

## ORGANISMAL BIOLOGY

# The sardine run in southeastern Africa is a mass migration into an ecological trap

Peter R. Teske<sup>1\*†</sup>, Arsalan Emami-Khoyi<sup>1†</sup>, Tirupathi R. Golla<sup>1†</sup>, Jonathan Sandoval-Castillo<sup>2†</sup>, Tarron Lamont<sup>3,4†</sup>, Brent Chiazzari<sup>5</sup>, Christopher D. McQuaid<sup>6</sup>, Luciano B. Beheregaray<sup>2</sup>, Carl D. van der Lingen<sup>7,8†</sup>

The KwaZulu-Natal sardine run, popularly known as the “greatest shoal on Earth,” is a mass migration of South African sardines from their temperate core range into the subtropical Indian Ocean. It has been suggested that this represents the spawning migration of a distinct subtropical stock. Using genomic and transcriptomic data from sardines collected around the South African coast, we identified two stocks, one cool temperate (Atlantic) and the other warm temperate (Indian Ocean). Unexpectedly, we found that sardines participating in the sardine run are primarily of Atlantic origin and thus prefer colder water. These sardines separate from the warm-temperate stock and move into temporarily favorable Indian Ocean habitat during brief cold-water upwelling periods. Once the upwelling ends, they find themselves trapped in physiologically challenging subtropical habitat and subject to intense predation pressure. This makes the sardine run a rare example of a mass migration that has no apparent fitness benefits.

Copyright © 2021  
The Authors, some  
rights reserved;  
exclusive licensee  
American Association  
for the Advancement  
of Science. No claim to  
original U.S. Government  
Works. Distributed  
under a Creative  
Commons Attribution  
NonCommercial  
License 4.0 (CC BY-NC).

## INTRODUCTION

Large-scale annual migrations occur in an extraordinary range of animals, from insects to the great whales. While the driving mechanisms of these migrations are varied and sometimes poorly understood, they often represent a way of optimizing conditions for breeding and adult fitness when these are in conflict. Often, populations may follow the availability of suitable food as it shifts in a geographically predictable fashion, as in the case of the mass migrations of the Serengeti (1). However, migrations may also allow adults to take advantage of high food availability in places, or at times, that are not suitable for breeding or for earlier ontogenetic stages, a strategy pursued by several Antarctic cetaceans (2). Species that depend on migrations for their survival are especially vulnerable because anything that disrupts this behavior threatens a huge proportion, or indeed all, of their populations. Understanding the drivers of migrations is key to predicting how or whether such behavior may alter with long-term changes in environmental conditions, such as those expected under anthropogenic climate warming.

One of the best documented annual marine mass migrations is the sardine run off the east coast of South Africa (3). While the so-called “greatest shoal on Earth” involves several small pelagic fish species and attracts huge numbers of opportunistic predators, including seabirds, gamefish, sharks, and marine mammals (4, 5), it primarily represents the migration of the southern African population of Pacific sardine, *Sardinops sagax* (Jenyns, 1842). This species is southern Africa’s commercially most important small pelagic fish.

Its range in this region includes a distinct Namibian stock (6, 7) and a South African population with two major high-density areas separated by the boundary between the Atlantic and Indian oceans, approximately at Cape Agulhas (Fig. 1) (8, 9). Sardines from these two South African regions spawn at different temperatures (10) and have different nursery areas (11, 12). These and other differences (13) suggest that they comprise distinct western and southern stocks (8, 14–18), but genetic studies have so far failed to confirm this (19).

The sardine run is a spawning migration of sardines from the eastern portion of their temperate South African core range to the subtropical east coast (Fig. 1) (3). It involves the movement of tens to hundreds of millions of sardines migrating to the northeast in high-density shoals by remaining associated with pockets of cool water that are sporadically uplifted onto the shallow continental shelf inshore of the warm, southward-flowing Agulhas Current (20, 21). During the migration, water temperatures may exceed the species’ preferred range (22), leading to thermal stress and reduced condition (20, 23). It has been suggested that the sardine run represents relic spawning behavior dating back to the previous glacial period, when what is now tropical Indian Ocean habitat may have been an important sardine nursery area (24). However, genetic data have so far failed to support the distinctness of sardines participating in the run from those in the remainder of the South African range (19, 25).

Here, we use genomic and transcriptomic data to test the hypothesis that the sardine run represents the spawning migration of a distinct stock (24). Such datasets have considerable potential to identify subtle population structure in high-dispersal species where traditional genetic methods suggest spatial homogeneity (26–29). We also tested the hypothesis of distinct regional sardine stocks in the species’ core range, which is important for management of the sardine fishery, and then used this information to investigate the origins of fish that participate in the sardine run.

## RESULTS

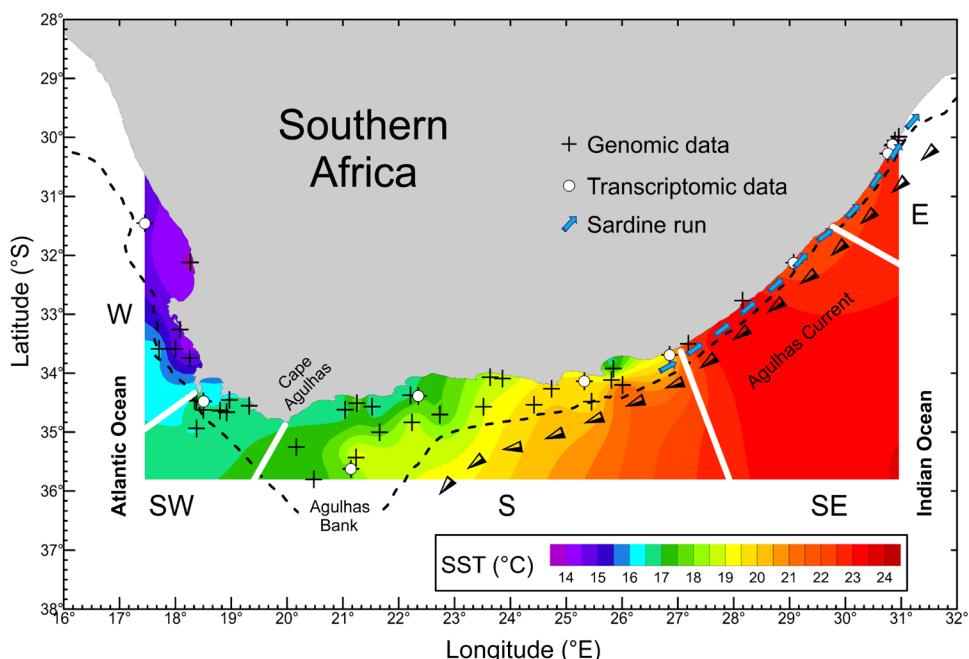
Genomic and transcriptomic data were generated from sardines captured throughout the species’ South African range. The SNPs (single-nucleotide polymorphisms) from each dataset were assigned either to a dataset of candidate SNPs potentially under selection or

<sup>1</sup>Centre for Ecological Genomics and Wildlife Conservation, Department of Zoology, University of Johannesburg, Auckland Park 2006, South Africa. <sup>2</sup>Molecular Ecology Lab, College of Science and Engineering, Flinders University, Adelaide, SA 5042, Australia. <sup>3</sup>Oceans and Coasts, Department of Forestry, Fisheries and the Environment, P.O. Box 52126, Victoria and Alfred Waterfront, Cape Town 8002, South Africa.

