SUPPLEMENTARY MATERIALS

High differentiation but low divergence: demographic history of grey reef sharks (*Carcharhinus amblyrhynchos*) across the Indian Ocean

Carolin Dahms[1*], Paolo Momigliano[1]

[1]School of Biological Sciences, the University of Hong Kong, Hong Kong SAR, China

*carolin.dahms.ac@gmail.com

Methods for Demographic modelling - model selection and parameter estimation

For each of the tested coalescent models including: three population models ('Stepping Stone' and 'Island') and four types of gene flow models - the strict isolation model (SI), the isolation with migration model (IM), the secondary contact model (SC) and the ancient migration model (AM)- and their modifications (ancestral expansions (AE), changes in effective population sizes (NeC) and heterogeneous migration rates across the genome (2M)), model parameters were optimised in 10 independent optimization routines, following methods as per Momigliano et al. (2021). Convergence was checked using the best replicate of each of the 10 independent optimizations. We selected two basic scenarios with the highest likelihood (SC and IM) for each population pair and tested among competing models with changes in migrations and population size using Likelihood Ratio Tests (LRT) and adjusting the D-statistics to account for possible effects of linkage (Coffman et al., 2016). The correct weighting of χ^2 distributions for the LRT were calculated according to Ota et al. (2000). Models with ancestral expansions were excluded as from the best model replicates, the parameter estimates for time of ancestral population expansion were hitting the lower bound (1×10^{-3}), suggesting convergence into the basic demographic scenario.

To estimate parameter uncertainties, from 100 bootstrapped SFS, generated in ANGSD, the best performing demographic model was estimated. The resulting parameters in coalescent units were scaled based on estimates of the ancestral effective population size (N_{ANC}) according to the manual of the *dadi* software (Gutenkunst et al., 2009). N_{ANC} was calculated as $\Theta/(4L\mu)$, where L is the total sequence length from which SNPs used in demographic analyses originated and $\mu = 1.93\times10^{-8}$. The total sequence length was calculated as the number of sites (variant and invariant) retained in ANGSD to calculate the two-dimensional SFS. Time parameters were scaled assuming a generation time of 16.4 years (Robbins, 2006). The absolute number of migrants (M_{12}, M_{21}) were calculated from scaled migration rates and population sizes from the relevant time period.

SUPPLEMENTARY TABLES

Supplementary Table S1: Divergence and diversity estimates for the Chagos and Juan dataset.

Central Indian Ocean comparison					
Population	***p***	**Tajima's D**	F_{ST}	d_{xy}	***dA***
Chagos	$0.0006 \pm 9.8\times10^{-6}$	-0.47 ±0.04	-	-	-
Indo	$0.0012 \pm 1.8\times10^{-5}$	-0.92 ±0.02	0.34 ±0.004	$0.0013 \pm 1.7\times10^{-5}$	$3.9\times10^{-4} \pm 1.2\times10^{-5}$

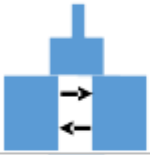
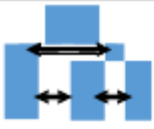
Ningaloo	0.0011 ±1.8×10^{-5}	-0.91 ±0.02	0.37 ±0.009	0.0013 ±2.6×10^{-5}	4.4×10^{-4} ±2.9×10^{-5}
South GBR	0.001 ±1.8×10^{-5}	-0.76 ±0.03	0.38 ±0.006	0.0013 ±1.8×10^{-5}	4.4×10^{-4} ±2.3×10^{-5}

West Indian Ocean comparison					
Population	***p***	**Tajima's D**	F_{ST}	d_{xy}	***dA***
Juan	0.0008 ±1.3×10^{-5}	-0.31 ±0.01	-	-	-
Bampton	0.0015 ±1.7×10^{-5}	-0.60 ±0.007	0.42 ±0.003	0.002 ±1.9×10^{-5}	7.1×10^{-4} ±2.4×10^{-5}
Orona	0.0016 ±2.2×10^{-5}	-0.59 ±0.01	0.43 ±0.003	0.002 ±1.9×10^{-5}	7.7×10^{-4} ±1.9×10^{-5}

