Morro Bay	1.111	500-600
Morro Bay	1.150	500-600
Morro Bay	0.363	500-600
Alamitos Bay	0.498	200-400
Alamitos Bay	0.256	200-400
Alamitos Bay	0.455	200-400
Alamitos Bay	0.578	200-400
Alamitos Bay	0.000	200-400
Banda	0.000	200-600
Banda	0.00	200-600
Banda	0.687	200-600
Banda	0.192	200-600
Banda	0.241	200-600
San Quintín	0.000	1000-1700
San Quintín	0.000	1000-1700
San Quintín	0.312	1000-1700
San Quintín	0.100	1000-1700
San Quintín	0.198	1000-1700
Vizcaíno	0.106	2000-5000
Vizcaíno	0.116	2000-5000
Vizcaíno	0.254	2000-5000
Vizcaíno	0.743	2000-5000
Vizcaíno	1.263	2000-5000
Devereaux Slough	0.424	250
Santa Barbara Channel	0.713	400
Carpenteria	0.952	100
Goleta Slough	1.330	75
Point Mugu	0.542	350
Catalina	0.298	350
Mandalay Canal	0.498	200
Ballona	0.862	120
Anaheim Bay	0.000	500
Los Penasquitos	0.571	175
Famosa Slough	0.437	250
Santa Margarita	0.305	985
Mission Bay	0.328	305

Table S4. Fixation indices. For *G. mirabilis* (A), *Q. y-cauda* (B), and *F. parvipinnis* (C) pairwise Fst values listed on the lower half of the table and significance indicated (p value < 0.01) on the upper half. Note that populations may be sample-limited for this metric, particularly in *Q. y-cauda*.

A	MOR	DEV	USB	MGU	BNA	FAM	HID	BAN	QTN	MAN	GNG
MOR		-	+	-	+	+	+	+	+	+	+
DEV	0.11259		-	-	-	-	+	-	+	-	+
USB	0.1585	-0.00284		-	-	+	+	+	+	-	+
MGU	0.02382	-0.02909	-0.00054		-	-	+	-	+	-	-
BNA	0.14001	0.04147	0.04251	-0.01184		+	+	-	+	-	+
FAM	0.15165	0.04154	0.086	0.03708	0.061		+	+	+	+	+
HID	0.36703	0.28492	0.37074	0.26191	0.30592	0.23939		+	+	+	+
BAN	0.11088	0.04977	0.07244	0.02616	0.05321	0.05381	0.35553		+	-	+
QTN	0.23894	0.16832	0.14758	0.15631	0.19337	0.15255	0.42318	0.08467		-	+
MAN	0.17583	0.06185	0.07306	0.02963	0.11049	0.11616	0.42224	-0.00465	0.04351		-
GNG	0.16641	0.11276	0.13273	0.08537	0.15241	0.09274	0.2881	0.05073	0.06413	-0.00933	

B	MOR	GOL	CAT	MDC	ALA	MRG	PSQ	MSN	QTN	MAN	GNG
MOR		+	-	-	+	+	-	-	+	+	-
GOL	0.18882		-	-	-	+	-	-	+	+	-
CAT	0.35014	0.35305		-	-	-	-	-	-	-	-
MDC	0.02749	0.0252	0.22865		-	-	-	-	-	-	-
ALA	0.14909	0.06344	0.21702	0.02549		-	-	-	-	-	-
MRG	0.2212	0.12796	0.22574	0.08269	0.01366		-	-	-	-	-
PSQ	0.16498	0.08628	0.20494	0.03161	-0.00175	-0.02079		-	-	-	-
MSN	0.07978	0.0898	0.45455	-0.03927	0.00809	0.04493	0.02279		-	-	-
QTN	0.21292	0.13986	0.25464	0.10784	0.056	0.04772	-0.01143	0.03424		-	-
MAN	0.23235	0.21966	0.25788	0.12299	0.08238	0.09591	0.0878	0.03725	0.08041		-
GNG	0.30393	0.21731	0.32613	0.14373	0.10027	0.02961	-0.01873	-0.00267	0.00494	0.0298	