<sup>4</sup>Department of Oceanography, University of Cape Town, Private Bag X3, Rondebosch 7700, South Africa. <sup>5</sup>Oceanographic Research Institute, P.O. Box 10712, Marine Parade, Durban 4056, South Africa. <sup>6</sup>Department of Zoology and Entomology, Rhodes University, Grahamstown 6140, South Africa. <sup>7</sup>Fisheries Management, Department of Forestry, Fisheries and the Environment, Private Bag X2, Vlaeberg 8012, South Africa. <sup>8</sup>Department of Biological Sciences and Marine Research Institute, University of Cape Town, Private Bag X3, Rondebosch 7700, South Africa.

\*Corresponding author. Email: pteske101@gmail.com

†These authors contributed equally to this work.



**Fig. 1. A map of the South African range of the southern African population of Pacific sardine, *Sardinops sagax*.** The map shows sites at which sardines were caught for genome and transcriptome sequencing. Colors represent mean sea surface temperatures (SSTs). The coastline was divided into five temperature-defined geographical regions (temperate core range: W, west; SW, southwest; S, south; SE, southeast; sardine run: E, east). Cape Agulhas is the boundary between the Atlantic and Indian oceans. The broken line represents the edge of the continental shelf, beyond which the sardines rarely disperse. The black and white arrows represent the approximate path of the Agulhas Current, which transports tropical Indian Ocean water southward and confines sardines participating in the run (blue arrows) to a narrow coastal band of cooler water.

a dataset of selectively neutral SNPs. Candidate SNPs were identified using consensus between  $F_{ST}$ -based genome scans and tests of genotype-environment associations (GEAs) using water temperature as the explanatory variable. Datasets of putatively selectively neutral SNPs were generated by removing a larger dataset of candidate SNPs identified using more relaxed settings from the complete datasets. Various methods to detect population structure were explored, but only those showing clear results are presented here in detail.

### Selectively neutral SNPs

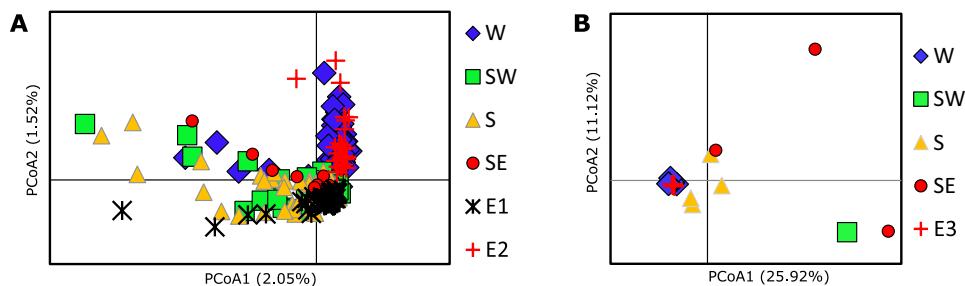
Methods that assign individuals to population clusters found no structure using the selectively neutral SNPs. The fastSTRUCTURE chooseK algorithm identified  $K = 1$  (i.e., a single cluster) for both the genomic and transcriptomic SNPs, and bar plots enforcing  $K = 2$  (two clusters) showed no clear trends (fig. S1). Cross-validation in ADMIXTURE using  $K = 1$  to 5 found the lowest cross-validation error for  $K = 1$  (table S1), and plots depicting  $K = 2$  also showed no trends (fig. S2). Although principal coordinates analyses (PCoAs) of the selectively neutral genomic data also failed to recover any distinct clusters, the ordination plot (Fig. 2A) showed interesting spatial trends. Sardines from the west coast were recovered as a largely distinct cluster that partially overlapped with other regions in the species' core range (southwest coast, south coast, and southeast coast; Fig. 2A). In contrast, the plot for the larger transcriptomic dataset produced a single cluster of west coast individuals (Fig. 2B). Fish from the 2015 sardine run (genomic data; Fig. 2A, E1) showed no clear affiliation with the individuals from any of four regions in the core range, whereas 2018 (genomic data; Fig. 2B, E2) and 2019 (transcriptomic data; Fig. 2B, E3) sardines clustered mostly among west coast sardines.

### Candidate SNPs

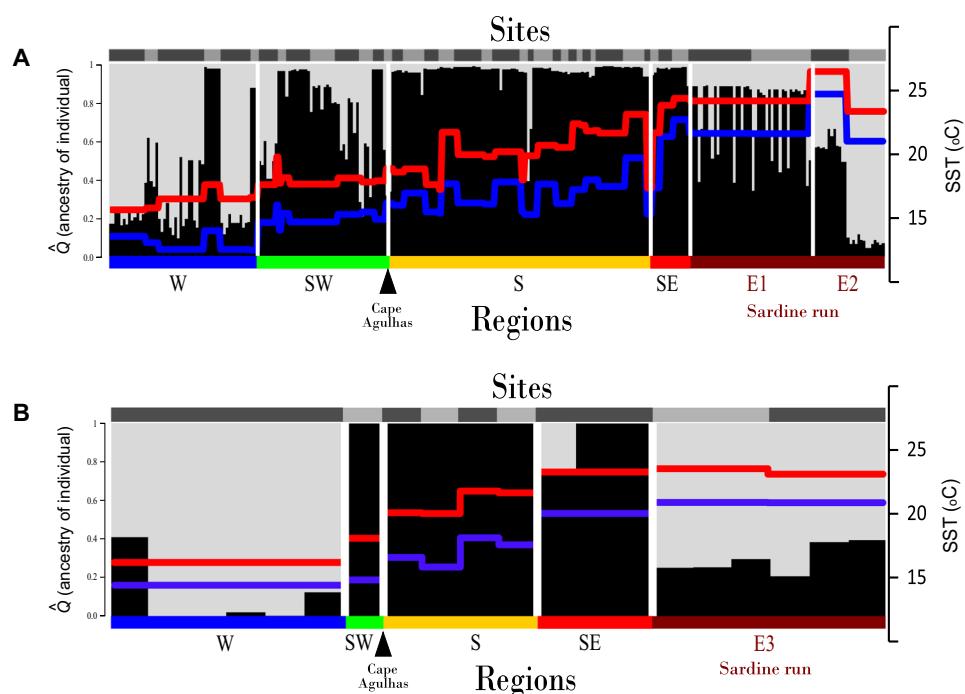
A STRUCTURE bar plot constructed using genomic candidate SNPs (Fig. 3A) provided more detailed information concerning the likely origin of sardine run participants based on each individual's proportional assignment to two population clusters (fig. S4), which are represented by gray and black colors. Sardines that were mostly assigned to the first cluster (gray) were particularly common in the cool-temperate Atlantic Ocean, while those mostly assigned to the second cluster (black) were common in the Indian Ocean. The southwest coast had a mixture of both clusters, with numerous individuals having intermediate ancestry.

