Science Advances
AAAS

Supplementary Materials for

Barriers to gene flow in the deepest ocean ecosystems: Evidence from global population genomics of a cosmopolitan amphipod

Johanna N. J. Weston *et al.*

Corresponding author: Johanna N. J. Weston, johanna.weston@whoi.edu

Sci. Adv. **8**, eabo6672 (2022)
DOI: 10.1126/sciadv.abo6672

The PDF file includes:

Figs. S1 to S10
Tables S1 to S9
References

Other Supplementary Material for this manuscript includes the following:

Data file S1

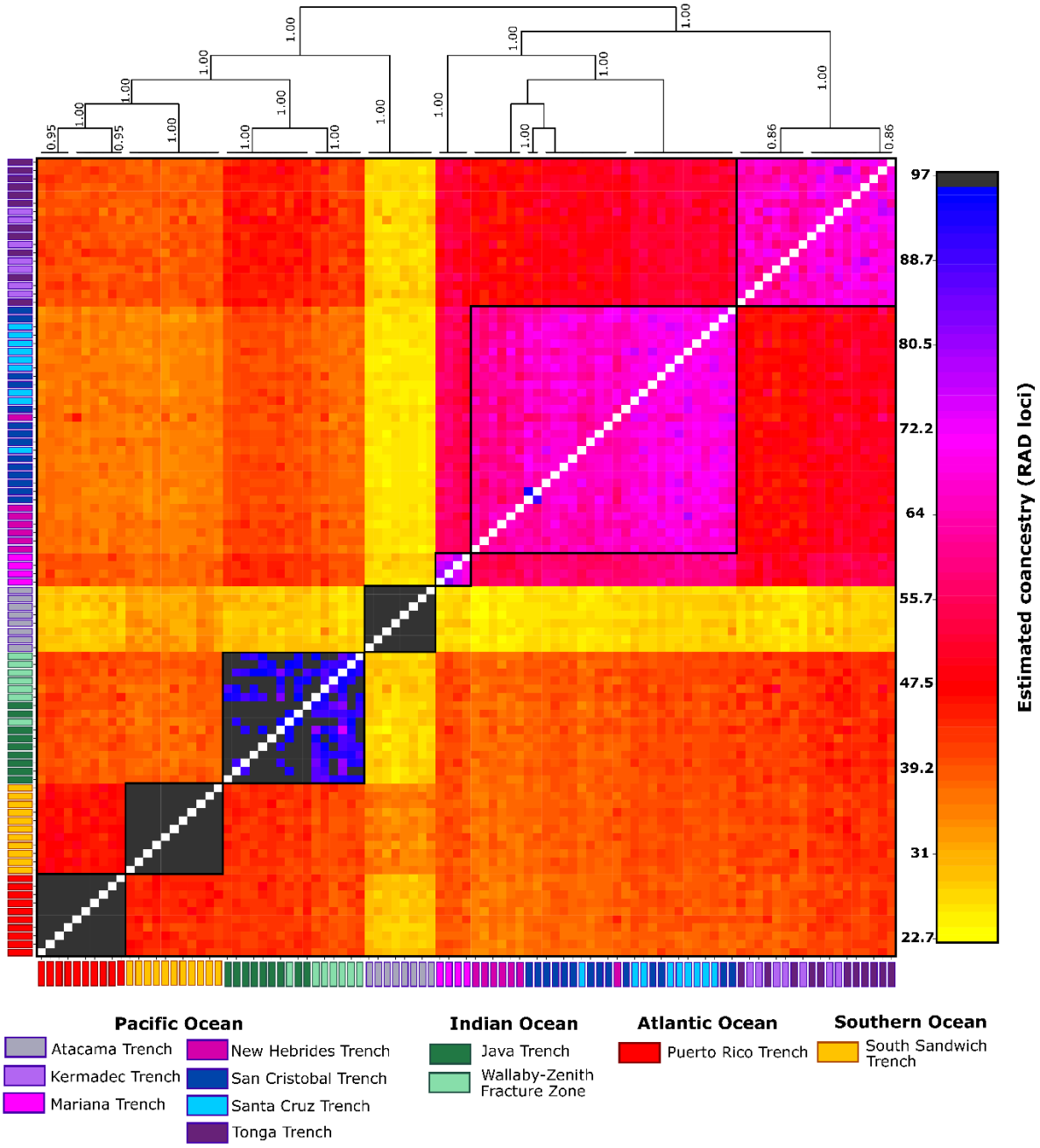

Fig. S1.

The fineRADstructure coancestry matrix resolves the same seven genetic clusters as in other population genetic structure analyses and identifies fine-scale coancestry among 11 hadal feature populations and 97 individuals. The input RAD dataset consists of all 4,611 loci and 58,730 SNP sites available at the end of the ipryad v.0.9.68 denovo pipeline, which has not been further filtered by VCFtools 0.1.16 and thinned to one SNP per loci.

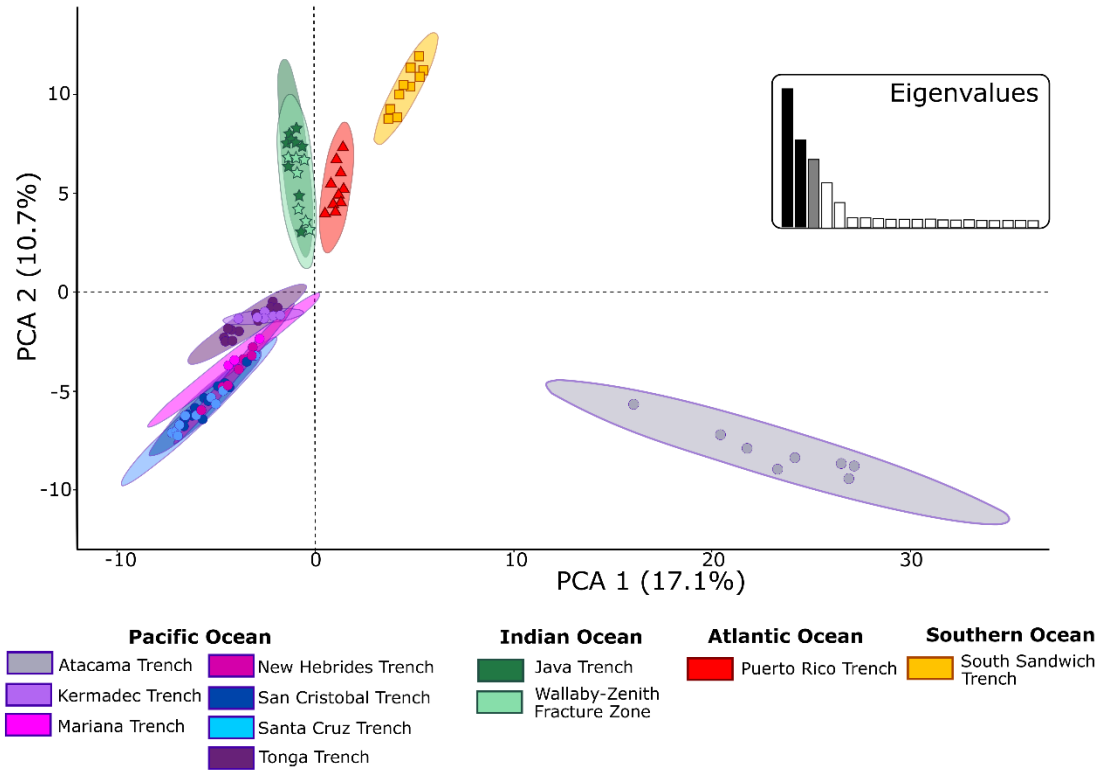

Fig. S2.