C	MOR	CAR	ALA	ANB	PSQ	BAN	QTN	MAN	GNG	OJO
MOR		+	+	+	+	+	+	+	+	+
CAR	0.12178		+	+	+	+	+	+	+	+
ALA	0.18952	0.1248		-	+	+	+	+	+	+
ANB	0.14079	0.06373	0.04448		-	+	+	+	+	+
PSQ	0.18514	0.0925	0.0784	0.03604		+	+	+	+	+
BAN	0.20903	0.1308	0.10186	0.11841	0.08601		+	+	+	+
QTN	0.36129	0.29639	0.23719	0.25691	0.26707	0.111		+	+	+
MAN	0.2809	0.19879	0.16387	0.15858	0.17707	0.1074	0.09602		-	-
GNG	0.23274	0.16499	0.1211	0.10075	0.12859	0.08157	0.1129	0.01897		-
OJO	0.26223	0.17876	0.13097	0.13189	0.13236	0.04938	0.07599	0.0082	0.0167	

Table S5. AICc regression values. Comparison of AICc values for linear, quadratic, and cubic regressions of northern and southern allele counts versus geographic distance. Yellow cells indicate the favoured regression for each allele set.

Taxon	AICc regression scores		
	linear	quadratic	cubic
G. mirabilis- North	369.045	368.924	368.88
G. mirabilis- South	297.835	299.836	301.451
Q. y-cauda- North	208.753	209.045	211.482
Q. y-cauda- South	177.798	179.985	181.337
F. parvipinnis- North	378.669	347.987	343.632
F. parvipinnis- South	354.81	354.537	335.681

Table S6. Refugium model values. Comparison of p-value and corrected Akaike Information Criterion (AICc) scores for different refugium scenarios (left column). Results are from Generalized Linear Models using the predictive variables identified via DFA (Maximum Polygon Area and Summed Habitat Area). Asterisks denote significant values, daggers denote models performed with Firth's Biased Adjustment estimates, double daggers denote models run with False Discovery Rate. After corrections, only the Vizcaíno and North Conception refugium model scenario is statistically significant (shown in yellow). A Vizcaíno only refugium is not supported.

Refugium Model Scenario	Habitat	
	p-value	AICc
Vizcaíno + N. Conception	0.0085*	18.0000
Vizcaíno + N. Conception†‡	0.0240*	12.9156
Vizcaíno + Morro Bay	0.0813	17.2383
Vizcaíno + Morro Bay + N. Conception	0.0599	23.8280
Vizcaíno + Morro Bay + N. Conception†‡	0.1297	25.3831
Vizcaíno + Morro Bay + N. Conception + LA Basin	0.2165	20.1049
Vizcaíno + LA Basin	0.0889	15.4952
Vizcaíno + LA Basin†‡	0.2716	24.9661
Vizcaíno + Santa Barbara Channel	0.0889	15.4952
Vizcaíno + Santa Barbara Channel†‡	0.2716	24.9661
Vizcaíno only	0.0433*	10.8000
Vizcaíno only†‡	0.1147	20.0641
San Quintín + N. Conception	0.1786	16.8895
Vizcaíno + San Quintín + N. Conception	0.0633	16.7370
Vizcaíno + San Quintín + N. Conception†‡	0.1146	25.1359

Table S7. Approximate Bayesian Computation (ABC) results. Posterior scores are shown for a two-refugium (scenario 1) and a one-refugium (scenario 2) model using both direct and logistic sampling of the posterior. A two-refugium model (scenario 1) is supported in all cases (higher posterior scores for each comparison are bolded). An estimation of the error rates (posterior predictive error) shows errors ranging from 14 % to 23 %.

		Gllicthys mirabilis	***Quietula y-cauda***	***Fundulus parvipinnis***
		(Scenario 1:2)		
Posterior (Direct)	*200 closest*	**0.60**:0.40	**0.60**:0.40	**0.71**:0.29
	500 closest	**0.59**:0.41	**0.54**:0.46	**0.66**:0.34
Posterior (Logistic)	*8,000 closest*	**0.97**:0.03	**0.73**:0.27	**0.98**:0.02
	20,000 closest	**0.94**:0.06	**0.70**:0.30	**0.98**:0.02
Posterior predictive error		computed over 1000 data sets		
	Direct:	0.15	0.23	0.16
	Logistic:	0.14	0.22	0.15