Sardines collected during the sardine run (E1 and E2) were of heterogeneous ancestry. The estimated ancestry proportion from the Indian Ocean cluster (black) was high for many individuals, but assignment probabilities were  $<0.8$ . This was rare on the south coast and more common in the cooler Atlantic Ocean waters west of Cape Agulhas (west and southwest). Evidence for the Atlantic ancestry of sardine run participants was particularly strong for the second site of E2 and for the sardine run participants for which transcriptomic data were generated (fig. S1C, E3). The fact that very few sardines from the warm-temperate south coast exhibited any appreciable proportion of their genomic ancestry from the cool-temperate Atlantic Ocean cluster (gray) illustrates that, although sardines in this area could potentially participate in the run, which commences on the eastern Agulhas Bank (Fig. 1), the number that do so is comparatively small.

While a PCoA ordination plot of the genomic candidate SNPs showed no clear trends (fig. S5A), the west coast individuals and E3 sardine run participants clustered together in the corresponding



**Fig. 2. PCoA ordination plots of sardine data from five geographical regions along the South African coastline.** Results are based on a standardized covariance matrix constructed from neutral genomic SNP data (A) and neutral transcriptomic SNP data (B). Geographical regions are indicated by W, west; SW, southwest; S, south; SE, southeast; and E, east, with E1, E2, and E3 representing the 2015, 2018, and 2019 sardine runs, respectively. Differentiation between W and E1 in (A) is less clear than is apparent here, as there is considerable overlap with west coast data points that are obscured in this figure (see fig. S3).



**Fig. 3. Population structure in the South African population of Pacific sardine, *S. sagax*.** Assignment of sardines to one of two population clusters (gray or black) using candidate loci from genome data (A) and transcriptome data (B). Each vertical bar represents a single individual. The 5th (blue) and 95th (red) percentiles of SST are superimposed. Gray shading on top of bar plots represents sampling sites, and labels at the bottom represent the five geographical regions shown in Fig. 1, with E1, E2, and E3 representing the 2015, 2018, and 2019 sardine runs, respectively.

plot for the transcriptomic dataset (fig. S5B). In contrast to the STRUCTURE results, ADMIXTURE cross-validation identified  $K = 1$  as the optimal number of clusters for both the genomic and transcriptomic candidate SNPs (table S1). Nonetheless, a  $K = 2$  bar plot for the genomic candidate SNPs (fig. S2B) showed subtle regional differentiation, with many individuals from the westernmost and easternmost sites having Atlantic (gray) ancestry proportions that were rare on the south coast. Moreover, a  $K = 2$  bar plot of the transcriptomic data (Fig. 3B) showed ancestry assignments that were similar to those in Fig. 3A. These included west coast individuals with more Atlantic (gray) cluster ancestry than in the STRUCTURE bar plot (up to 100%) and all south coast sardines with 100% Indian Ocean (black) cluster ancestry.

## DISCUSSION

Large-scale annual migrations are a characteristic of many species and often represent a way of optimizing conditions for breeding and adult fitness when these are in conflict. In terms of spectacle and the sheer magnitude of the event, the sardine run is comparable to East Africa's great wildebeest migration, but while the drivers of the latter are well known, the same cannot be said for the sardine run. Biological explanations for this migration into what is apparently unsuitable habitat include relic spawning behavior (20), equatorward movement of juvenile fish (30), natal homing (20), a feeding migration facilitated by habitat extension (22), and northward herding of sardines by predators (22). Early studies of the physical environment speculated that the sardine run was enabled by the establishment of a

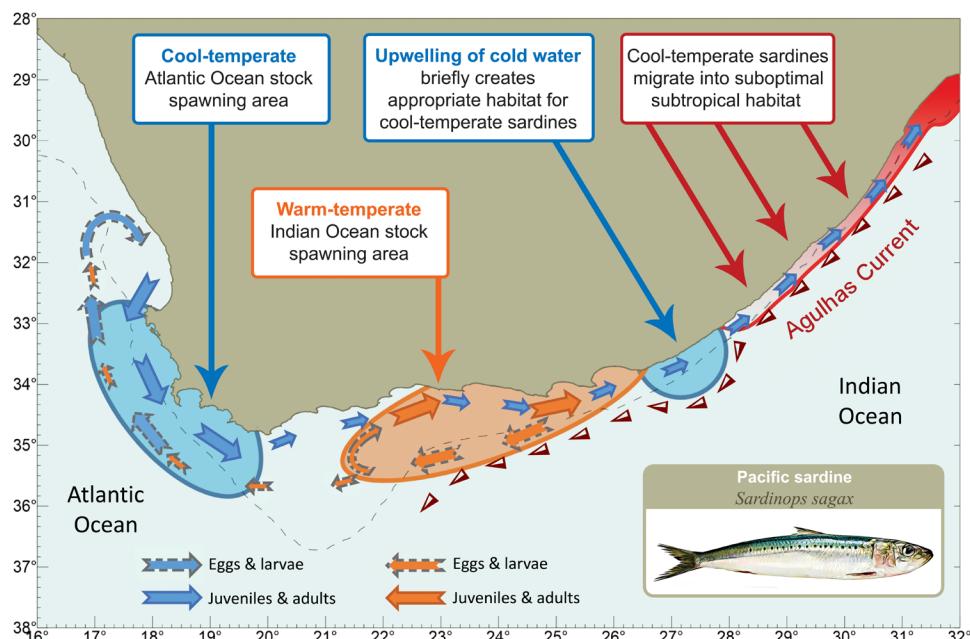
northeastward-flowing countercurrent inshore of the Agulhas Current on the east coast during winter (31) or that it represents a response to the expansion of suitable habitat during cooler winter conditions (22). Neither hypothesis has been strongly supported by subsequent research (21, 32–34).

Our study provides new insights into the causes of the sardine run. First, it provides the first molecular support for a previously proposed two-stock hypothesis based on nongenetic data (8, 10–18). This suggests that the previous evidence for stocks following the definition of Ihssen *et al.* (35) as being “intraspecific groups of randomly mating individuals with temporal and spatial integrity” is not merely a result of phenotypic plasticity in regional assemblages affected by different environmental conditions but that it has a heritable basis. The finding that spatial population structure in candidate SNPs is linked to the region’s temperature-defined biogeographical provinces (36, 37) rather than isolation by geographic distance previously reported for selectively neutral SNPs (38) has important implications for the management of the sardine fishery. Despite admixture between the two stocks that is facilitated by dispersal in both directions, a single-stock strategy can result in population declines, as regional stocks adapted to specific temperature ranges are overexploited. Although it was found that the sardine run comprises a heterogeneous assemblage of fish with ancestry from both the western and southern stocks, ancestry proportions indicated that these migrants originate mostly from the species’ cool-temperate core area in the Atlantic Ocean.

Our results further show that sardines participating in the run are not a distinct east coast population that mingles with the south coast stock during summer and then separates from it in winter. Instead, the primarily Atlantic origin of these sardines supports a

previous hypothesis based on morphological similarities between east coast sardines and those from the western portion of the species’ range (31). A preference of these sardines for colder water not only explains why only a fraction of the sardines present on the south coast participates in the run (23) but also explains why cold water upwelling off the southern east coast is a prerequisite for annual sardine runs to occur.