Principal Component Analysis (PCA) with visualization of the first two principal components with all 11 feature populations based on 2,933 SNPs.

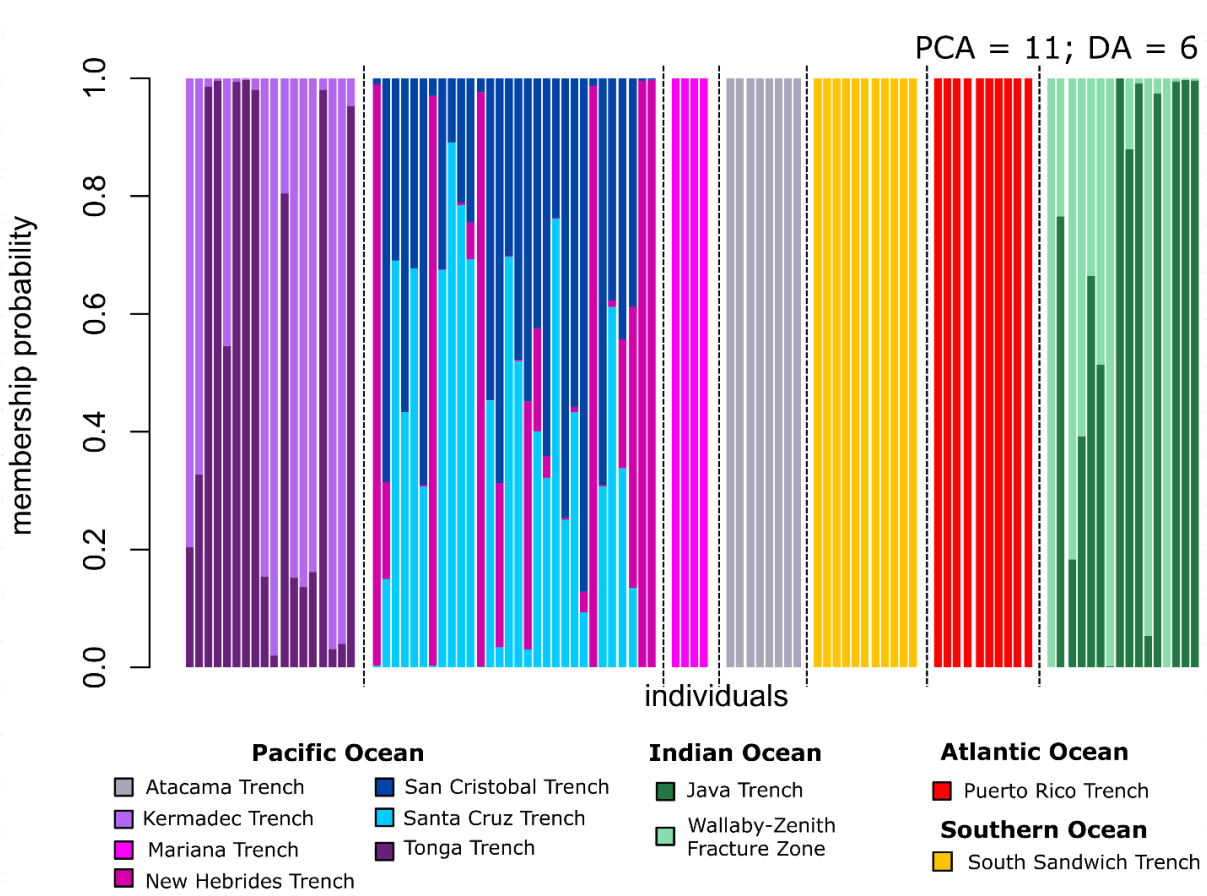

Fig. S3.

Discriminant Analysis of Principal Components (DAPC) that resolved seven genetic clusters among 11 hadal feature populations based on 2,933 SNPs.

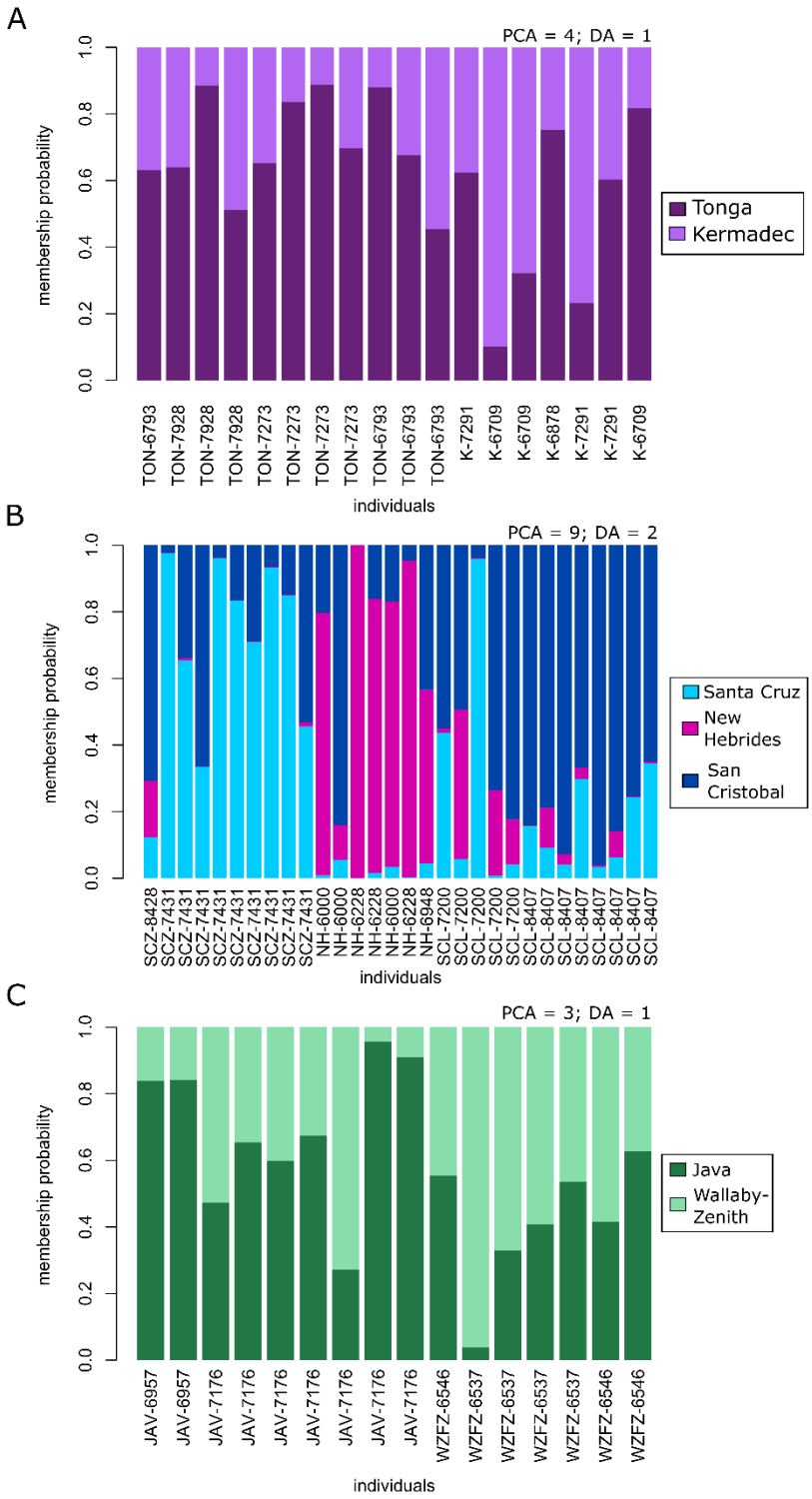

Fig. S4.