Supplementary Table S2: Parameter estimates from the best-performing demographic models. *Coalescent demographic models were run in moments and selected according to likelihood ratio tests with the Godambe information matrix.*

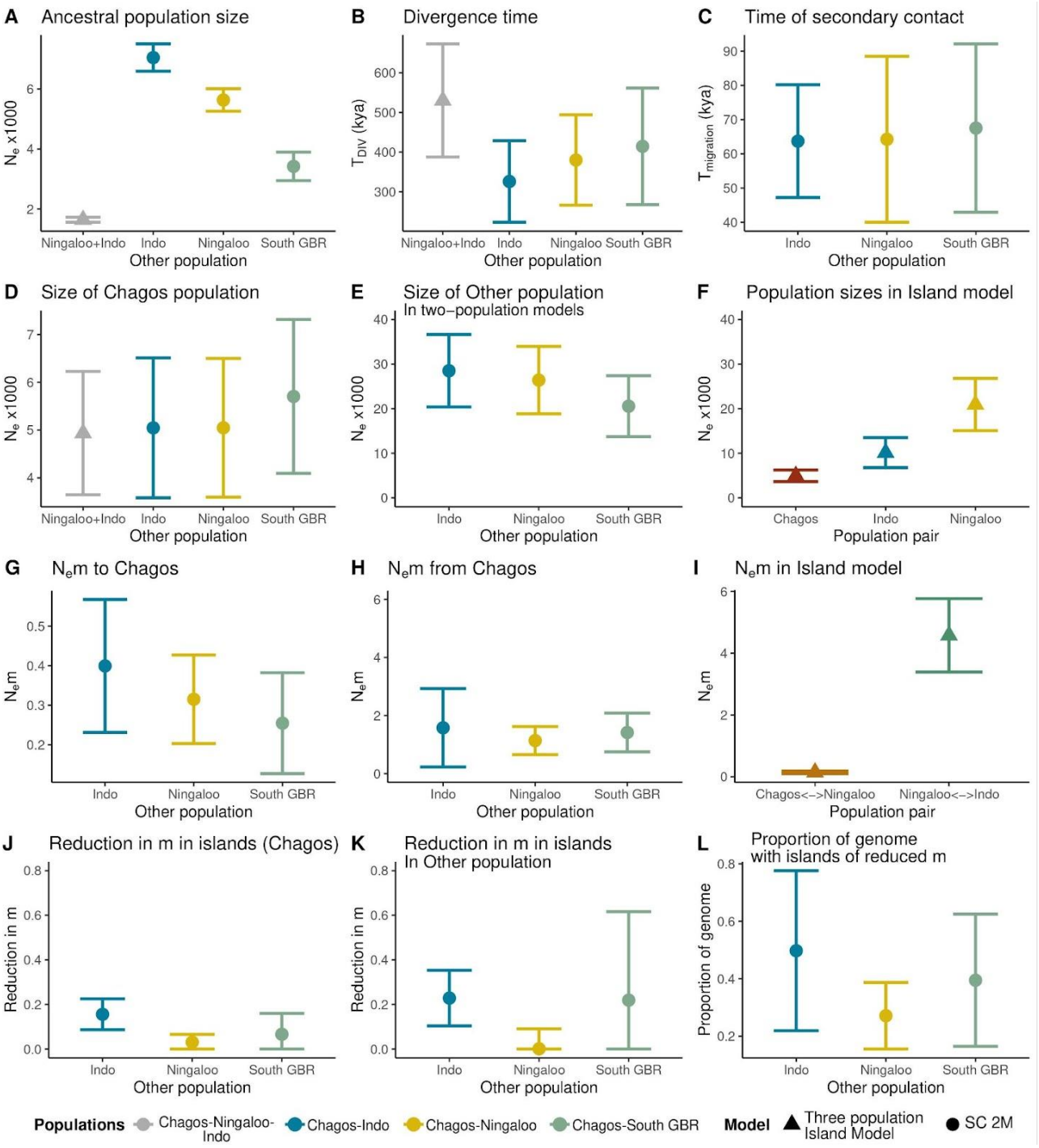
Central Indian Ocean comparison													
Populations	**Model**	**T_{DIV}**	**T_{SC}**	**N_{ANC}**	**nu_1 (outgroup)**	**nu2**	**nu3**	**NeM1 (to outgroup)**	**NeM2 (from outgroup)**	**NeMA (outgroup)**	**i1**	**i2**	**P**
Chagos-Indo	SC 2M	389,694 ±55,815	63712 ±8,407	7,050 ±232	5,046 ±747	28,520 ±4,149	-	0.399 ±0.08	1.58 ±0.69	-	0.16 ±0.03	0.23 ±0.06	0.5 ±0.14
Chagos-Ningaloo	SC 2M	444,388 ±63,097	64,260 ±12,363	5,636 ±191	5,048 ±741	26,408 ±3,858	-	0.315 ±0.06	1.14 ±0.25	-	0.03 ±0.02	0.001 ±0.04	0.27 ±0.06