The sequence of events that results in a sardine run is briefly summarized in light of our evidence (Fig. 4). During winter, shelf waters off the eastern south coast can temporarily become cooler than those off the western south coast (39), providing appropriate conditions that cause migrants of cool-temperate Atlantic ancestry to aggregate on the eastern south coast. Here, cooling of the shelf environment due to the uplift of cold, nutrient-rich waters by cyclonic eddies (21, 34, 40–42), and wind-driven advection and vertical mixing that promote shoreward movement of these cold waters (34, 42), result in conditions of elevated productivity (43) that briefly resemble those found during summer upwelling in the Atlantic Ocean (39). The intermittent nature and random timing of these events provide an explanation for why the sardine run does not occur at the same time or with the same intensity every year. In addition, the cyclonic eddies do not always affect the shelf regions to the same degree, with some eddies resulting in only minimal cooling of the shelf waters (34). Ekman veering that results in the uplifting of cooler water at the shelf edge (34, 42) may facilitate the northward movement of sardines at depths of 100 to 200 m. The introduction of upwelled water, with temperatures that are closer to the western sardines’ preferred thermal tolerance range, may stimulate sardine run participants to separate from the warm-temperate southern stock and follow these transient pockets of cooler water up the east coast.



**Fig. 4. Stock structure of Pacific sardine, *S. sagax*, in South African waters and sequence of events that results in a sardine run.** The spawning area in the Atlantic Ocean (blue) is numerically dominated by cool-temperate sardines (gray in Fig. 3), and the spawning area in the Indian Ocean (orange) is dominated by warm-temperate sardines (black in Fig. 3). There is considerable exchange between these areas, with eggs and larvae from the Indian Ocean stock primarily moving westward and juveniles and adults of both stocks moving eastward. Upwelling on the southeast coast attracts cool-temperate sardines present on the south coast, which follow the cooler water as it is transported northward. When the upwelling ceases, these sardines eventually find themselves in an ecological trap of suboptimal subtropical habitat.

Cessation of upwelling eventually results in these sardines finding themselves in subtropical waters that exceed their preferred thermal range. Thermal stress, coupled with physical exhaustion and generally lower food availability (compared to their typical west coast environment), results in poor body condition (20, 23). Moreover, avoidance of the tropical water of the Agulhas Current concentrates them on the narrow continental shelf off the east coast (Fig. 4) and makes them an easy target for predators.

The presence of sardine eggs in the plankton confirms their survival off the east coast until austral summer (44). At best, this would make the sardine run a migration of west coast sardines into east coast habitat where conditions are temporarily suitable for feeding and spawning during a time when such conditions do not exist on the south coast. At worst, it constitutes a costly navigation error into an ecological trap sensu Robertson and Hutto (45), where sardines show a preference for one habitat over another and fitness will differ between habitats and ultimately be lower in the preferred habitat. Specifically, the migration is driven by initially favorable environmental cues, which lure sardines with a preference for colder temperatures into habitat that eventually becomes unfavorable. Here, their fitness and survival are ultimately compromised because of physiologically challenging high temperatures and intense predation (3–5, 20, 22).

Long-term increases in sea surface temperature (SST) along the South African east coast that are driven by anthropogenic climate warming (46) increasingly make this region less hospitable to sardine [but see (47)]. Under the representative concentration pathway 4.5 scenario and using National Oceanic and Atmospheric Administration's (NOAA's) Coupled Model Intercomparison Project 5 ensemble (48), temperatures off the east coast during April to June (the period during which the sardine run starts) are predicted to increase by almost 1°C by 2055 and to show substantially increased variability (fig. S6). Given the colder water origins of sardines participating in the run, the projected warming could lead to the cessation of the sardine run in the next few decades (49). Despite the huge numbers of fish involved, the sardine run involves only a small portion (<10%) of South Africa's *S. sagax* population (8). Although cessation of the sardine run would mean the loss of one of nature's most spectacular migrations, the effects on the population as a whole are likely to be negligible.

## MATERIALS AND METHODS

### Sampling design

Population structure in the South African population of Pacific sardine, *S. sagax*, was assessed using genomic and transcriptomic data from sardines collected throughout the species' core range and from three sardine runs. Genomic data were generated from tissue samples obtained from 284 sardines collected at 40 locations throughout the species' range (table S2). To confirm the genomic findings independently, transcriptome data were generated using RNA extracted from liver tissue of 20 individuals collected at nine locations (table S2). Sardines from the west coast to the southeast coast were collected as part of pelagic research surveys conducted in autumn/winter (May and June) and spring/summer (October and November) in 2014 and 2015. All of these were used to generate genomic data, with a subset of 14 individuals from seven sites also used to generate transcriptome data. East coast samples for genomic analyses were obtained from artisanal fishers during the 2015 and 2018 sardine runs. As all these samples failed the quality screening for RNA sequencing,

we obtained six additional samples from the 2019 sardine run that were not used for genomic analyses. For each sardine run, samples originated from two different locations that were sampled at different times (table S2).

### Genome assembly

Genomic libraries were prepared on the basis of a double-digest restriction site-associated DNA sequencing approach, ddRADseq (50), using the restriction enzymes Sbf I-HF and Mse I (New England Biolabs), and one of 96 unique 6-base pair (bp) barcodes was ligated to each individual library. We multiplexed 48 or 96 samples per lane, and libraries were randomly assigned to each of five Illumina HiSeq 2500 (Illumina Inc., San Diego, USA) lanes (single end, 100 bp). Raw sequences were demultiplexed using the process\_radtags script from Stacks v2.4 (51); barcodes and RAD tags were then trimmed, and reads with low average base quality ( $Q > 20$ ) were removed using Trimmomatic v0.39 (52). Using dDocent v2.19 (53), the remaining reads were then de novo assembled in a reference catalog. This catalog was used to call SNPs and genotype each sample. The obtained variants were filtered using VCFtools v1.19 (54) following Teske *et al.* (28). To reduce the effect of linkage disequilibrium, we removed one SNP from each pair of SNPs with  $r^2 > 0.8$  (55). Details about the different steps in the pipeline are listed in table S3.

### Transcriptome assembly

The transcriptome data were generated from RNA using sardine livers that were cut into small pieces using a sterile scalpel blade and stored in RNAlater solution (Thermo Fisher Scientific). Although RNA sequence data are often used for gene expression analyses, we considered this approach unsuitable because the sardine livers could not be preserved under controlled conditions. For example, while some sardine livers from the pelagic surveys were preserved immediately, some of the sardine run fish had died at least an hour before preservation. Total RNA was extracted from each liver sample using a combination of mechanical homogenization with TRIzol and QIAGEN RNeasy purification kit (QIAGEN, Hilden, Germany). Then, cDNA libraries were constructed from each extraction, indexed separately, and sequenced on an Illumina HiSeq 4000 platform (Illumina Inc., San Diego, USA) following the manufacturer's instructions for  $2 \times 150$  paired-end chemistry.