Discriminant Analysis of Principal Components (DAPC) based on 2,933 SNPs for **(A)** the Tonga and Kermadec trench populations; **(B)** the Santa Cruz, New Hebrides, and San Cristobal trench populations; and, **(C)** the Java Trench and Wallaby-Zenith Fracture Zone populations.

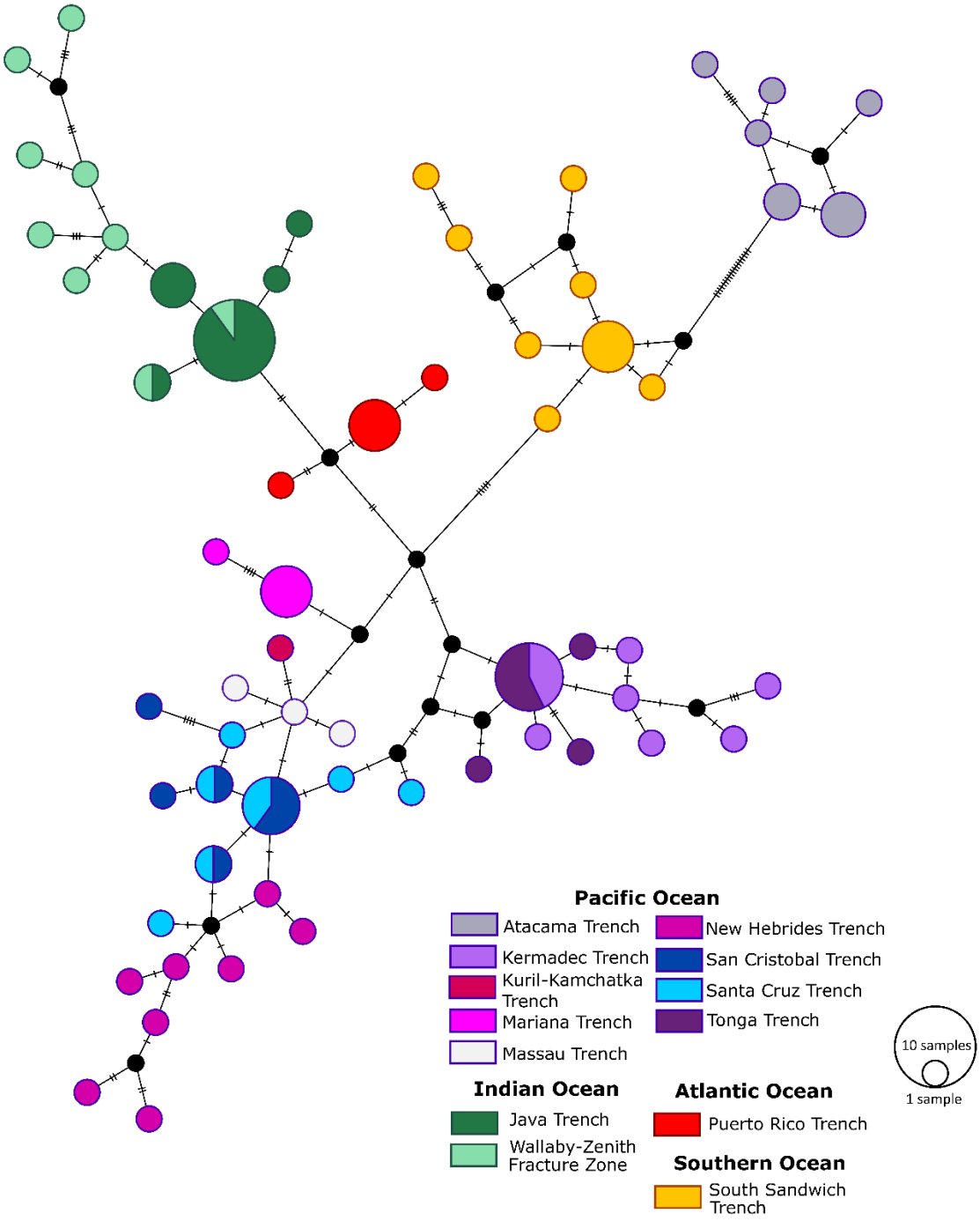

Fig. S5.

Statistical parsimony haplotype network of *Bathycallisoma schellenbergi* based on a concatenated dataset (16S rRNA + COI; 774 bp). Colors represent an individual hadal feature, with color shade groups by the ocean. The circle size is proportional to the number of individuals with a given haplotype. The lines between haplotypes denote the number of base pair differences between haplotypes. The small black dots represent predicted but not directly sampled haplotypes.