Chagos-South GBR	SC 2M	482085 ±80,614	67,524 ±12,543	3,422 ±242	5,704 ±822	20,562 ±3,488	-	0.255 ±0.06	1.42 ±0.34	-	0.07 ±0.05	0.22 ±0.2	0.39 ±0.12
Chagos-Ningaloo-Indo	SymSplit	530,201 ±72,786	-	1,642 ±42	4,936 ±659	10,124 ±1,721	-	0.145 ±0.02	4.58 ±0.61*		-	-	-
Western Indian Ocean comparison													
Populations	**Model**	**T_{DIV}**	**T_{SC}**	**N_{ANC}**	**nu_1 (Juan)**	**nu2 (Bampton)**	**nu3 (Orona)**	**NeM1 (Juan-Bampton)**	**NeM2 (Bampton-Orona)**	**NeMA (Juan-Anc)**	**i1**	**i2**	**P**
Juan-Bampton	SC 2M	833,101 ±136,300	62,556 ±11,147	2,375 ±850	5,960 ±997	29,607 ±4,648	-	0.28 ±0.05	1.64 ±0.29	-	0.16 ±0.03	0.23 ±0.06	0.5 ±0.14
Juan-Orona	SC 2M	874,748 ±118,901	63,868 ±10,003	2,877 ±1,312	6,218 ±948	29,360 ±4,539	-	0.26 ±0.04	1.77 ±0.31	-	0.03 ±0.02	0.001 ±0.04	0.27 ±0.06
Juan-Bampton-Orona	TwoSplits	604,992 ±86,976	-	5,413 ±1,057	5,913 ±762	11,761 ±1,648	18,391 ±2,725	0.22 ±0.0*	6.73 ±0.0*	<0.0001	-	-	-