The quality of raw sequences was checked using FastQC (56). Leading and trailing low-quality bases (with Phred score <3), sliding 4-bp windows with average Phred score <20 Phred, and Illumina adapter contaminations were removed using Trimmomatic v0.36 (52). A reference transcriptome for liver tissue was assembled de novo in Trinity v2.8.6 by using default settings in the Inchworm, Chrysalis, and Butterfly plugins (57). The quality of the assembled reference transcriptome was evaluated in QUAST v4.6.3 (58). Quality-filtered sequences were mapped against assembled transcripts using the semi-global aligner BWA-MEM v0.7.12 (59). Resulting Sequence Alignment Map (SAM) files were converted to binary format and sorted in Samtools v1.6 (60). Only sequences with a minimum alignment quality score of  $\geq 30$  Phred (probability of erroneous alignment  $\leq 0.001$ ) in the SAM files were selected for downstream variant calling.

A combination of Samtools v1.6 mpileup and VarScan v2.3.9 mpileup2snp commands (61) was used to detect variant sites within sardine transcriptome. In VarScan, the minimum number of reads to call a variant site was set to eight reads, the *P* value was set to a

maximum 0.001, and all variants with >90% support on one strand (forward or reverse reads) were removed. Indels and variant sites with more than 20% missing data were removed using VCFtools (54). The resulting Variant Format Calling (VCF) file was indexed using the tabix command in Samtools, and sites that showed a linkage disequilibrium coefficient ( $r^2$ ) >80% (55) and genotyping quality less than 50 were removed in BCFtools v1.6.33 (62). Details about the trinity assembly are shown in table S4. Details on the functional content and core biochemical pathways of the liver transcripts are provided in (63).

### Identification of candidate and selectively neutral SNPs

Population structure was assessed with the genome data using both selectively neutral and candidate adaptive SNPs. The latter were identified using consensus between  $F_{ST}$ -based genome scans and GEAs using water temperature as the explanatory variable, while datasets of neutral SNPs were generated by removing a larger candidate dataset identified using more relaxed settings. Outlier detection by means of genome scans was performed using the program BayeScan v2.1 (64), using a burn-in of 100,000 iterations, 20 pilot runs of 5000 iterations, a thinning interval of 100, and 10,000 recorded iterations, with prior odds (PO) set to 10 and a false discovery rate (FDR) of 0.05. To identify SNPs under environmental selection (i.e., GEAs), we compiled data for the following predictor variables for each site: SST, salinity, dissolved oxygen, nitrate, phosphate, and silicate. It was assumed that during the last month of each sardine's life, the fish did not move beyond the  $0.25^\circ \times 0.25^\circ$  area in which it was caught. For each location, NOAA optimally interpolated  $0.25^\circ$  resolution daily SST data (65) were extracted for 30 days leading up to each fish catch. Several descriptors, including minimum, maximum, and mean SST over the 30-day period, as well as the 5th and 95th percentiles, were computed. Salinity at 5 m depth was obtained from the World Ocean Atlas (WOA) (66) using the objectively interpolated climatological (2005–2017) mean at  $\frac{1}{4}^\circ$  resolution. The salinity of each sample was assigned on the basis of the  $\frac{1}{4}^\circ$  block in which it was collected or to the nearest  $\frac{1}{4}^\circ$  block. Data for the remaining environmental variables, namely, dissolved O<sub>2</sub>, nitrate, phosphate, and silicate (all at 5 m depth) were also obtained from the WOA, in this case, using objectively interpolated climatological means (all available years) at  $1^\circ$  resolution. As there were several sardine samples collected in  $1^\circ$  blocks for which there were no WOA data (particularly inshore), contour plots for each environmental variable were generated in Surfer ([www.goldensoftware.com/products/surfer](http://www.goldensoftware.com/products/surfer)) using available data around the coast and within the region  $16^\circ\text{E}$  to  $32^\circ\text{E}$  and  $29^\circ\text{S}$  to  $38^\circ\text{S}$ . The environmental parameter value for each sample location was assigned on the basis of the contour plots.

Linear regression analyses identified strong correlations between all pairs of variables ( $P < 0.01$  for salinity versus nitrate and  $P < 0.001$  for all others; table S5), which precluded the use of multivariate methods for identifying candidate loci (e.g., redundancy analysis). For that reason, we used the 5th and 95th percentiles of SST as environmental predictor variables, as these data were available at and before the time of capture and the univariate GEA method gINLAnd (67) to identify loci under selection. This method corrects for geographic structure and population history and thus reduces the occurrence of false positives. Although correcting for geographic distance also reduces the signal from true positives (68), this was considered necessary given the strong correlations between

environmental parameters and geography along the South African coastline. It is well known that coarse resolution SST products can illustrate large biases close to the coast (69) and do not always adequately resolve smaller transient features (21), particularly in narrow shelf regions such as the east coast of South Africa. However, the SST data provided sufficient distinction between the broadly defined geographical regions (Fig. 1). Specifically, there was a clear west-to-east gradient (Fig. 1), with west coast sites being coldest and east coast sites being warmest.

To minimize the occurrence of false positives, only loci with a  $\log_{10}$  Bayes factor ( $\log\text{BF}$ ) of  $>1.3$  ("strong support") (70) were considered to be under selection. Geographical coordinates were corrected for coastline curvature using the melfuR (<https://rdrr.io/github/pygmyperch/melfuR/>) script viamaris, with extent.buffer set to 5 and default settings for all other parameters, followed by the mass v7.3-47 (71) script isoMDS, with  $k$  set to 2. In their core range, sardines are normally constrained to an SST of  $<20^\circ\text{C}$  (22) but may experience warmer conditions during the sardine run (20). To reduce the effect of sardines being present at unfavorably high SST, we excluded all sardines caught on the east coast (i.e., north of  $32^\circ\text{S}$ ) from the gINLAnd analyses. The procedure for identifying outlier loci from the transcriptome data was identical, but since no loci were identified using genome scans, we selected a more stringent  $\log\text{BF}$  of  $>2.2$  ("overwhelming support") for the gINLAnd analyses.

Datasets of selectively neutral loci were created by relaxing conditions for detection of candidate loci (BayeScan: PO = 1, FDR = 0.05; gINLAnd:  $\log\text{BF} = 0.25$ , i.e., "hardly worth mentioning") and removing the resulting larger number of candidate loci from the complete dataset. The final genomic dataset comprised 8296 loci, of which 63 and 198 loci were identified as candidate loci using gINLAnd and BayeScan, respectively. Eleven loci that were identified by both methods were used in subsequent analyses. The transcriptome dataset comprised 14,973 loci, 234 of which were identified as candidate loci. Using more relaxed conditions for detection of candidate loci to be removed from the genomic dataset to create a dataset of neutral loci, 85 and 495 loci were identified using gINLAnd and BayeScan, respectively. Of these, 26 were shared, resulting in a total of 554 unique loci that were removed from the complete dataset, with 7742 loci remaining. A total of 1960 putative candidate loci were removed from the transcriptome dataset to create a dataset of neutral loci, but as this exceeded the column limit of GenAlEx, the dataset was further reduced to 8191 loci (corresponding to 16,384 columns).