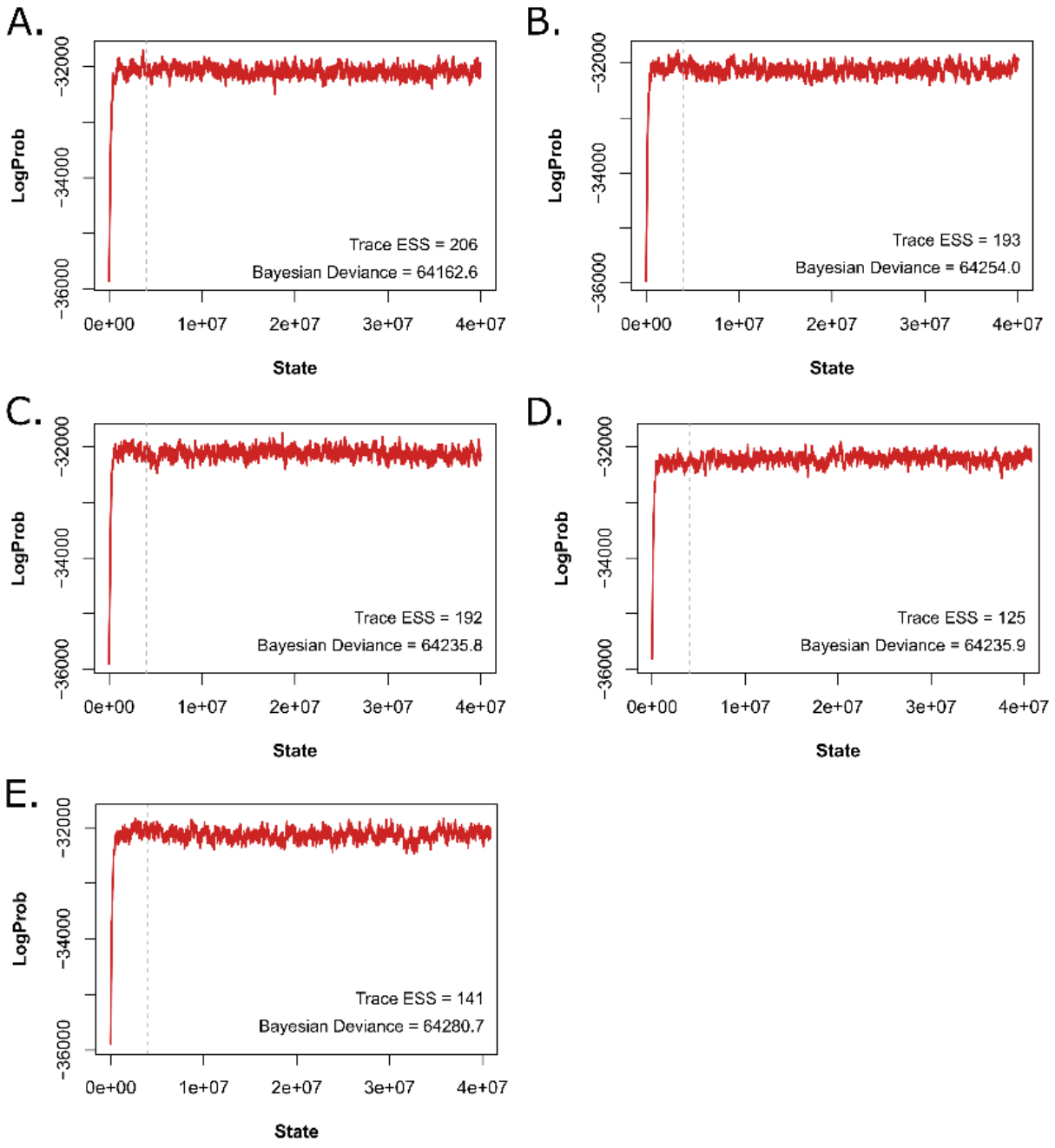

Fig. S6.

The log probabilities from five repeated BayesAss v3.04 runs for migration estimates among the *a priori* populations in the Santa Cruz, San Cristobal, and New Hebrides trenches (30 individuals). Each run includes Tracer ESS value and Bayesian Deviance, calculated with the 'calculateDeviance.R' script in Meirmans (2014). Each analysis was run for 40,000,000 MCMC iterations, 4,000,000 burn-in, and sampling interval of 5,000. Mixing parameters were set at dM=0.45, dA=1, and dF=0.1. Run 1 **(A)** was selected as the best run.

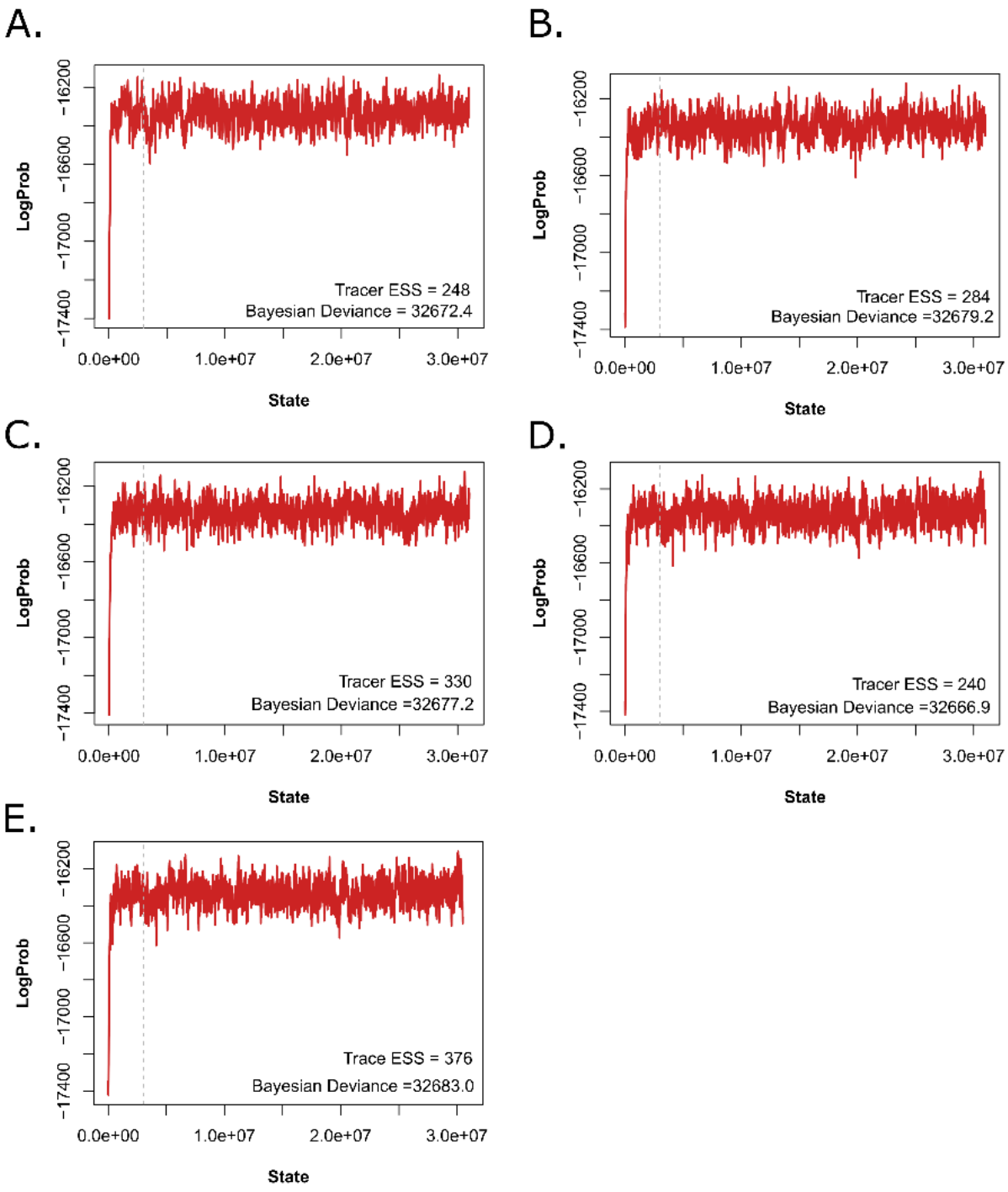

Fig. S7.

The log probabilities from five repeated BayesAss v3.04 runs for migration estimates between the *a priori* populations in the Java Trench and Wallaby-Zenith Fracture Zone (16 individuals). Each run includes Tracer ESS value and Bayesian Deviance, calculated with the 'calculateDeviance.R' script in Meirmans (2014). Each analysis was run for 31,000,000 MCMC iterations, 1,000,000 burn-in, and sampling interval of 5,000. Mixing parameters were set at dM=0.3, dA=0.9, and dF=0.1. Run 4 **(D)** was selected as the best run.