SUPPLEMENTARY FIGURES

Graphic representation	Demographic scenarios	Number of populations	Modifiers	Number of free parameters
	SI Strict Isolation. No migration	2	None	3
	AM Ancient migration	2	None	6
	IM Isolation with Migration. One change in N_e in daughter populations	2	2M	Basic = 5 2M = 8
	IM_{AE} Isolation with Migration. One change in N_e in the ancestral population	2	2M	Basic= 7 2M = 10
	IM_{NeC} Isolation with Migration.	2	2M	Basic = 6 2M = 9
	SC Secondary Contact	2	2M	Basic = 6 2M = 9
	SC_{AE} Secondary Contact. One change in N_e in the ancestral population	2	2M	Basic = 8 2M = 11
	SC_{NeC} Secondary Contact. One change in N_e in daughter populations	2	2M	Basic = 7 2M = 12
	SymSplit Simultaneous split, symmetric migration between adjacent populations	3	None	3
	TwoSplits Two splits, symmetric migration between adjacent populations	3	None	3

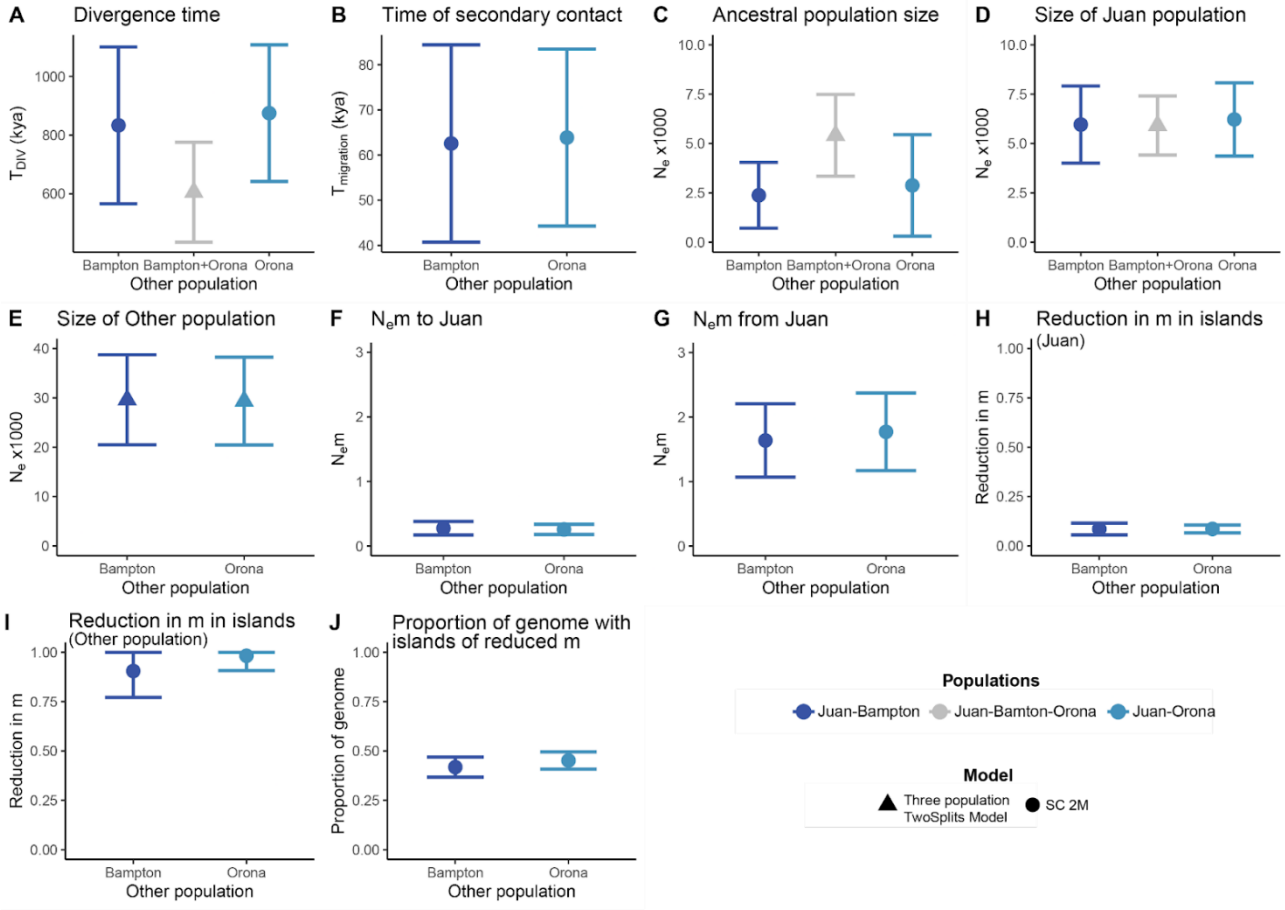
Supplementary Figure S1: Description of demographic model parameters. *Two and three population models run in moments software to infer the demographic history. For the two best performing scenarios (IM and SC) according to the Akaike Information Criterion, we also tested all model combinations with heterogeneous migration rates (2M) and instantaneous changes in effective population sizes in ancestral (AE) and/or daughter populations (NeC). All models assume an ancestral population splitting into two daughter populations at T1. The strict isolation (SI), and isolation with migration (IM) models have only time parameter, with IM allowing for migration*

throughout (black arrows); while in the SC (secondary contact) model migration begins at time of secondary contact (T2) and vice versa for the AM (ancient migration) model.