### Assessment of population structure

Population structure using neutral SNPs was assessed using fastSTRUCTURE (72), ADMIXTURE (73), and PCoA. For fastSTRUCTURE, default settings were used, and  $K$  was identified using the chooseK algorithm. ADMIXTURE was run with default settings, and the most suitable  $K$  was identified by comparing the magnitude of cross-validation errors. The PCoA was conducted in GenAlEx v6.5 (74). Raw distance data were converted to standardized covariance matrices using the algorithm of Orlóci (75).

Population structure for candidate SNPs was assessed using the same methods as above, except that the fastSTRUCTURE analysis was replaced by an analysis using STRUCTURE 2.3.4 (76), with the "locprior" model (77) implemented. We specified a burn-in of 1 million generations followed by 10 million iterations and assessed genetic structure for one to four clusters ( $K$ ). For each value of  $K$ , 10 runs were conducted, and consistency of results was assessed in

CLUMPAK (78). The same program was used to generate CLUMPP (79) bar plots, and both  $L(K)$  (76) and  $\Delta K$  (80) in STRUCTURE HARVESTER (81) were used to identify the best  $K$  for each dataset.

## SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <https://science.org/doi/10.1126/sciadv.abf4514>

## REFERENCES AND NOTES

1. J. G. C. Hopcraft, J. M. Morales, H. L. Beyer, M. Börner, E. Mwangomo, A. R. E. Sinclair, H. Oliff, D. T. Haydon, Competition, predation, and migration: Individual choice patterns of Serengeti migrants captured by hierarchical models. *Ecol. Monogr.* **84**, 355–372 (2014).
2. E. J. Torres-Romero, I. Morales-Castilla, M. Á. Olalla-Tárraga, Bergmann's rule in the oceans? Temperature strongly correlates with global interspecific patterns of body size in marine mammals. *Glob. Ecol. Biogeogr.* **25**, 1206–1215 (2016).
3. C. D. van der Lingen, J. C. Coetze, L. Hutchings, Overview of the KwaZulu-Natal sardine run. *Afr. J. Mar. Sci.* **32**, 271–277 (2010).
4. S. F. Dudley, G. Cliff, Influence of the annual sardine run on catches of large sharks in the protective gillnets off KwaZulu-Natal, South Africa, and the occurrence of sardine in shark diet. *Afr. J. Mar. Sci.* **32**, 383–397 (2010).
5. S. H. O'Donoghue, L. Drapeau, V. M. Peddemors, Broad-scale distribution patterns of sardine and their predators in relation to remotely sensed environmental conditions during the KwaZulu-Natal sardine run. *Afr. J. Mar. Sci.* **32**, 279–291 (2010).
6. G. G. Newman, Migration of the Pilchard (*Sardinops ocellata*) in Southern Africa. *Investigational Report, Division of Sea Fisheries, South Africa* **86**, 1–6 (1970).
7. C. Lett, J. Veitch, C. D. van der Lingen, L. Hutchings, Assessment of an environmental barrier to transport of ichthyoplankton from the southern to the northern Benguela ecosystems. *Mar. Ecol. Prog. Ser.* **347**, 247–259 (2007).
8. J. C. Coetze, C. D. van der Lingen, L. Hutchings, T. P. Fairweather, Has the fishery contributed to a major shift in the distribution of South African sardine? *ICES J. Mar. Sci.* **65**, 1676–1688 (2008).
9. H. S. Grantham, E. T. Game, A. T. Lombard, A. J. Hobday, A. J. Richardson, L. E. Beckley, R. L. Pressey, J. A. Huggett, J. C. Coetze, C. D. van der Lingen, S. L. Petersen, D. Merkle, H. P. Possingham, Accommodating dynamic oceanographic processes and pelagic biodiversity in marine conservation planning. *PLOS ONE* **6**, e16552 (2011).
10. N. Mhlongo, D. Yemane, M. Hendricks, C. D. van der Lingen, Have the spawning habitat preferences of anchovy (*Engraulis encrasicolus*) and sardine (*Sardinops sagax*) in the southern Benguela changed in recent years? *Fish. Oceanogr.* **24**, 1–14 (2015).
11. D. C. M. Miller, C. L. Moloney, C. D. van der Lingen, C. Lett, C. Mullon, J. G. Field, Modelling the effects of physical–biological interactions and spatial variability in spawning and nursery areas on transport and retention of sardine *Sardinops sagax* eggs and larvae in the southern Benguela ecosystem. *J. Mar. Syst.* **61**, 212–229 (2006).
12. A. M. McGrath, J. C. Hermes, C. L. Moloney, C. Roy, G. Cambon, S. Herbet, C. D. van der Lingen, Investigating connectivity between two sardine stocks off South Africa using a high-resolution IBM: Retention and transport success of sardine eggs. *Fish. Oceanogr.* **29**, 137–151 (2020).
13. T. Sakamoto, C. D. van der Lingen, K. Shiari, T. Ishimura, Y. Geja, J. Petersen, K. Komatsu, Otolith  $\delta^{18}\text{O}$  and microstructure analyses provide further evidence of population structure in sardine *Sardinops sagax* around South Africa. *ICES J. Mar. Sci.* **77**, 2669–2680 (2020).
14. C. D. van der Lingen, J. C. Coetze, H. Demarq, L. Drapeau, T. P. Fairweather, L. Hutchings, An eastward shift in the distribution of southern Benguela sardine. *GLOBEC Int. Newslett.* **11**, 17–22 (2005).
15. C. D. van der Lingen, M. D. Durholtz, T. P. Fairweather, Y. Melo, "Spatial variability in biological characteristics of southern Benguela sardine and the possible existence of two stocks. *Marine and Coastal Management Report No. MCM/2009/SWG-PEL/39* (2009).
16. C. L. de Moor, D. S. Butterworth, Assessing the South African sardine resource: Two stocks rather than one? *Afr. J. Mar. Sci.* **37**, 41–51 (2015).
17. C. L. de Moor, D. S. Butterworth, C. D. van der Lingen, The quantitative use of parasite data in multistock modelling of South African sardine (*Sardinops sagax*). *Can. J. Fish. Aquat. Sci.* **74**, 1895–1903 (2017).
18. C. D. van der Lingen, L. F. Weston, N. N. Ssempe, C. C. Reed, Incorporating parasite data in population structure studies of South African sardine *Sardinops sagax*. *Parasitology* **142**, 156–167 (2015).
19. S. L. Hampton, "Multidisciplinary investigation into stock structure of small pelagic fishes in southern Africa," thesis, University of Cape Town (2014).
20. J. C. Coetze, D. Merkle, L. Hutchings, C. D. van der Lingen, M. van den Berg, M. D. Durholtz, The 2005 KwaZulu-Natal sardine run survey sheds new light on the ecology of small pelagic fish off the east coast of South Africa. *Afr. J. Mar. Sci.* **32**, 337–360 (2010).