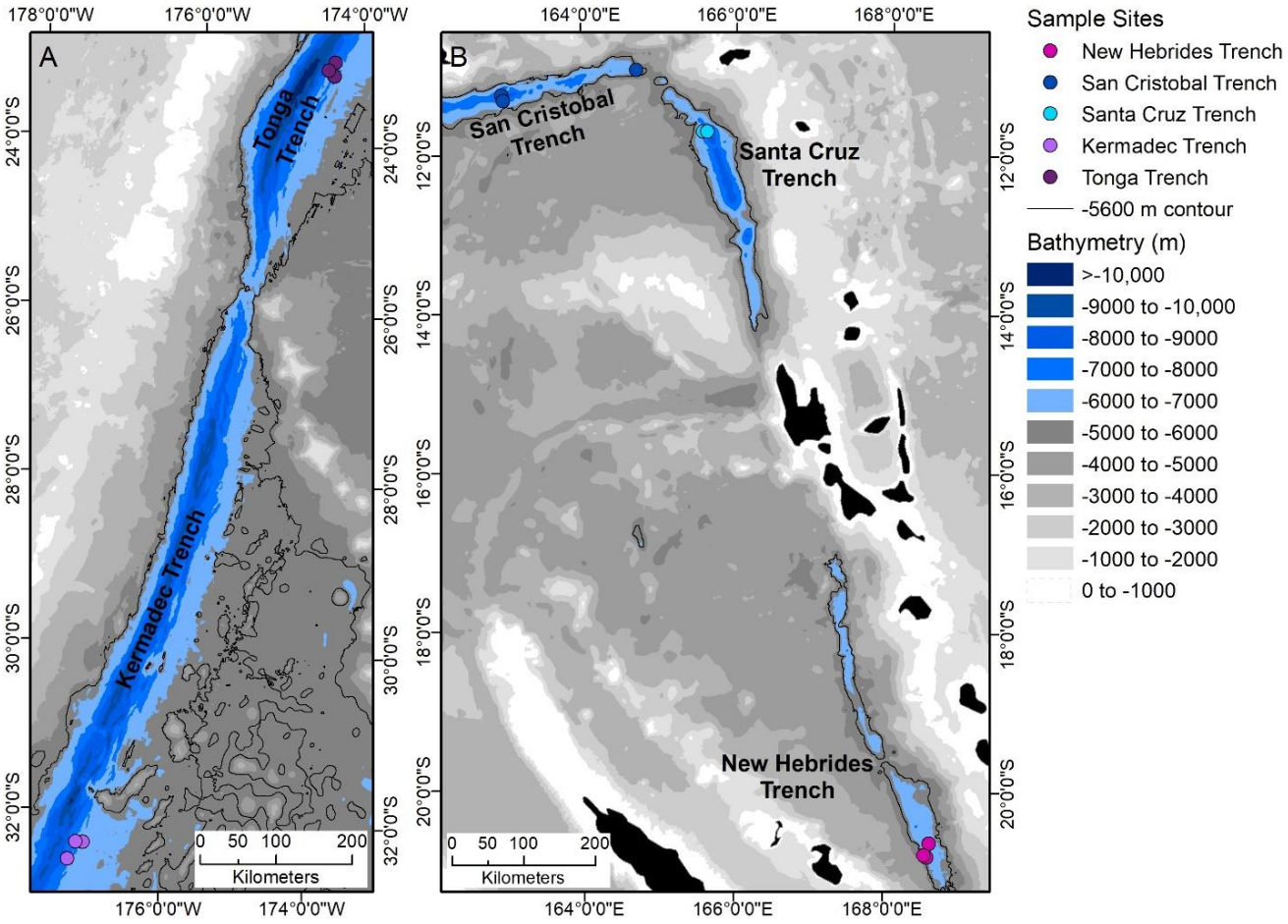

Fig. S8.

Sampling sites of the hadal amphipod, *Bathycallisoma schellenbergi*, in **(A)** the Kermadec and Tonga trenches, and **(B)** the San Cristobal, Santa Cruz, and New Hebrides trenches. Hadal depths are represented by the blue palette. The 5600 m depth contour is indicated by black lines, where this depth corresponds with the shallowest depth record for *B. schellenbergi* (*11*). All ocean elevation data was sourced from the GEBCO Compilation Group (2021). The map was created in ArcGIS v10.7.

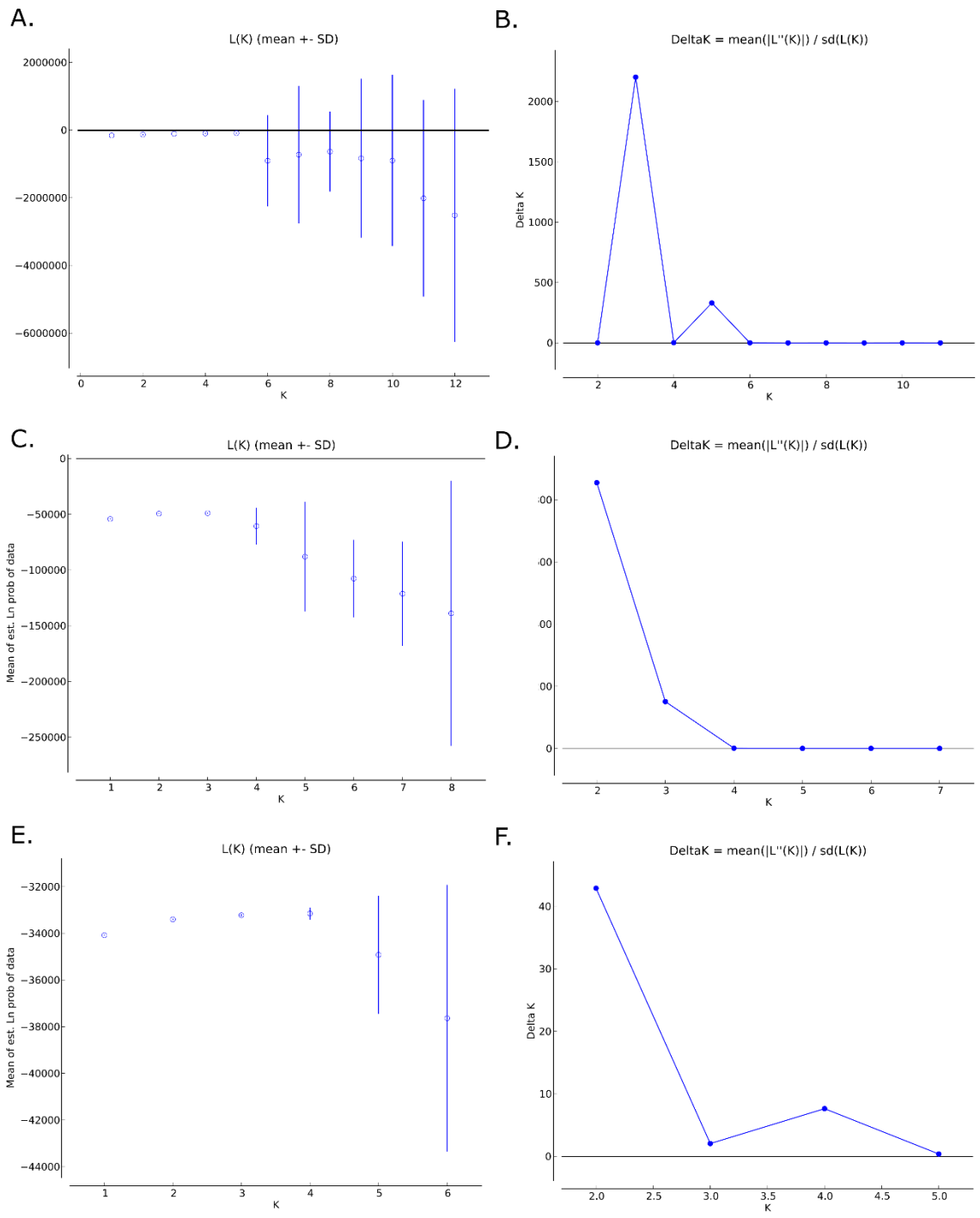

Fig. S9.

STRUCTURE HARVESTER outputs for the log probabilities (ln Pr(X|K) for STRUCTURE analysis on 2,933 SNPs. **(A-B)** populations from 11 hadal features; **(C-D)** populations from six hadal features in the western Pacific Ocean; and, **(E-F)** populations from four hadal features in the western central Pacific Ocean.