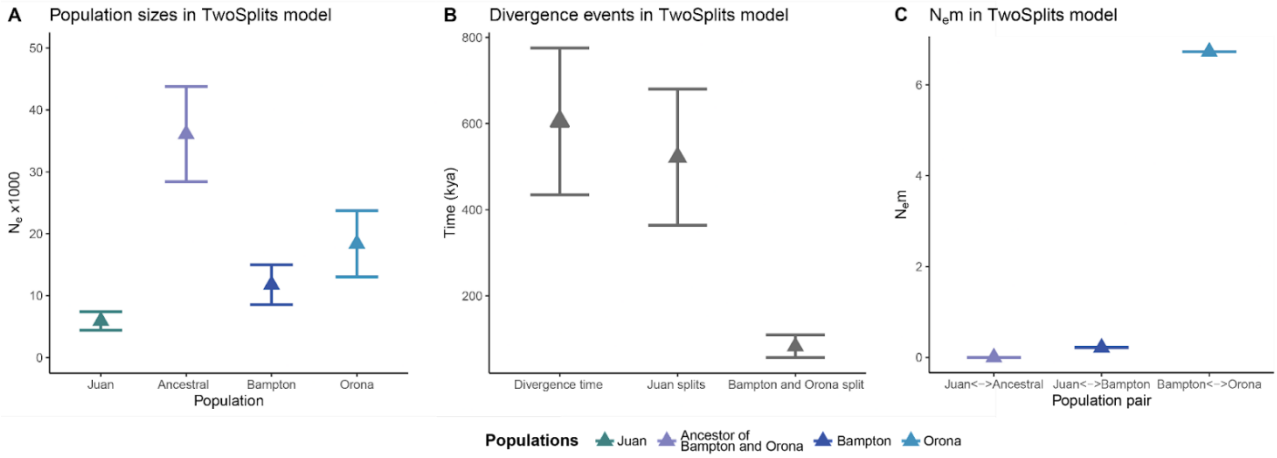


Supplementary Figure S2: Parameter estimates for two and three population models for the central Indian Ocean comparison. *Best performing two-population scenarios (circles) between Chagos and three populations from eastern Indian Ocean (Ningaloo), and the Pacific (Indo and South Great Barrier Reef) from demographic histories inferred by moments all suggested recent secondary contact (SC) with heterogeneous migration rates (2M). Parameters from the best performing three population scenario with simultaneous split (triangles) are shown to compare*

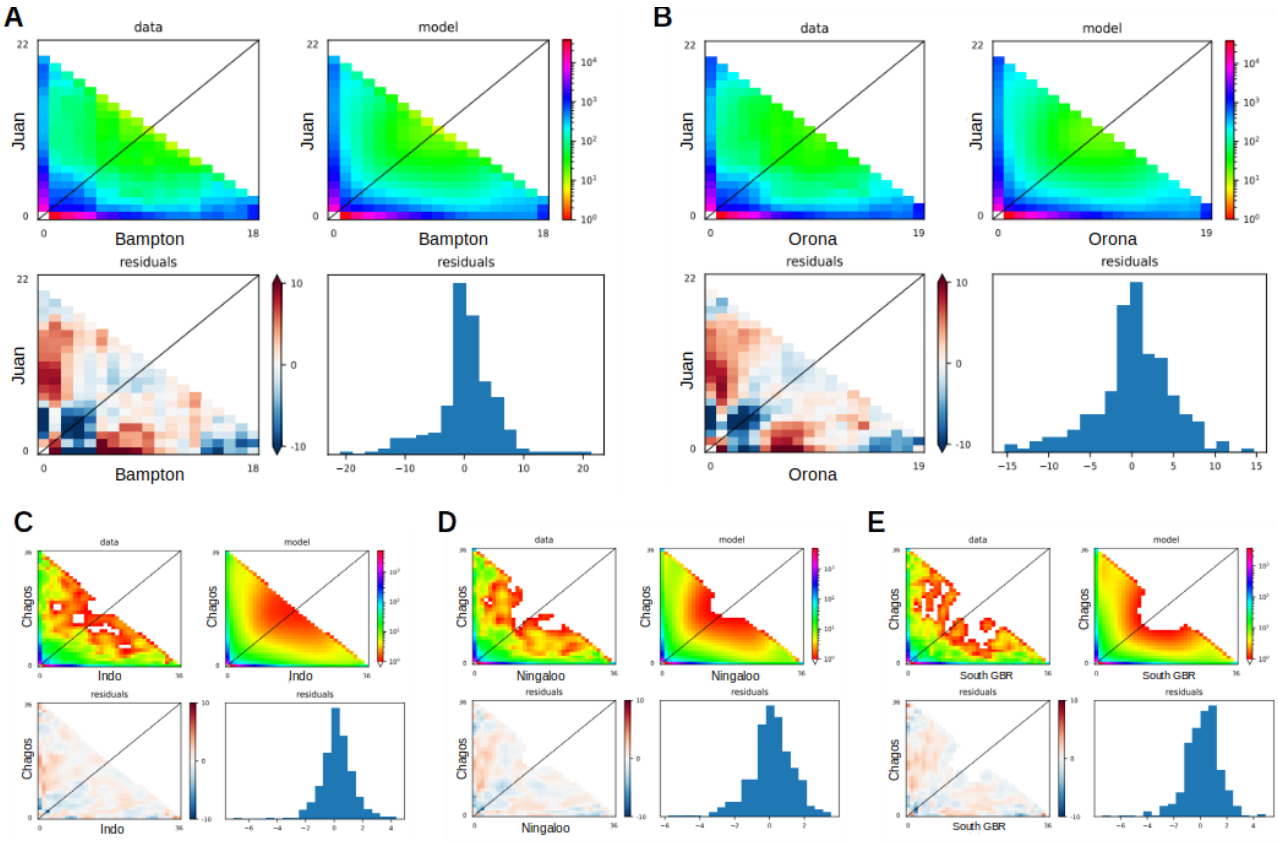
estimates. In the three-population model, migration rates are symmetrical between adjacent populations, while N_em in two population rates are asymmetrical and represent the number of migrants per generation making up the receiving population. Mean demographic history parameters were estimated from 100 model inferences, and parameters scaled with a mutation rate of $\mu = 1.93\times10^{-8}$ and a generation time of 16.4 years.