21. M. J. Roberts, C. D. van der Lingen, C. Whittle, M. van den Berg, Shelf currents, lee-trapped and transient eddies on the inshore boundary of the Agulhas Current, South Africa: Their relevance to the KwaZulu-Natal sardine run. *Afr. J. Mar. Sci.* **32**, 423–447 (2010).
22. M. J. Armstrong, P. Chapman, S. F. J. Dudley, I. Hampton, P. E. Malan, Occurrence and population structure of pilchard *Sardinops ocellatus*, round herring *Etrumeus whiteheadi* and anchovy *Engraulis capensis* off the east coast of southern Africa. *South Afr. J. Mar. Sci.* **11**, 227–249 (1991).
23. C. D. van der Lingen, M. Hendricks, M. D. Durholtz, G. Wessels, C. Mtengwane, Biological characteristics of sardine caught by the beach-seine fishery during the KwaZulu-Natal sardine run. *Afr. J. Mar. Sci.* **32**, 309–330 (2010).
24. P. Fréon, J. C. Coetze, C. D. van der Lingen, A. D. Connell, S. H. O'Donoghue, M. J. Roberts, H. Demarq, C. G. Attwood, S. J. Lamberth, L. Hutchings, A review and tests of hypotheses about causes of the KwaZulu-Natal sardine run. *Afr. J. Mar. Sci.* **32**, 449–479 (2010).
25. B. Chiazzari, "Population connectivity of sardines (*Sardinops sagax*) of the KZN sardine run using meristic, morphological and genetic data," MSc thesis, University of KwaZulu-Natal (2014).
26. C. Carreras, V. Ordóñez, L. Zane, C. Kruschel, J. Nasto, E. Macpherson, M. Pascual, Population genomics of an endemic Mediterranean fish: Differentiation by fine scale dispersal and adaptation. *Sci. Rep.* **7**, 43417 (2017).
27. J. Sandoval-Castillo, N. A. Robinson, A. M. Hart, L. W. S. Strain, L. B. Beheregaray, Seascapes genomics reveals adaptive divergence in a connected and commercially important mollusc, the greenlip abalone (*Haliotis laevigata*), along a longitudinal environmental gradient. *Mol. Ecol.* **27**, 1603–1620 (2018).
28. P. R. Teske, J. Sandoval-Castillo, T. R. Golla, A. Emami-Khoyi, M. Tine, S. von der Heyden, L. B. Beheregaray, Thermal selection as a driver of marine ecological speciation. *Proc. Roy. Soc. B* **286**, 20182023 (2019).
29. I. Coscia, S. B. Wilmes, J. E. Ironside, A. Goward-Brown, E. O'Dea, S. K. Malham, A. D. McDevitt, P. E. Robins, Fine-scale seascape genomics of an exploited marine species, the common cockle *Cerastoderma edule*, using a multi-modelling approach. *Evol. Appl.* **13**, 1854–1867 (2020).
30. D. H. Davies, The South African pilchard (*Sardinops ocellata*). Migration 1950–55. *Investigational Report, Division of Sea Fisheries, South Africa* **24**, 1–52 (1956).
31. D. Baird, Seasonal occurrence of the pilchard *Sardinops ocellata* on the east coast of South Africa. *Investigational Report, Division of Sea Fisheries, South Africa* **96**, 1–19 (1971).
32. E. H. Schumann, Low frequency fluctuations off the Natal coast. *J. Geophys. Res. Oceans* **86**, 6499–6508 (1981).
33. W. S. Goschen, E. H. Schumann, K. S. Bernard, S. E. Bailey, S. H. P. Deyzel, Upwelling and ocean structures off Algoa Bay and the south-east coast of South Africa. *Afr. J. Mar. Sci.* **34**, 525–536 (2012).
34. C. S. Russo, T. Lamont, G. C. O. Tutt, M. A. van den Berg, R. G. Barlow, Hydrography of a shelf ecosystem inshore of a major Western boundary current. *Estuar. Coast. Shelf Sci.* **228**, 106363 (2019).
35. P. E. Ihssen, H. E. Booke, J. M. Casselman, J. M. McGlade, N. R. Payne, F. M. Utter, Stock identification: Materials and methods. *Can. J. Fish. Aquat. Sci.* **38**, 1838–1855 (1981).
36. P. R. Teske, S. von der Heyden, C. D. McQuaid, N. P. Barker, A review of marine phylogeography in southern Africa. *South Afr. J. Sci.* **107**, 43–53 (2011).
37. A. T. Lombard, Marine component of the National Spatial Biodiversity Assessment for the development of South Africa's National Biodiversity Strategic and Action Plan (2004).
38. P. R. Teske, T. R. Golla, J. Sandoval-Castillo, A. Emami-Khoyi, C. D. van der Lingen, S. von der Heyden, B. Chiazzari, B. J. van Vuuren, L. B. Beheregaray, Mitochondrial DNA is unsuitable to test for isolation by distance. *Sci. Rep.* **8**, 8448 (2018).
39. L. Hutchings, C. D. van der Lingen, L. J. Shannon, R. J. M. Crawford, H. M. S. Verheyen, C. H. Bartholomae, A. K. van der Plas, D. Louw, A. Kreiner, M. Ostrowski, Q. Fidel, R. G. Barlow, T. Lamont, J. Coetze, F. Shillington, J. Veitch, J. C. Currie, P. M. S. Monteiro, The Benguela Current: An ecosystem of four components. *Prog. Oceanogr.* **83**, 15–32 (2009).
40. J. R. E. Lutjeharms, O. Boebel, P. C. F. van der Vaart, W. P. M. de Ruijter, T. Rossby, H. L. Bryden, Evidence that the Natal Pulse involves the Agulhas Current to its full depth. *Geophys. Res. Lett.* **28**, 3449–3452 (2001).
41. J. R. E. Lutjeharms, O. Boebel, H. T. Rossby, Agulhas cyclones. *Deep-Sea Res. II* **50**, 13–34 (2003).
42. G. M. Leber, L. M. Beal, S. Eliot, Wind and current forcing combine to drive strong upwelling in the Agulhas Current. *J. Phys. Oceanogr.* **47**, 123–134 (2017).
43. H. Demarq, R. Barlow, L. Hutchings, Application of a chlorophyll index derived from satellite data to investigate the variability of phytoplankton in the Benguela ecosystem. *Afr. J. Mar. Sci.* **29**, 271–282 (2007).
44. A. D. Connell, A 21-year ichthyoplankton collection confirms sardine spawning in KwaZulu-Natal waters. *Afr. J. Mar. Sci.* **32**, 331–336 (2010).
45. B. A. Robertson, R. L. Hutto, A framework for understanding ecological traps and an evaluation of existing evidence. *Ecology* **87**, 1075–1085 (2006).
46. M. Rouault, B. Pohl, P. Penven, Coastal oceanic climate change and variability from 1982 to 2009 around South Africa. *Afr. J. Mar. Sci.* **32**, 237–246 (2010).

47. M. Krug, *Oceans and Coasts Annual Science Report, 2020*, S. P. Kirkman, J. A. Huggett, T. Lamont, M. C. Pfaff, Eds. (Department of Forestry, Fisheries and the Environment, 2021).

48. J. D. Scott, M. A. Alexander, D. R. Murray, D. Swales, J. Eisched, The Climate Change Web Portal: A system to access and display Climate and Earth System Model Output from the CMIP5 Archive. *Bull. Am. Meteorol. Soc.* **97**, 523–530 (2016).