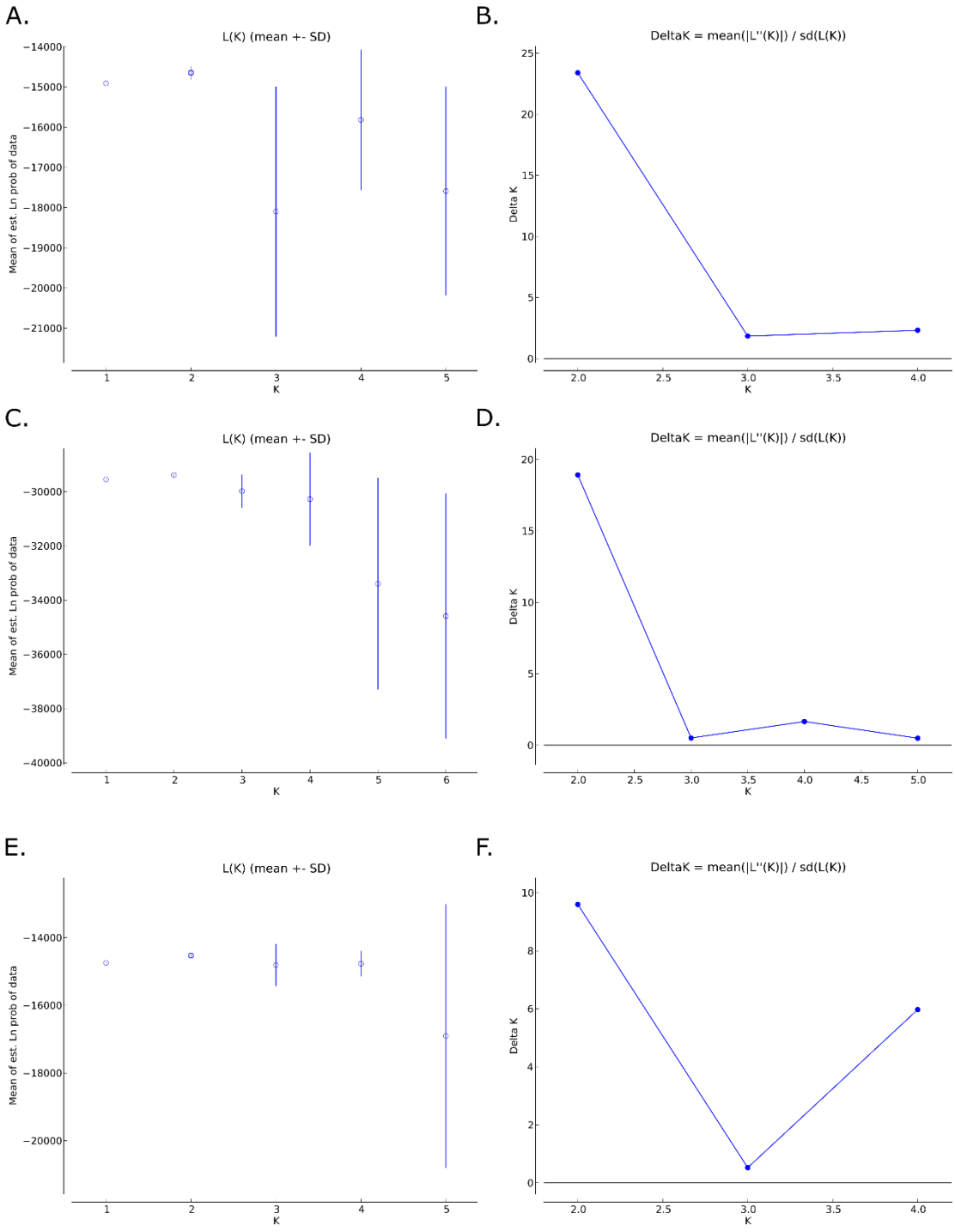

Fig. S10.

STRUCTURE HARVESTER outputs for the log probabilities (ln Pr(X|K) and the ΔK for STRUCTURE analysis on 2,933 SNPs. **(A-B)** *a priori* Kermadec and Tonga trench populations; **(C-D)** *a priori* Santa Cruz, San Cristobal, and New Hebrides trench populations; and, **(E-F)** *a priori* Java Trench and Wallaby-Zenith Fracture Zone populations.

	Atacama	Diamantina	Java	Kermadec	Mariana	New Hebrides	Puerto Rico	San Cristobal	Santa Cruz	South Sandwich	Tonga	Wallaby-Zenith
Atacama		14,907	16,316	10,125	16,305	12,000	15,935	13,062	12,797	5,499	10,332	15,909
Diamantina	14,907		2,862	7,541	7,108	7,445	18,376	8,173	8,198	8,677	8,322	1,271
Java	16,961	2,862		9,129	4,246	7,201	20,915	6,205	6,490	11,526	7,567	1,818
Kermadec	10,125	7,541	9,129		6,504	1,928	20,915	3,109	2,907	10,721	1,091	8,238
Mariana	16,305	7,108	4,246	6,504		4,577	20,915	3,407	3,640	17,375	6,085	6,064
New Hebrides	12,000	7,445	7,201	1,928	4,577		21,405	1,199	1,046	12,649	1,773	8,165
Puerto Rico	15,935	18,376	20,915	20,915	20,915	21,405		22,605	22,451	10,630	20,201	19,118
San Cristobal	13,062	8,173	6,205	3,109	3,407	1,199	22,605		288	13,830	2,730	6,530
Santa Cruz	12,797	8,198	6,490	2,907	3,640	1,046	22,451	288		13,628	2,467	8,924
South Sandwich	6,785	8,677	11,526	10,721	17,375	12,649	10,630	13,830	13,628		11,444	9,861
Tonga	10,332	8,322	7,567	1,091	6,085	1,773	20,201	2,730	2,467	11,444		9,043
Wallaby-Zenith	15,909	1,271	1,818	8,238	6,064	8,165	19,118	6,530	8,924	9,861	9,043	

Table S1.

Geographic distance matrix (km) using the semi-major axis of the WGS84 reference system as the default radius of the Earth for the 12 sampled hadal features (*95*). Where a continental landmass separated any sampled features, the geographic distance was measured directly in a 3-dimensional global scene generated in ArcGIS Pro (version 2.9). In instances where a feature had more than one sampling site, the deepest sampling sites were selected as the geographical reference point for the feature (Data S1).

Run Type	All individuals	Poor-performing individuals removed
No. individuals	120	97
Starting SNP site count	11,481	58,730
Min depth 6x and site genotyped >70% individuals	9,887	80,731
Minor allele frequency 0.05	1,356	9,258
Removed over-represented loci	1,239	9,014
Maximum of 4 SNP per tag	798	5,913
Thin loci to 1 SNP per tag	419	2,933
Individual <50% missingness	100	97
Min depth 6x and site genotyped >70% individuals	11,246	
Minor allele frequency 0.05	1,991	
Removed over-represented loci	1,831	
Maximum of 4 SNP per tag	1,080	
Thin loci to 1 SNP per tag	498	
Individual <50% missingness	98	

Table S2.