Supplementary Figure S3: Parameter estimates for the western Indian Ocean comparison. *Best performing two-population scenarios (circles) between Juan and three populations from the western Pacific (Bampton and Orona) inferred by moments all suggested recent secondary contact (SC) with heterogeneous migration rates (2M). Parameters from the best performing three population TwoSplits scenario (triangles), where Juan splits prior to the split between the Pacific populations, are shown to compare estimates. In the three-population model, migration rates are symmetrical between adjacent populations, while N_em in two population rates are asymmetrical and represent the number of migrants per generation making up the receiving population. Mean demographic history parameters were estimated from 100 model inferences, and parameters scaled with a mutation rate of $\mu = 1.93\times10^{-8}$ and a generation time of 16.4 years.*



Supplementary Figure S4: Parameter estimates from the TwoSplits scenario for the western Indian Ocean comparison. *The best performing three-population scenarios between Juan and three populations from the western Pacific (Bampton and Orona) inferred by moments suggested the western Indian Ocean population (Juan) splitting prior to the split between the Pacific populations. Migration rates are symmetrical between adjacent populations, and represent the number of migrants per generation making up the receiving population.*



Supplementary Figure S5: Model fit and SFS. *Inferred and observed Site Frequency Spectra (SFS) and residuals of best fitting two dimensional demographic models for population pairs with the western Indian Ocean population Juan (A,B), and central Indian Ocean population Chagos (C-E) obtained with moments v.1.1.0 (Jouganous et al., 2017).*

References

Coffman, A. J., Hsieh, P. H., Gravel, S., & Gutenkunst, R. N. (2016). Computationally efficient composite likelihood statistics for demographic inference. *Molecular Biology and Evolution*, *33*(2), 591–593.

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, *5*(10), e1000695.

Jouganous, J., Long, W., Ragsdale, A. P., & Gravel, S. (2017). Inferring the joint demographic history of multiple populations: Beyond the diffusion approximation. *Genetics*, *206*(3), 1549–1567.

Momigliano, P., Florin, A.-B., & Merilä, J. (2021). Biases in demographic modeling affect our understanding of recent divergence. *Molecular Biology and Evolution*, *38*(7), 2967–2985.

Robbins, W. D. (2006). *Structure of the grey reef shark (Carcharhinus amblyrhynchos) and the white tip reef shark (Triaenodon obesus)(Fam. Charcharhinidae). PhD thesis, James Cook University.*