49. J. Augustyn, A. Cockroft, S. Kerwath, S. Lambeth, J. Gitahaiga-Mwicigi, G. Pitcher, M. Roberts, C. van der Lingen, L. Auerswald, in *Climate Change Impacts on Fisheries and Aquaculture: A Global Analysis*, B. Phillips, M. Pérez-Ramírez, Ed. (John Wiley & Sons, 2018), pp. 479–522.

50. B. K. Peterson, J. N. Weber, E. H. Kay, H. S. Fisher, H. E. Hoekstra, Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLOS ONE* **7**, e37135 (2012).

51. J. Catchen, P. A. Hohenlohe, S. Bassham, A. Amores, W. A. Cresko, Stacks: An analysis tool set for population genomics. *Mol. Ecol.* **22**, 3124–3140 (2013).

52. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

53. J. B. Puritz, C. M. Hollenbeck, J. R. Gold, dDocent: A RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ* **2**, e431 (2014).

54. P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, 1000 Genomes Project Analysis Group, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

55. W. G. Hill, A. Robertson, Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* **38**, 226–231 (1968).

56. S. Andrews, FastQC, 2017; [www.bioinformatics.babraham.ac.uk/projects/fastqc/](http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

57. B. J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li, M. Lieber, M. D. MacManes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C. N. Dewey, R. Henschel, R. D. LeDuc, N. Friedman, A. Regev, De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).

58. A. Gurevich, V. Saveliev, N. Vyahhi, G. Tesler, QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).

59. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 [q-bio.GN] (2013).

60. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

61. D. C. Koboldt, K. M. Steinberg, D. E. Larson, R. K. Wilson, E. R. Mardis, The next-generation sequencing revolution and its impact on genomics. *Cell* **155**, 27–38 (2013).

62. H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).

63. A. Emami-Khoyi, R. Le Roux, M. G. Adair, D. M. Monsanto, D. C. Main, S. P. Parbhoo, C. M. Schnelle, C. D. van der Lingen, B. Jansen van Vuuren, P. R. Teske, Transcriptomic diversity in the livers of South African sardines participating in the annual sardine run. *Genes* **12**, 368 (2021).

64. M. Foll, O. Gaggiotti, A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics* **180**, 977–993 (2008).

65. R. W. Reynolds, T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, M. G. Schlax, Daily high-resolution-blended analyses for sea surface temperature. *J. Clim.* **20**, 5473–5496 (2007).

66. H. E. Garcia, T. P. Boyer, O. K. Baranova, R. A. Locarnini, A. V. Mishonov, A. Grodsky, C. R. Paver, K. W. Weathers, I. V. Smolyar, J. R. Reagan, D. Seidov, M. M. Zweng, *World Ocean Atlas 2018: Product Documentation*, A. Mishonov, Ed. (National Centers for Environmental Information, 2018).

67. G. Guillot, R. Vitalis, A. le Rouzic, M. Gautier, Detecting correlation between allele frequencies and environmental variables as a signature of selection. A fast computational approach for genome-wide studies. *Spat. Stat.* **8**, 145–155 (2014).

68. S. Yeaman, K. A. Hodgins, K. E. Lotterhos, H. Suren, S. Nadeau, J. C. Degner, K. A. Nurkowski, P. Smets, T. Wang, L. K. Gray, K. J. Liepe, A. Hamann, J. A. Holliday, M. C. Whitlock, L. H. Rieseberg, S. N. Aitken, Convergent local adaptation to climate in distantly related conifers. *Science* **353**, 1431–1433 (2016).

69. J. Xie, J. Zhu, Y. Li, Assessment and inter-comparison of five high-resolution sea surface temperature products in the shelf and coastal seas around China. *Cont. Shelf Res.* **28**, 1286–1293 (2008).

70. A. J. Drummond, R. R. Bouckaert, *Bayesian Evolutionary Analysis with BEAST* (Cambridge Univ. Press, 2015).

71. W. N. Venables, B. D. Ripley, *Modern Applied Statistics with S* (Springer, ed. 4, 2002).

72. A. Raj, M. Stephens, J. K. Pritchard, fastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics* **197**, 573–589 (2014).

73. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).

74. R. Peakall, P. E. Smouse, GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* **28**, 2537–2539 (2012).

75. L. Orlóci, *Multivariate Analysis in Vegetation Research* (Dr. W. Junk B.V. Publishers, ed. 2, 1987).

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

77. M. J. Hubisz, D. Falush, M. Stephens, J. K. Pritchard, Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Resour.* **9**, 1322–1332 (2009).

78. N. M. Kopelman, J. Mayzel, M. Jakobsson, N. A. Rosenberg, I. Mayrose, CLUMPAK: A program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* **15**, 1179–1191 (2015).

79. M. Jakobsson, N. A. Rosenberg, CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**, 1801–1806 (2007).

80. G. Evanno, S. Regnaut, J. Goudet, Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* **14**, 2611–2620 (2005).

81. D. A. Earl, B. M. von Holdt, STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **4**, 359–361 (2012).

**Acknowledgments:** We are grateful to Y. Geja and other Fish Team members (DFFE) for collecting sardine and to the Centre for High Performance Computing (CHPC) for the use of computational resources. We also thank R. Waples and an anonymous reviewer for comments on an earlier version of this manuscript. Ethical clearance for sampling was granted by the Faculty of Science Ethics Committee, University of Johannesburg (approval code: 19022015).

**Funding:** NRF CSUR (grant no. 87702) to P.R.T., C.D.v.d.L., C.D.M., and L.B.B.; ARC Future Fellowship (FT130101068) to L.B.B.; NRF (grant no. 64801) to C.D.M.; and DFFE funding to T.L. and C.D.v.d.L. **Author contributions:** Conceptualization: P.R.T., C.D.M., L.B.B., and C.D.v.d.L. Analysis: P.R.T., A.E.-K., T.R.G., J.S.-C., T.L., and C.D.v.d.L. Visualization: P.R.T., T.L., and C.D.v.d.L. Supervision: P.R.T. Writing (original draft): P.R.T. Writing (review and editing): P.R.T., B.C., T.L., C.D.M., L.B.B., and C.D.v.d.L. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Genomic data are available from figshare (<https://figshare.com/s/0fd2e9e601e8a4a50348>).

Submitted 27 October 2020

Accepted 26 July 2021

Published 15 September 2021

10.1126/sciadv.abf4514

**Citation:** P. R. Teske, A. Emami-Khoyi, T. R. Golla, J. Sandoval-Castillo, T. Lamont, B. Chiazzari, C. D. McQuaid, L. B. Beheregaray, C. D. van der Lingen, The sardine run in southeastern Africa is a mass migration into an ecological trap. *Sci. Adv.* **7**, eabf4514 (2021).

## The sardine run in southeastern Africa is a mass migration into an ecological trap

Peter R. TeskeArsalan Emami-KhoyiTirupathi R. GollaJonathan Sandoval-CastilloTarron LamontBrent  
ChiazzariChristopher D. McQuaidLuciano B. BeheregarayCarl D. van der Lingen

*Sci. Adv.*, 7 (38), eabf4514. • DOI: 10.1126/sciadv.abf4514

### View the article online

<https://www.science.org/doi/10.1126/sciadv.abf4514>

### Permissions

<https://www.science.org/help/reprints-and-permissions>