Summary of strict SNPs filtering and cleaning with VCFtools 0.1.16 to generate the final dataset constructed with ipryad 0.9.68

	Atacama (Bs1)	Java (Bs4)	Kermadec (Bs5)	Kuril-Kamchatka (Bs6)	Massau (Bs6)	Mariana (Bs6)	New Hebrides (Bs6)	Puerto Rico (Bs2)	San Cristobal (Bs6)	Santa Cruz (Bs6)	South Sandwich (Bs2)	Tonga (Bs5)	Wallaby-Zenith (Bs4)
Atacama (Bs1)	0.000-0.005												
Java (Bs4)	0.041-0.051	0.000-0.005											
Kermadec (Bs5)	0.016-0.056	0.010-0.025	0.000-0.011										
Kuril-Kamchatka (Bs6)	0.047-0.052	0.014-0.018	0.009-0.018	-									
Massau (Bs6)	0.045-0.052	0.012-0.018	0.007-0.018	0.001-0.003	0.002-0.004								
Mariana (Bs6)	0.048-0.057	0.012-0.023	0.010-0.027	0.005-0.012	0.003-0.012	0.000-0.007							
New Hebrides (Bs6)	0.046-0.063	0.016-0.032	0.010-0.028	0.005-0.018	0.003-0.018	0.007-0.023	0.000-0.014						
Puerto Rico (Bs2)	0.043-0.050	0.003-0.012	0.0126-0.025	0.0108-0.016	0.009-0.016	0.010-0.023	0.012-0.028	0.000-0.007					
San Cristobal (Bs6)	0.045-0.054	0.014-0.023	0.007-0.021	0.003-0.009	0.001-0.009	0.005-0.016	0.001-0.018	0.010-0.0216	0.000-0.013				
Santa Cruz (Bs6)	0.045-0.054	0.014-0.023	0.007-0.019	0.003-0.009	0.001-0.009	0.005-0.018	0.001-0.019	0.010-0.019	0.001-0.016	0.002-0.011			
South Sandwich (Bs2)	0.034-0.049	0.016-0.032	0.0144-0.034	0.014-0.028	0.012-0.028	0.018-0.032	0.016-0.034	0.016-0.0324	0.016-0.034	0.014-0.032	0.000-0.016		
Tonga (Bs5)	0.041-0.050	0.012-0.019	0-0.014	0.009-0.012	0.007-0.012	0.0108-0.023	0.0108-0.028	0.012-0.023	0.007-0.018	0.007-0.019	0.014-0.032	0.000-0.007	
Wallaby-Zenith (Bs4)	0.043-0.059	0-0.018	0.012-0.028	0.014-0.028	0.012-0.028	0.012-0.027	0.0162-0.036	0.003-0.019	0.014-0.032	0.014-0.030	0.016-0.034	0.012-0.0288	0.0000.020

Table S3.

Minimum and maximum pairwise p-distances for each hadal feature population of *Bathycallisoma schellenbergi* based on COI alignment (559 bp). In brackets denotes the assigned clade name

	Atacama (Bs1)	Java (Bs4)	Kermadec (Bs5)	Kuril-Kamchatka (Bs6)	Massau (Bs6)	Mariana (Bs6)	New Hebrides (Bs6)	Puerto Rico (Bs2)	San Cristobal (Bs6)	Santa Cruz (Bs6)	South Sandwich (Bs2)	Tonga (Bs5)	Wallaby-Zenith (Bs4)	Diamantina (Bs4)
Atacama (Bs1)	0.000-0.050	0.006-0.050	0.0061-0.055	0.006-0.052	0.006-0.052	0.006-0.057	0.006-0.063	0.012-0.050	0.006-0.054	0.006-0.054	0.006-0.049	0.012-0.050	0.006-0.059	0.006-0.027
Java (Bs4)		0.000-0.019	0.006-0.025	0-0.018	0-0.018	0-0.023	0-0.032	0.006-0.012	0-0.023	0-0.023	0-0.032	0.006-0.019	0-0.018	0-0.013
Kermadec (Bs5)			0.000-0.032	0.006-0.018	0.006-0.018	0.006-0.027	0.006-0.028	0-0.025	0.006-0.021	0.006-0.019	0.006-0.034	0-0.014	0.006-0.028	0.006-0.034
Kuril-Kamchatka (Bs6)				-	0-0.003	0-0.012	0-0.018	0.0064-0.016	0-0.009	0-0.009	0-0.028	0.006-0.012	0-0.028	0-0
Massau (Bs6)					0.000	0-0.012	0-0.018	0.006-0.016	0-0.009	0-0.009	0-0.028	0.0064-0.012	0-0.028	0-0
Mariana (Bs6)						0.000-0.013	0-0.023	0.006-0.023	0-0.016	0-0.018	0-0.032	0.006-0.023	0-0.027	0-0.006
New Hebrides (Bs6)							0.000-0.013	0.006-0.028	0-0.018	0-0.019	0-0.034	0.006-0.028	0-0.036	0-0.006
Puerto Rico (Bs2)								0.000	0.006-0.021	0.006-0.019	0.006-0.032	0-0.023	0.006-0.019	0.006-0.006
San Cristobal (Bs6)									0.000-0.013	0-0.016	0-0.034	0.006-0.018	0-0.032	0-0.006
Santa Cruz (Bs6)										0.000-0.006	0-0.032	0.006-0.019	0-0.030	0-0
South Sandwich (Bs2)											0.000-0.013	0.006-0.032	0-0.034	0-0
Tonga (Bs5)												0.000-0.013	0.006-0.028	0.006-0.013
Wallaby-Zenith (Bs4)													0.000-0.013	0-0.006

Table S4.

Minimum and maximum pairwise p-distances for each hadal feature population of *Bathycallisoma schellenbergi* based on 16S alignment (215 bp). In brackets denotes the assigned clade name.

	Atacama	Java	Kermadec	Mariana	New Hebrides	Puerto Rico	San Cristobal	Santa Cruz	South Sandwich	Tonga	Wallaby -Zenith
Atacama		0.747	0.748	0.760	0.743	0.772	0.743	0.739	0.768	0.752	0.757
Java	0.747		0.365	0.416	0.395	0.490	0.392	0.389	0.515	0.364	0.059
Kermadec	0.748	0.365		0.251	0.225	0.462	0.217	0.214	0.483	0.040	0.371
Mariana	0.760	0.416	0.251		0.167	0.504	0.158	0.169	0.528	0.257	0.425
New Hebrides	0.743	0.395	0.225	0.167		0.482	0.051	0.052	0.506	0.224	0.403
Puerto Rico	0.772	0.490	0.462	0.504	0.482		0.484	0.477	0.514	0.472	0.492
San Cristobal	0.743	0.392	0.217	0.158	0.051	0.484		0.019	0.493	0.203	0.398
Santa Cruz	0.739	0.389	0.214	0.169	0.052	0.477	0.019		0.490	0.205	0.398
South Sandwich	0.768	0.515	0.483	0.528	0.506	0.514	0.493	0.490		0.484	0.527
Tonga	0.752	0.364	0.040	0.257	0.224	0.472	0.203	0.205	0.484		0.373
Wallaby-Zenith	0.757	0.059	0.371	0.425	0.403	0.492	0.398	0.398	0.527	0.373	

Table S5.

F_{ST} pairwise calculated with the Hudson's estimator among the 11 hadal feature populations of *Bathycallisoma schellenbergi* populations based on 2,933 SNPs.

		Migrants TO		
		Santa Cruz	**San Cristobal**	**New Hebrides**
Migrants FROM	**Santa Cruz**	0.795 +/- 0.085	0.179 +/- 0.087	0.026 +/- 0.047
	San Cristobal	0.917 +/- 0.062	0.917 +/- 0.062	0.021 +/- 0.070
	New Hebrides	0.033 +/- 0.059	0.176 +/- 0.119	0.710 +/- 0.117
		Java	**Wallaby-Zenith**	
Migrants FROM	**Java**	0.970 +/- 0.054	0.037 +/- 0.054	
	Wallaby-Zenith	0.030 +/- 0.065	0.963 +/- 0.065	

Table S6.

Summary of BayesAss v3.04 estimated migration rates among the three STRUCTURE identified genetic clusters of *Bathycallisoma schellenbergi*. 95% confidence intervals are constructed as the mean +/- 1.96 x standard deviation.

					Bathycallisoma schellenbergi by hadal feature												
		Hirondellea dubia	*Hirondellea gigas*	*Paracallisoma* sp.	Atacama	Java	Kermadec	Kuril-Kamchatka	Massau	Mariana	New Hebrides	Puerto Rico	San Cristobal	Santa Cruz	South Sandwich	Tonga	Wallaby-Zenith
	Hirondellea dubia		5.08	5.21	4.82	4.80	4.87	4.87	4.84	4.90	5.05	4.72	4.94	4.95	4.93	4.85	4.90
	Hirondellea gigas	15.60		5.08	6.59	6.17	6.17	6.21	6.17	6.20	6.22	6.15	6.16	6.17	6.30	6.13	6.28
	Paracallisoma sp.	15.99	15.60		5.11	5.22	5.18	5.16	5.16	5.31	5.24	5.27	5.24	5.24	5.23	5.16	5.30
Bathycallisoma schellenbergi by hadal feature	Atacama	14.82	20.24	15.69		1.06	0.98	1.14	1.12	1.20	1.26	1.06	1.15	1.17	0.91	1.03	1.17
	Java	14.74	18.96	16.03	3.26		0.36	0.36	0.34	0.35	0.50	0.12	0.39	0.39	0.46	0.34	0.15
	Kermadec	14.96	18.94	15.91	3.02	1.09		0.27	0.25	0.34	0.41	0.39	0.31	0.29	0.47	0.11	0.44
	Kuril-Kamchatka	14.96	19.08	15.86	3.49	1.09	0.83		0.07	0.16	0.24	0.27	0.12	0.14	0.44	0.24	0.47
	Massau	14.87	18.95	15.86	3.44	1.05	0.78	0.21		0.15	0.22	0.26	0.11	0.13	0.43	0.23	0.46
	Mariana	15.06	19.04	16.30	3.69	1.07	1.05	0.49	0.45		0.32	0.31	0.22	0.23	0.54	0.34	0.45
	New Hebrides	15.51	19.11	16.09	3.87	1.52	1.25	0.72	0.68	0.98		0.44	0.20	0.22	0.58	0.41	0.57
	Puerto Rico	14.51	18.89	16.20	3.25	0.37	1.21	0.84	0.80	0.94	1.36		0.34	0.35	0.49	0.37	0.23
	San Cristobal	15.18	18.93	16.09	3.55	1.20	0.94	0.37	0.33	0.67	0.62	1.05		0.13	0.49	0.29	0.50
	Santa Cruz	15.19	18.95	16.11	3.58	1.20	0.90	0.43	0.39	0.70	0.67	1.07	0.39		0.49	0.29	0.49
	South Sandwich	15.14	19.34	16.06	2.79	1.42	1.45	1.36	1.32	1.65	1.77	1.49	1.51	1.51		0.45	0.54
	Tonga	14.91	18.82	15.85	3.18	1.06	0.34	0.74	0.70	1.04	1.25	1.15	0.89	0.90	1.39		0.46
	Wallaby-Zenith	15.06	19.28	16.28	3.61	0.45	1.37	1.46	1.41	1.40	1.74	0.72	1.53	1.51	1.66	1.41	

Table S7. Preliminary estimates for divergence times based on mean COI pairwise percent differences. Below diagonal are divergence time estimates in millions of years based on the most adopted divergence rate of 1.4%/MY (*46*). Above diagonal are divergence time estimates in millions of years based on using a recalibrated divergence rate for Arctic Malacostraca of 4.5%/MY (*96*).

Species	Location	16S	COI	Reference
Hirondellea dubia	Puerto Rico Trench	OK641755	OK639161	(*48*)
Hirondellea gigas	Mariana Trench	OK641760	OK639181	(*48*)
Paracallisoma sp.		KX034319	KX365241	(*97*)
Bathycallisoma schellenbergi	Kuril-Kamchatka Trench	MN228750	MN346352	(*15*)
	Massau Trench	MF598978	MF598998	(*98*)
	Massau Trench	MF598979	MF598997	(*98*)
	Massau Trench	MF598980	MF598996	(*98*)
	Tonga Trench	AY256969	AY256968	(*99*)
	Kermadec Trench	KP308148	KP713939	(*14*)
	New Hebrides Trench	KP456060	KP713938	(*14*)
	New Hebrides Trench	KP456061	KP713937	(*14*)
	Wallaby-Zenith Fracture Zone	MN251314	MN262165	(*23*)

Table S8.

Genetic sequences that were publicly available in GenBank and were in the mtDNA phylogenetic analysis. The first three species represent outgroups.

------- ipyrad params file (v.0.9.68)----------	
ct87_13_final	## [0] [assembly_name]: Assembly name. Used to name output directories for assembly steps
~/ipyrad/bathy-all/analysis-params_single-def	## [1] [project_dir]: Project dir (made in curdir if not present)
	## [2] [raw_fastq_path]: Location of raw non-demultiplexed fastq files
	## [3] [barcodes_path]: Location of barcodes file
~/ipyrad/bathy-all/data_demulti_post-drop/*1_.fq.gz	## [4] [sorted_fastq_path]: Location of demultiplexed/sorted fastq files
denovo	## [5] [assembly_method]: Assembly method (denovo, reference)
	## [6] [reference_sequence]: Location of reference sequence file
rad	## [7] [datatype]: Datatype (see docs): rad, gbs, ddrad, etc.
TGCAGG,	## [8] [restriction_overhang]: Restriction overhang (cut1,) or (cut1, cut2)
5	## [9] [max_low_qual_bases]: Max low quality base calls (Q<20) in a read
33	## [10] [phred_Qscore_offset]: phred Q score offset (33 is default and very standard)
6	## [11] [mindepth_statistical]: Min depth for statistical base calling
6	## [12] [mindepth_majrule]: Min depth for majority-rule base calling
100	## [13] [maxdepth]: Max cluster depth within samples
0.87	## [14] [clust_threshold]: Clustering threshold for de novo assembly
0	## [15] [max_barcode_mismatch]: Max number of allowable mismatches in barcodes
2	## [16] [filter_adapters]: Filter for adapters/primers (1 or 2=stricter)
35	## [17] [filter_min_trim_len]: Min length of reads after adapter trim
2	## [18] [max_alleles_consens]: Max alleles per site in consensus sequences
0.05	## [19] [max_Ns_consens]: Max N's (uncalled bases) in consensus
0.05	## [20] [max_Hs_consens]: Max Hs (heterozygotes) in consensus
68	## [21] [min_samples_locus]: Min # samples per locus for output
0.2	## [22] [max_SNPs_locus]: Max # SNPs per locus
8	## [23] [max_Indels_locus]: Max # of indels per locus
0.5	## [24] [max_shared_Hs_locus]: Max # heterozygous sites per locus
0, 0, 0, 0	## [25] [trim_reads]: Trim raw read edges (R1>, <R1, R2>, <R2) (see docs)
0, 0, 0, 0	## [26] [trim_loci]: Trim locus edges (see docs) (R1>, <R1, R2>, <R2)
*	## [27] [output_formats]: Output formats (see docs)
~/ipyrad/bathy-all/pop_postdrop	## [28] [pop_assign_file]: Path to population assignment file
	## [29] [reference_as_filter]: Reads mapped to this reference are removed in step 3

Table S9.

Params file for full ipyrad (v.0.9.68) analysis with 97 *Bathycallisoma schellenbergi* individuals.

Data S1. (separate file)

Summary of specimens used for the mtDNA and RAD-seq data analysis. For each amphipod, the collection information, GenBank and NCBI data accession numbers, and the RAD-seq outputs from different steps along the SNP discovery processing pipeline are provided.